ATMÓSFERA

Revista trimestral publicada por el
Instituto de Ciencias de la Atmósfera y Cambio Climático
de la Universidad Nacional Autónoma de México

Volumen 36
Número 2
1 de abril de 2023

CONTENTS

MÉXICO 2023

INFORMATION FOR CONTRIBUTORS

ATMOSFERA seeks contributions on theoretical, empirical and applied research in all the areas of atmospheric sciences, with emphasis on meteorology, climatology, aeronomy, physics, chemistry, and aerobiology. Interdisciplinary contributions are also accepted; especially those related with oceanography, hydrology, ecology, forestry, glaciology, agriculture, environmental pollution, and other topics related to economy and society as they are affected by climate variability and change, and atmospheric hazards.

ATMOSFERA publishes articles, short contributions and correspondence in English. The submission of a contribution to be considered an article is assumed to indicate that no similar paper, other than an abstract or preliminary report, has been or will be submitted for publication elsewhere. Short contributions should not exceed 1000 words. Correspondence is intended for discussion of papers published in the journal and should not exceed 250 words.

Manuscripts. Please submit an electronic copy of your final version of the manuscript, preferably prepared with MS Word or compatible software, through the journal's webpage at http://www.revistascca.unam.mx/atm/index.php/atm/about/submissions#onlineSubmissions. It should be formatted with wide margins, double-spacing and all pages numbered. Footnotes to the text should be avoided. When symbols in mathematical work could be in doubt, the authors should submit a list identifying typographically the symbols used. Equations will be numbered in parentheses to the right.

Title, name, address. Provide the title in full and suggest a short title for headers on odd-numbered pages. Author's name, affiliation, e-mail and present postal address will not be included in the document, but they should be captured in the "Submission Metadata" section.

Abstract. All articles shall include an abstract, both in Spanish and in English. Abstracts are not required for short contributions.

Text. The text should be divided into sections, each with a separate heading.

Acknowledgements. Brief.

Appendices. Supplementary materials and lengthy mathematical analyses, whose details are subordinate to the main theme of the paper, should normally appear in an appendix. Each appendix should have a title.

Units, abbreviations and symbols. The International System of Units (m, kg, s, K) should be used. Only standardized terms that have been generally accepted should be used. Unfamiliar abbreviations must be defined when first used. Particular care should be taken to ensure that all symbols are clearly identified. The minus sign should be –.

References. Articles should be fully documented with references in the text and a reference listing ordered alphabetically by author's names at the end of the paper. Digital object identifiers (DOIs) should be provided whenever possible for each reference. Authors are encouraged to adjust to the citation system. Examples:

Articles in journals: Hoerling M.P. and Kumar A., 2002. Atmosphere response patterns associated with tropical forcing. J. Climate 15, 2184-2203.

Chapters in books: Magaña V. and Conde C., 2003. Climate variability and climate change and their impacts on the freshwater resources in the border region: A case study for Sonora, México. In: Climate and water-transboundary challenges in the Americas (Díaz H.F. and Morehouse B.S., Eds.). Kluwer Academic Publishers, Dordrecht, 373-393.

Articles in proceedings: Gallardo L., Córdova A.M., Rojas M., Quintana J., Alcafuz R. and Ramos I., 2004. Stratosphere-troposphere exchange (STE) over Easter Island and Cerro Tololo stations in South America: analysis and simulation of an intensive sounding campaign. 8th International Global Atmospheric Chemistry Conference, Christchurch, New Zealand, September 2004.

Books: Thompson R.D., 1998. *Atmospheric processes and systems.* Routledge, New York, 194 pp.

References cited in the text should indicate, within parenthesis, the last name of each author and the year of publication. If there are more than two authors, the last name of the first author should be followed by et al. When the name forms part of the content of the article, and does not appear solely as a reference, the year of publication is parenthesized. For example:

...a weekly cycle has been observed in surface temperature (Gordon, 1994).

...Gordon (1994) has observed a weekly cycle in surface temperature.

Illustrations and tables. Vector graphics should be supplied in a vector format (eps, ps, ai, etc.) or as pdfs. Photographs should be provided as tiff or jpeg images (with at least 300 dpi); the same applies to line drawings (with at least 1000 dpi). Graphics created in Microsoft Office applications should be supplied in their original format. Figure captions should be self-explanatory and printed on a separate sheet of paper. All figures should be numbered and specifically mentioned in the text by number. Please number all tables with Roman numerals and include a short title.

Proofs. Copies of the proofs will be sent to the author. Revised proofs should be returned to the publisher without delay. The only changes that could be accepted in proofs are on grammar or related to printing errors.

Authors can print copies of their articles from: http://www.revistascca.unam.mx/atm/index.php/atm.

ATMÓSFERA

Revista trimestral publicada por el
Instituto de Ciencias de la Atmósfera y Cambio Climático
de la Universidad Nacional Autónoma de México

MÉXICO 2023

Atmósfera is included in: Web of Science (WoS), SCOPUS, SCISEARCH, Research Alert, GeoBase, Current Contents Connect, and Aquatic Sciences and Fisheries Abstracts (ASFA).

Registrada en el Sistema de Clasificación de Revistas Mexicanas de Investigación Científica y Tecnológica del Consejo Nacional de Ciencia y Tecnología (CONACyT), e indizada en PERIÓDICA.

Manuscripts should be submitted through our website at: https://www.revistascca.unam.mx/atm/.

Subscriptions and all correspondence should be sent to:
Editor of *Atmósfera*, Instituto de Ciencias de la Atmósfera y Cambio Climático, UNAM, Circuito Exterior, Ciudad Universitaria, 04510 Coyoacán, Ciudad de México. Tel.: 52 (55) 5622-4240 Ext. 81888; fax: 52 (55) 5616-0789; e-mail: editorial@atmosfera.unam.mx

Subscription information: Institutional rate: US $90.00 per volume (4 issues); Private rate: US $70.00 per volume (4 issues).

The complete texts of Atmósfera are available in digital format at:
https://www.revistascca.unam.mx/atm/; Redalyc: https://redalyc.org; SciELO: https://www.scielo.org.mx/.

Atmósfera , vol. 36, núm. 2, abril-junio de 2023, es una publicación trimestral editada por la Universidad Nacional Autónoma de México, Ciudad Universitaria, 04510 Coyoacán, Ciudad de México, teléfono (55) 5622-4240, ext.81888, https://www.revistascca.unam.mx/atm/. Editor responsable: Graciela B. Raga. Reserva de derechos al uso exclusivo del título núm. 001559/94 otorgado por el Instituto Nacional del Derecho de Autor. ISSN: 0187-6236. Certificado de licitud de título y contenido 7719 y 5464 otorgados por la Comisión Calificadora de Publicaciones y Revistas Ilustradas de la Secretaría de Gobernación.

Printed in Mexico.

Atmósfera 36(2), 183-205 (2023)
https://doi.org/10.20937/ATM.53013

Relationship between rainfall and streamflow in the La Plata Basin: Annual cycles, interdecadal and multidecadal variability

Carla GULIZIA[1,2,3]* and Inés CAMILLONI[1,2,3]

[1]*Universidad de Buenos Aires, Facultad de Ciencias Exactas y Naturales, Departamento de Ciencias de la Atmósfera y los Océanos, C1428 Ciudad Autónoma de Buenos Aires, Argentina.*
[2]*CONICET-Universidad de Buenos Aires, Centro de Investigaciones del Mar y la Atmósfera (CIMA), C1428 Ciudad Autónoma de Buenos Aires, Argentina.*
[3]*CNRS-IRD-CONICET-Universidad de Buenos Aires, Instituto Franco-Argentino para el Estudio del Clima y sus Impactos (IRL 3351 IFAECI), C1428 Ciudad Autónoma de Buenos Aires, Argentina.*
*Corresponding author: gulizia@cima.fcen.uba.ar

Received: December 4, 2020; accepted: July 26, 2021

RESUMEN

Se analiza la relación entre la variabilidad de la precipitación y los caudales en la cuenca del Plata (LPB) a lo largo de un amplio rango de escalas temporales. La LPB se divide en seis subcuencas asociadas a los principales ríos regionales (Paraguay, Paraná, Uruguay e Iguazú). Se analiza la amplificación de la respuesta del caudal para evaluar en qué medida la variabilidad de las descargas de los ríos puede explicarse por fluctuaciones de la precipitación. Se analizan los ciclos anuales medios correspondientes al período 1931-2010 y a cada una de las décadas que lo integran. La mayoría de las estaciones de aforo presentan cambios interdecadales en los caudales. Se aplica un filtro de promedio móvil de 11 años a las series anuales normalizadas. Las series filtradas de caudales exhiben un porcentaje considerablemente mayor de la varianza explicada, destacando el predominio de la variabilidad de baja y media frecuencias en comparación con la de precipitación. Las descargas de los ríos muestran una densidad espectral mayor en las frecuencias interdecadales/multidecadales en comparación con el análisis de precipitación. Para avanzar en la comprensión de la relación precipitación-caudal se emplea un enfoque estadístico simple con resultados prometedores: los espectros de caudal se derivan directamente del espectro de precipitación, transformados por un operador de "cuenca". Se asume que las subcuencas actúan sobre la precipitación como integradores espaciotemporales que operan como filtros de flujo bajo, como medias móviles. Los espectros correspondientes a los caudales se simulan asumiendo que siguen un proceso de medias móviles autorregresivas y se ajustan efectivamente a las observaciones.

ABSTRACT

The aim of this study is to understand the interaction between rainfall and streamflow variability in the La Plata basin (LPB) along a wide range of timescales. The LPB is divided in six sub-basins associated to the main regional rivers (Paraguay, Paraná, Uruguay and Iguazú). The amplification of the streamflow response is addressed in order to evaluate to what extent river discharges variability can be explained by precipitation fluctuations. Mean annual cycles corresponding to the period 1931-2010 and to each decade of this interval are analyzed. Streamflow interdecadal changes are observed in most of the gauging stations. In addition, an 11-year moving-average filter is applied to the normalized annual time series. Results exhibit a considerable higher percentage of explained variance in the streamflow filtered series, highlighting the predominance of medium and low frequencies variability present in these compared to those of precipitation. Consistently, river discharges show higher spectral density in the interdecadal/multidecadal frequencies compared to precipitation analysis. A simple statistical approach to advance in the understanding of the complex rainfall-streamflow physical relationship is addressed with promising results: streamflow spectrums are derived directly from the precipitation spectrum, transformed by a "basin" operator, characteristic of the basin itself. It is assumed that watersheds act on precipitation as spatiotemporal integrators operating as low-pass filters, like a moving

average. Streamflow power spectrums are simulated assuming that the underlying process is an autoregressive moving average. Considering the sub-basin areal-averaged precipitation time series as the only input, results show that simulated streamflow spectrums fit effectively the observations at the sub-basin scale.

Keywords: La Plata basin, rainfall, streamflow, climate variability, South America.

1. Introduction

The La Plata basin (LPB) is the fifth largest basin in the world, and the second in South America, after the Amazon basin. It covers 3.2 million km^2 (Pasquini and Depetris, 2007), comprising southern and southeastern Brazil, southeastern Bolivia, a great part of Uruguay, all of Paraguay and an extensive area of central and northern Argentina. With 57 cities of over 110 000 inhabitants, it currently concentrates more than 70% of those countries' gross domestic product (CIC, 2016). There are at least 20 cities with more than 500 000 inhabitants, including the capitals of four of the countries that make up the LPB: Buenos Aires, Brasilia, Asunción, and Montevideo. In addition, the city of São Paulo, in Brazil, is located on one of the tributaries of the Paraná River and it is one of the largest megalopolis and industrial concentrations in the world, with more than 20 million people (CIC, 2017). Other economic activities along the LPB are based on agriculture, livestock and the production of hydroelectric energy. The LPB hydroelectric potential is estimated in 93 000 MW, of which 66% has already been used or is going to be used (CIC, 2016).

Different aspects of the hydroclimatic variability and trends in the LPB have been addressed by several authors (García and Vargas, 1996; Tucci and Clarke, 1998; Krepper et al., 2006; Doyle and Barros, 2011; Maciel et al., 2013; Antico et al., 2014; Boulanger et al., 2016; Lovino et al., 2018, Shi et al., 2019, Zanin and Satyamurty, 2020a, b). In the last decades the discharge interannual variability of the main rivers of the LPB and its climate forcing have also been analyzed (Mechoso and Pérez-Iribarren, 1992; Depetris et al, 1996; García and Vargas, 1996, 1998; Genta et al., 1998; Robertson and Mechoso, 1998, 2000; Bischoff et al., 2000; Camilloni and Barros, 2000, 2003; Berri et al., 2002; Krepper et al., 2003, among others). Krepper and García (2004), applying a low-passed-filter to anomalies of monthly precipitation in the LPB, concluded that the Uruguay, Paraná and Paraguay rivers behavior is a clear reflection of the precipitation at the interannual frequency.

Focusing on longer time scales, Genta et al. (1998) examined the long discharge series of the major rivers in southeastern South America (SESA) searching for long-period changes. Robertson and Mechoso (1998) analyzed the annual discharges of Negro, Paraguay, Paraná and Uruguay rivers for the period 1911-1993, identifying a non-linear trend and near decadal component, particularly notorious over the Paraná and Paraguay rivers, where high discharges were associated to cool sea surface temperature (SST) over the tropical North Atlantic. Meanwhile, Robertson and Mechoso (2000) distinguished a north-south dipole in streamflow anomalies with a predominant interdecadal component. These authors found that the differences in river discharges were consistent with the dipolar structure of the atmospheric vertical motion anomalies related to the South Atlantic Convergence Zone (SACZ): enhanced streamflows to the north (the Paraná and Paraguay rivers) and diminished discharges to the south (the Uruguay and Negro rivers), coinciding with years in which the SACZ intensified.

Moreover, Krepper et al. (2006, 2008) analyzed the response of the Paraguay River basin to seasonal precipitation as well as the low frequency response of the upper Paraná River basin. Results from both studies show that the lower Paraguay River basin presents a near decadal cycle while the Paraná River in the Posadas discharge gauging station is modulated by a cycle of around nine years. Maciel et al. (2013) studied the multi-annual variability of observed streamflows for Paraná, Paraguay and Negro rivers during the 20th century. Results from their analysis presented significant low frequency variability with increasing trends that showed a clear seasonality in the three rivers. Antico et al. (2014), using a novel empirical decomposition, found that the Paraná River modes consist of annual and intra-annual oscillations related to the rainfall seasonality, and interannual to

interdecadal changes associated to modes of variability such as El Niño/Southern Oscillation (ENSO), the North Atlantic Oscillation (NAO), and the Interdecadal Pacific Oscillation. Non-linear trends of the river streamflows were also linked by the authors to global warming.

River discharges represent an integrated response to catchment heterogeneity and spatial variability of the key hydrological processes present in a watershed. In this sense, they reflect the climatic inputs and contain more information of climate variability than that provided by the scarce observational precipitation and temperature records (García and Vargas, 1996). Although rainfall and streamflow are intrinsically related, their relationship is complex and not all the precipitation signals may be reflected in the river discharges. Moreover, several authors have observed that the percentage change and the low frequency variability presented on streamflow tends to be amplified compared to that of precipitation (Chiew et al., 1995; Berbery and Barros, 2002; Krepper et al., 2006). In the LPB, precipitation enhancements have produced positive variations of about 35% in the total discharges (Genta et al., 1998; Berbery and Barros, 2002) as well as in the flood frequency events in the Paraná River (Camilloni and Barros, 2003). Milliman et al. (2008) examined annual discharges of 137 rivers located in representative areas of the world for the period 1951-2000 and showed that the streamflow trends of most analyzed rivers were forced by changes in precipitation. Meanwhile, Dai et al. (2009) studied 200 of the main rivers worldwide for the period 1948-2004 and found significant streamflow trends for 64 of them explained by precipitation trends.

Since fluctuations in rainfall cannot be linearly reflected into streamflow variations (Chiew et al., 1995), further studies focused on understanding the regional hydrological variability in the context of climate variability and change are essential in order to generate more reliable information that may be relevant to hydrologists and water resources managers. Consequently, this paper intends to better understand the relationship between rainfall and river discharges on different timescales (including a lag analysis of their annual cycles and their interdecadal/multidecadal variability) in the LPB through the application of statistical processes analysis. In the present study, it is assumed that watersheds act on precipitation as spatiotemporal integrators operating as low-pass filters, like a moving average. Thus, streamflow power spectrums at sub-basins scale will be simulated assuming that the underlying process is an autoregressive moving average (ARMA).

The article is structured in five sections. Section 2 consists in the description of the data and methods employed in the present study. Section 3 presents the statistical characterization of precipitation and streamflow variability on different time scales whereas section 4 shows the implications of streamflow temporal variability in applying a transfer function to precipitation time series. Finally, Section 5 summarizes results and conclusions.

2. Data and methods

2.1 Data and region of study

The LPB lies between the Andes mountains to the west, and the Brazilian Plateau and Sierra do Mar" to the northeast and east (see Fig. 1). The western boundary of the LPB includes the Andes mountains, with altitudes varying from 1000 to 4000 masl. The eastern boundary of the basin has a mean altitude of 1000 masl, although it can be as high as 1500 masl in the east to less than 2 masl in the most downstream part of the basin (Barros et al., 2006; CIC, 2016). For the present study, the LPB was divided in six sub-basins as shown in Figure 1: Paraguay, Pantanal, Uruguay, Iguazú, lower and upper Paraná. This subdivision was based on the rainfall catchment areas which were delimited by topographic information. The Paraná's catchment area is the most extensive one and covers more than 80% of the LPB area. On the other hand, the Uruguay River basin is the only large river in the LPB that it is not tributary of the Paraná. It comprises around 10% of the LPB total extension. Moreover, the Paraguay River basin, which contains a vast swamp called the Pantanal, has an elevation that rarely exceeds 70 masl, and its gradient is typically lower than 1.5 cm km^{-1} (Coronel et al., 2006). The slopes in this basin are considerably smaller compared to the eastern portion of the LPB.

The river gauge stations considered as closing points of each watershed were selected from Saurral et al. (2013), except for the upper and lower Paraná River where nearby stations with a longer period

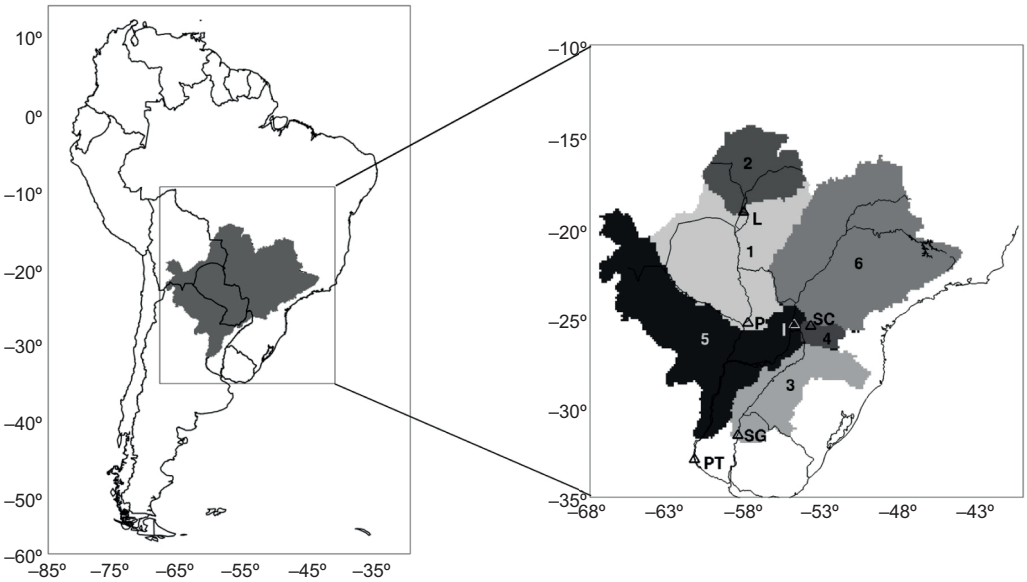

Fig. 1. Left panel: lower Paraná catchment area in South America. Right panel: catchment areas of the six sub-basins within the La Plata Basin (1: Paraguay, 2: Pantanal, 3: Uruguay, 4: Iguazú, 5: lower Paraná, 6: upper Paraná). Locations of the corresponding closing points are indicated: P: Pilcomayo, L: Ladario, SG: Salto Grande, SC: Salto Caxias, PT: Paraná Timbúes and I: Itaipú.

of available data were used. Monthly naturalized streamflow data derived from the gauging stations considered as closing points (see Table I) were gathered from Brazil's Operador Nacional do Sistema Elétrico (National Electricity System Operator), which provides the streamflow in hydroelectric power plants, and from Argentina's Base de Datos Hidrológica Integrada del Sistema Nacional de Información Hídrica (Integrated Hydrological Data Base of the National Water Information System). The delimitation of each sub-basin and the location of their closing points are shown in Figure 1. The study period is 1931-2010 for all gauging stations except for Ladario in the Paraguay River, where data are available for a shorter period (1931-1998). Monthly precipitation was estimated for each sub-basin as the areal average of the gridded dataset derived from the Global Precipitation Climatology Centre version 6 (GPCC V6) (Schneider et al., 2015). This dataset represents adequately the key spatial patterns and the temporal variability of precipitation over the study region (Gulizia and Camilloni, 2015).

In order to describe the main hydroclimatological features of the LPB sub-basins annual mean values of the areal average precipitation rate ($\overline{P}$), streamflow ($\overline{Q}$), evapotranspiration ($\overline{ET}$), the ratios ($\overline{Q}\,\overline{P}^{-1}$) (δ)

Table I. Gauging stations considered as closing points for each sub-basin. Their geographical location and the associated river are also indicated.

Sub-basin	River	Closing Point	Latitude	Longitude
Paraguay	Paraguay	Pilcomayo	25.15º S	57.31º W
Pantanal	Paraguay	Ladario	19.00º S	57.59º W
Lower Paraná	Paraná	Paraná Timbúes	32.72º S	60.73º W
Upper Paraná	Paraná	Itaipú	25.24º S	54.35º W
Uruguay	Uruguay	Salto Grande	31.38º S	57.95º W
Iguazú	Iguazú	Salto Caxias	25.32º S	53.29º W

and (γ) were calculated for the 1931-2010 period. $\overline{ET}$ was estimated as the difference between $\overline{P}$ and $\overline{Q}$ for each sub-basin, assuming a simplified water balance approach in which changes in terrestrial water storage were considered negligible (Walter et al., 2004; Huntington and Billmire, 2014). Thus, the estimation of $\overline{ET}$ is simplified as follows: $\overline{ET} = \overline{P} - \overline{Q}$. It is noteworthy to mention that in such an estimate the effect of water recycling may be ignored as, for example, precipitation reaching the ground can be evaporated before streamflow occurs. Table II presents the area of each sub-basin along with *P*, *Q*, *ET*, and . Particularly, the importance of $\overline{ET}$ stands out in the Paraguay and Paraná rivers where γ is high over the whole basins with values ranging from 72% to 88%. On the other hand, the Uruguay and Iguazú basins are the smallest ones but with the highest *Q*/*P* ratios (39% and 57%, respectively (Table II). In the particular case of the Iguazú basin, most of the rainfall contributes with surface runoff.

2.2 Methods

The hydrological regimes for each of the six sub-basins are characterized through the annual cycles of the observed streamflow as well as the mean areal averaged precipitation. This analysis was performed for the complete study period and for each of the eight decades that comprise it.

Spectral analysis based on the maximum entropy method (MEM) (Ulrych and Bishop, 1975; Wu, 2012) was applied to study extensively the complex relationship between observed streamflow and the sub-basin precipitation areal averages. The maximum entropy procedure consists of looking for the probability distribution that maximizes the entropy of the information, subject to the information limitations (Wu, 2012). The analysis was particularly focused to explore the amplification of medium and low frequencies variability (e.g., interdecadal and multidecadal variability, respectively) in streamflow compared to precipitation. This method was previously applied in several studies to analyze air temperature and precipitation cycles and to capture streamflow low frequency variability modes and pseudo-periodic patterns (Leite and Peixoto, 1995; Maciel et al., 2013; Li et al., 2015). The main feature of MEM, compared to the other available methodologies, is to assume that the time series of interest follow an autoregressive process of order *p* (AR[*p*]) (whose coefficients are estimated from observations) and takes advantage of the exact mathematical relationship between these and the theoretical spectral density of the AR processes (Box and Jenkins, 1994; Ghil et al., 2002) for extracting a smooth spectrum.

The correlation structure of continuous variable time series (e.g., monthly streamflow and precipitation) can often be represented successfully using a class of time series models known as Box-Jenkins models, after the classic text by Box and Jenkins (1994). The simplest Box-Jenkins model is the first-order autoregression, or AR(*1*), model. As the name suggests, one way of viewing the AR(*1*) model is as a simple linear regression where the predictand is the value of the time series at time $t+1$ (x_{t+1}) and the predictor is the current value of the time series (x_t).

Table II. Mean values of precipitation ($\overline{P}$), streamflow ($\overline{Q}$) and evapotranspiration ($\overline{ET}$), estimated as the difference between precipitation and streamflow) considering the 1931-2010 period. Also indicated are δ (streamflow to precipitation ratio), γ (evapotranspiration to precipitation ratio) and the catchment area for each of the six sub-basins.

Sub-basin (closing point)	$\overline{P}$ ($m^3\ s^{-1}$)	$\overline{Q}$ ($m^3\ s^{-1}$)	$\overline{ET}$ (m^3 s)	δ (%)	γ (%)	Area km^2
Paraguay (Pilcomayo)	29 007	3495	25 512	12	88	7.87E+05
Pantanal (Ladario)	10 693	1321	9372	12	88	2.52E+05
Lower Paraná (Paraná Timbúes)	89 229	15 803	73 426	18	82	2.25E+06
Upper Paraná (Itaipú)	36 129	10 239	25 890	28	72	7.88E+05
Uruguay (Salto Grande)	12 344	4806	7539	39	61	2.43E+05
Iguazú (Salto Caxias)	2334	1339	995	57	43	3.86E04

For the present study, higher order autoregressions, in which the parameter estimation is more complex, were applied. For instance, the regression equation predicting x_{t+1} can be expanded to include data values progressively further back in time as predictors. The estimation of the p autoregressive parameters is most easily done using the set of equations relating them to the autocorrelation function, which are known as the Yule-Walker equations (Box and Jenkins, 1994; Wilks, 2006a). These calculations were done applying the Yule-Walker function provided in Matlab software.

The MEM framework was applied in a univariate way for analyzing areal average rainfall and streamflow time series, respectively. Given x_t a time series and using the lag operator L definition such that $Lx_t = x_{t-1}$, by induction it can be shown that:

$$L^k x_t = x_{t-k} \tag{1}$$

where k $\in Z$ and t = 1,⋯, n.

Therefore, the AR(p) can be expressed as follows:

$$\phi(L)x_t = \varepsilon_t \tag{2}$$

where $\phi(L)= 1 - \sum^{p}_{k=1} \alpha_k L^k$ is the autoregressive polynomial with (α_1, α_2, α_p) the model parameters and p is a non-negative integer. ε_t is white noise consisting of a sequence of independent zero-mean unit-variance Gaussian random variables. For the present study, the time series x_t expressed in Eq. (2) will correspond to areal average precipitation P_t and observed streamflow Q_t, respectively.

Similarly, a moving average process of order q or MA(q) can be expressed briefly as in Eq. (3) using the polynomial of order q in the lag operator, $\Theta(L)$.

$$x_t = \Theta(L)\varepsilon_t \tag{3}$$

where $\Theta(L) = \sum^{q}_{j=1} \beta_j L^j$.

In order to understand to what extent precipitation variability is able to explain streamflow variability, the streamflow power spectrum was simulated by extending the MEM regular framework, assuming that the underlying process is an autoregressive moving-average of order p and q (ARMA(p,q)) (Eq. [4]). Under a linear model, and considering that precipitation follows an AR(p) process (i.e., the temporal dependence of precipitation itself), it may be assumed that the river discharges Q_t follow an ARMA(p,q) process instead.

$$\phi(L)Q_t = \Theta(L)\varepsilon_t \tag{4}$$

The moving-averages weights were obtained based on a spatiotemporal weighted sum of rainfall, on all prior periods and the entire sub-basin, respectively. Particularly, under MEM framework, the latter also implies that streamflow spectrum can be obtained from the spectrum of precipitation ($S_P(f)$) (see Eq. [5]) transformed by a mathematical operator which is essentially a low-pass filter whose exhaustive mathematical description is obtained from the weights associated with the space-time integration, characteristic of the sub-basin itself. Under the assumptions described above the simulated spectral density for streamflow ($\hat{S_Q}(f)$) is derived in Eq. (6):

$$S_P(f) = \frac{1}{\left|\phi(e^{2i\Pi f})\right|^2} \tag{5}$$

$$\hat{S}_Q(f) = \frac{\left|\Theta(e^{2i\Pi f})\right|^2}{\left|\phi(e^{2i\Pi f})\right|^2} = \left|\Theta(e^{2i\Pi f})\right|^2 \times S_P(f) \tag{6}$$

A key aspect to perform this analysis consists in estimating these weights in space and time from streamflow and areal average rainfall time series. By evaluating the autocorrelation coefficients, it can be determined the spectral density, which is associated to the most random or least predictable process that has the same autocorrelation coefficients. In terms of the information theory (Shannon, 1949), this corresponds to the concept of maximum entropy, hence the name of the method. When applying MEM, as the order increases the number of peaks in the power spectrum increases too. Therefore, determination of the AR order p is crucial for analyzing time series spectrum. Otherwise, if the order chosen is too large, spurious peaks may appear. An upper bound of $N2^{-1}$ is generally taken, where N is the length of the time series (Ghil et al., 2002). From the available model selection criteria that are commonly used to choose the order of the AR model, in the present study the Akaike Information Criterion (AIC) (Akaike, 1974) was applied which involves the maximum likelihood

added to a penalty that is an increasing function of the number of estimated parameters. In computing AIC, it is worth mentioning that the Yule-Walker estimates do not produce maximum likelihood estimates, so the likelihood computed here might not be "maximal", but it would be an approximation of it. The penalty aims at avoiding overfitting since adding more parameters would improve the likelihood (l), but at the same time increase its complexity (Wilks, 2006b). The order p chosen as the most appropriate corresponds to the minimum AIC as expressed in Eq. (7). For the present study, Table III presents the AIC selected orders p and q used for the univariate AR and for the ARMA processes analysis, respectively.

$$\hat{p} = argmin\{AIC_p\} = argmin\ \{-2\log lp + 2p\} \quad (7)$$

The most relevant feature of the filter process is that the output spectral density over a given frequency ($\hat{S_Q}$ (f)) (Eq. [6]) is the product of the square of the modulus of the transfer function and the input spectral density over that given frequency (S_P (f)). Applying a transfer function on a time series implies two main effects: on one hand, it can modify the relevance of some or all frequencies in the variability decomposition. Secondly, by performing the spectral decomposition of the P_t time series, then its individual spectral components will be multiplied by the complex valued spectral transfer function, which thus alters their phase, that is, the moment at which a given frequency is dominant.

Under the statistical framework previously described, the delay between rainfall and river discharges as a function of the different frequencies is also calculated for each sub-basin. The observed rainfall-streamflow annual cycles lags are compared with the simulated delays obtained when the frequency equals unity, thus, corresponding to the annual cycle analysis. The latter is performed in order to examine the reliability to capture the correct streamflow regimes when applying the transfer function, detailed above, to areal average precipitation time series. The precipitation and streamflow linear trends are also evaluated for each sub-basin and gauging station, respectively. These are calculated applying the linear regression technique using least square estimates. Their statistical significance is evaluated using the t-Student test considering 90, 95 and 99% significant levels.

Table III. Order p (months) for the AR model selected for precipitation and streamflow time series and order q (months) for the ARMA model.

Sub-basin	Order p (Precipitation)	Order p (Streamflow)	Order q
Paraguay	13	27	13
Pantanal	26	25	14
Lower Paraná	36	25	24
Upper Paraná	48	38	23
Uruguay	15	28	10
Iguazú	18	16	10

3. Characterization of the precipitation and streamflow temporal variability

The precipitation in the sub-basins of the LPB are dominated by different mechanisms. The tropical basins, which cover the Paraguay (which includes the Pantanal) and the Paraná rivers, are mainly affected by the SACZ, though the southern portions of these basins are also influenced by the South American Low-Level Jet (SALLJ) (Zanin and Satyamurty, 2020a). Precipitation over the subtropical basin, such as in the Uruguay and Iguazú rivers, depend on the SALLJ (Robertson and Mechoso, 2000). In this sense, Salio et al. (2002) showed that the moisture transported by SALLJ represents 55% of the summer precipitation over northeastern Argentina and 35% over subtropical La Plata basin. Figure 2 shows the annual cycle of the areal average precipitation over the six sub-basins of the LPB for the period 1931-2010 and for each decade. The Paraguay, Pantanal and lower and upper Paraná sub-basins show a comparable and well-defined precipitation annual cycle with maxima during the austral summer, associated to the South American Monsoon (SAM) system (Su and Lettenmaier, 2009). The most intense precipitation falling over the Paraná basin is associated with the summer occurrence of the SACZ in the northeastern sector of this basin. In winter, the SACZ vanishes and rainfall is very low in this region. The SACZ also drives the annual cycle dominating the seasonality of the Paraguay River (Antico et al., 2014). On the contrary, Uruguay and Iguazú sub-basins present different annual precipitation regimes of smaller am-

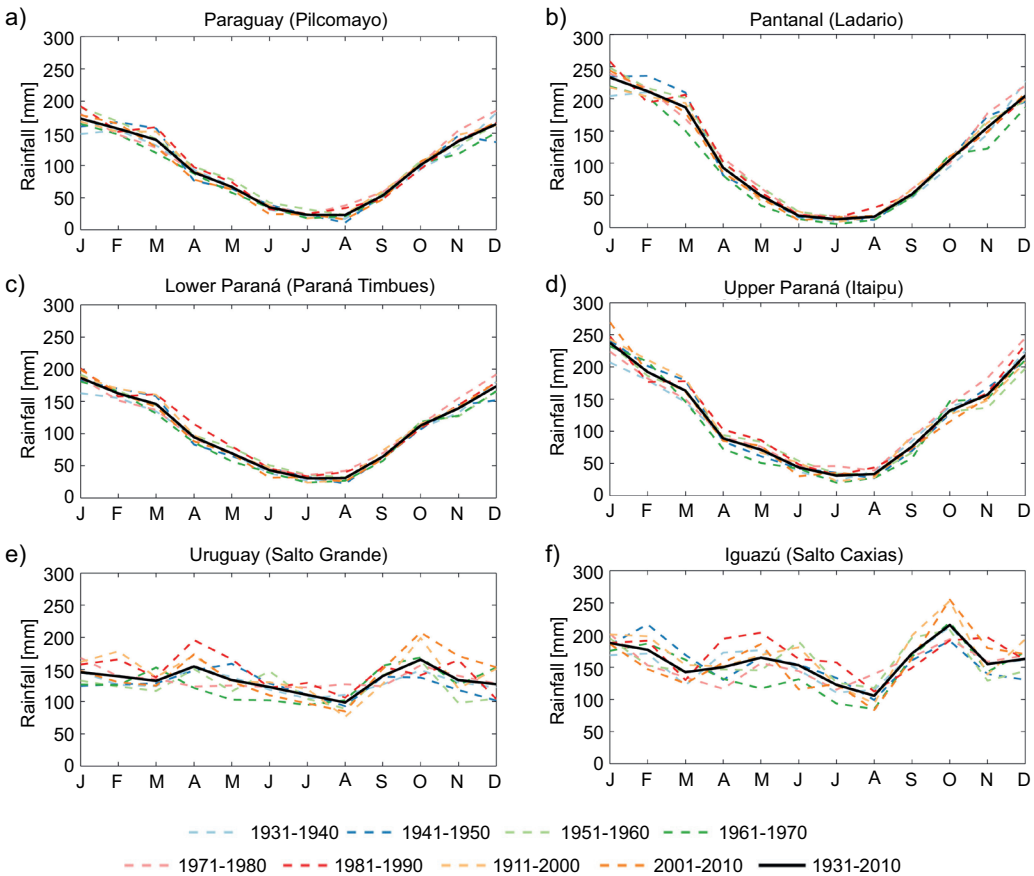

Fig. 2. Mean areal average rainfall annual cycles (mm) corresponding to 1931-2010 and to each of the decades that comprise the complete study period, for the six sub-basins: (a) Paraguay, (b) Pantanal, (c) lower Paraná, (d) upper Paraná, (e) Uruguay, (f) Iguazú. For reference, in parenthesis, the closing points considered for each sub-basin are also indicated.

plitude and not as clearly defined as in the rest of the LPB sub-basins. In the Iguazú basin, local maxima are present in summer, winter and spring. The summer one reflects the rainfall intensification associated with the SACZ. The remaining two are a result of the incursion of cold fronts, coming from the South Atlantic Ocean, in winter and spring (García and Vargas, 1996; Camilloni and Barros, 2000; Antico et al., 2014), which are related to a high frequency of cyclogenesis with higher water vapor content (Gan and Rao, 1991; Rao et al., 1996). Particularly, the Uruguay basin has the highest extratropical influence (García and Mechoso, 2005). Though there is a regularity of precipitation along the year, the largest amount of precipitation takes place in the northern portion (Nery et al., 2006).

Figure 2 also depicts precipitation annual cycles considering the different decades comprising the entire 80-years period. In the Paraná and Paraguay rivers sub-basins, the rainfall regimes are similar between the studied decades. Some variations of small magnitude are observed during the austral winter months, when precipitation is minimum (these variations in the magnitude are better perceived when applying a logarithmic scale into the y-axis of Fig. 2 [not shown]). Along the Uruguay and Iguazú rivers, much more marked interdecadal variations in the precipitation regimes are observed, varying not only in magnitude but also in the occurrence of the lowest and highest values.

Figure 3 presents the annual cycle of the streamflow at the selected gauging stations in each sub-basin

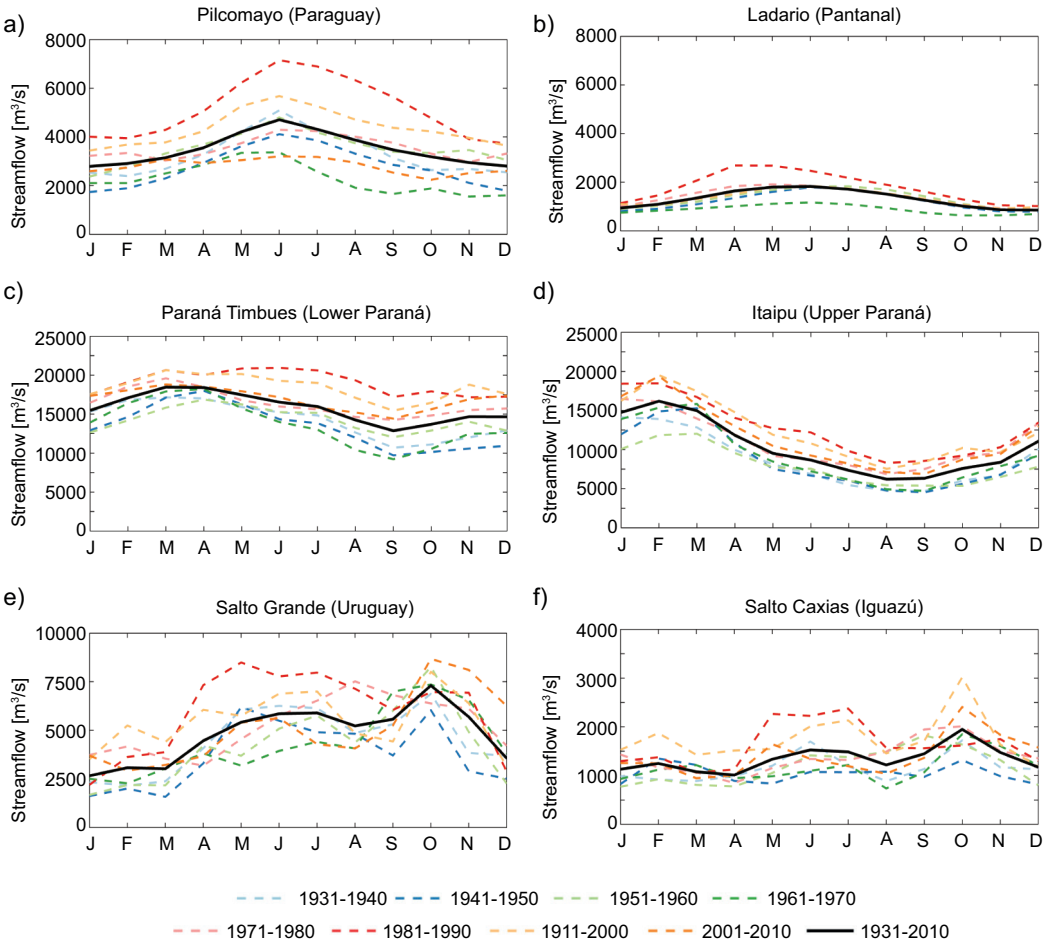

Fig. 3. Mean streamflow annual cycles (m^3 s^{-1}) corresponding to 1931-2010 and to each of the decades that comprise the complete study period, for the six gauging stations: (a) Pilcomayo, (b) Ladario, (c) Paraná Timbúes, (d) Itaipú, (e) Salto Grande, (f) Salto Caxias. The sub-basins to which each gauging station correspond are indicated in parenthesis.

for the same periods as in Figure 2. Three main streamflow regimes can be distinguished. First, the Paraguay River at Pilcomayo and Ladario exhibit a well-defined seasonal cycle with maxima during the austral winter and minima between spring and summer months. Secondly, streamflow of the Paraná River at the Itaipú and Paraná Timbúes gauging stations shows the maximum values between February and March (and the minima occur between August and September). Finally, the Uruguay and Iguazú rivers at Salto Grande and Salto Caixas display irregular annual cycles. The Iguazú River annual cycle has a small range because the regional precipitation regime varies little throughout the year. The Uruguay River maximum values occur around June-July and October, while the smaller discharges can be observed between December and March, consistent with previous results (Pasquini and Depetris, 2007; Saurral, 2010). A weakened (intensified) SACZ is associated with a strong (weak) SALLJ, which tends to transport more (less) moisture into the Uruguay River basin. As a result, stronger and more significant streamflow anomalies have been observed in the weak SACZ phase (e.g., strong SALLJ) as shown by Robertson and Mechoso (2000).

The average streamflow lag as a response to precipitation can be analyzed in each sub-basin by comparing the annual cycles corresponding to the

complete study period (Figs. 2 and 3). The lag in the streamflow in the Paraguay River sub-basins is five months. It is important to highlight that the upper portion of the Paraguay River basin, that is the Pantanal sub-basin, has a very flat topography and, thus, suffers from seasonal floods explaining the obtained damped effect of about five months of delay between peak rainfall and peak streamflow (Berbery and Barros, 2002, and references therein; Su and Lettenmaier, 2009). For the Paraná basin this lag is about one month in the upper sub-basin and two months in the lower one, in relation to the peak precipitation during the SAM season. These results are consistent with previous studies as reviewed by Zanin and Satyamurty (2020a). In addition, the comparison of the Paraguay and Paraná streamflow annual cycles shows that the maximum in the Paraguay River is lagged by about four months from the corresponding one to the Paraná River, as a result of the high water storage capacity of the Pantanal sub-basin (García and Vargas, 1996; Camilloni and Barros, 2000). Considering the temporal resolution used in the present study (i.e., monthly data), no apparent lag is observed in Uruguay and Iguazú rivers sub-basins. Particularly, the Uruguay River basin presents heterogeneous topography including valleys and tributaries of short extension; thus, it exhibits a quick response to precipitation (Berbery and Barros, 2002; Zanin and Satyamurty, 2020a). The months of peak precipitation and streamflows in each sub-basin, and the lag between them, are included in the Table SI in the supplementary material. Additionally, for complementary information, Figure SI includes the heatmaps of the precipitation-streamflow correlations at different lags. The larger delays are evident in the Paraná and Paraguay rivers basins (e.g., higher correlations are indicative of a larger delay), while smaller ones are identified in the Uruguay and Iguazú rivers basins (e.g., higher correlations when considering almost no delay between rainfall and streamflow). Overall, the observed precipitation-streamflow delays are mainly dependent on basin features such as vegetation, soil type, topography, areal extension, among others. These, together with the resulting evapotranspiration and the rest of the hydrological cycle components, drive to a large extent the observed delays.

Furthermore, the analysis based on the streamflow annual cycles for each of the decades that comprise the entire period of study shows that most of the rivers present interdecadal changes in the annual cycle regime in magnitude throughout the 12 months and, in some cases, also in the time of occurrence of the peak (Fig. 3). García and Mechoso (2005) analyzed quasi-periodicities in the time series of river streamflows in the LPB identifying a quasi-oscillation of about nine years. This oscillation was related to SST anomalies in the Atlantic Ocean. Particularly the Paraguay River at Pilcomayo is the only case where the higher mean monthly streamflow is observed during the same month (June) for all the studied decades. In addition, variations in magnitude are also observed between the eight decades. The latter river at Ladario gauging station, as well as the Paraná River at both Paraná Timbúes and Itaipú, present maxima with an interdecadal lag of no more than three months, depending on the studied decade, compared to the peak month in the 1931-2010 mean annual cycles. Interdecadal changes in the occurrence of the maximum streamflow are observed in the Uruguay and Iguazú rivers, consistent with the analysis of precipitation at these sub-basins. The SALLJ presents a decadal-multidecadal variability which in turn affects the atmospheric circulation and precipitation patterns over the region (Jones and Carvalho, 2018; Zanin and Satyamurty, 2020b). The observed changes in the annual cycles may be also influenced by the long-term trends as these reinforce the magnitudes of low-level winds and precipitation anomalies during decadal-to-multidecadal variability of the SALLJ. Furthermore, Maciel et al. (2013) showed that pseudo-cycles of eight-nine years appear in Uruguay River at seasonal time scales.

Finally, most of the rivers present the greatest discharges within the recent decades of the study period, and this is observed particularly for the decade 1981-1990, which is probably associated to the extreme El Niño 1982-83 event (Camilloni and Barros, 2000, 2003; Barros et al., 2004; Camilloni, 2005), since the precipitation in the study region is above normal during the warm ENSO phase (Ropelewski and Halpert, 1987, 1989; Boulanger et al., 2005). The higher occurrence of El Niño events over the end of the 20th century compared to the beginning of the study period could be leading the observed decadal difference. This is associated with the positive phase of the Pacific Decadal Oscillation (PDO) (Deser et

al., 2004), the leading pattern of SSTs anomalies in the North Pacific basin. There is a similarity between the ENSO and PDO spatial patterns, except for the relative magnitude of SST anomalies in the Pacific (ENSO emphasizes the equatorial region while PDO the North Pacific). Particularly, the PDO was predominantly positive between 1925 and 1946, negative between 1947 and 1976, and positive since 1977 until 1999 (Mantua et al., 1997). Since 2000, the PDO has been variable with shorter period phases, possibly explained by global warming (Zhang and Delworth, 2016; Li et al., 2020). A cold phase lasted four years (1999-2002), three years of a cold one afterwards, followed by a neutral phase until 2007, and then back again into a cold one through 2013 (Newmann et al., 2016). The last PDO phase shift registered was in 2014, when it turned into a strongly positive warm phase. In addition to the PDO, the Atlantic Multidecadal Oscillation (AMO) is also relevant to the multi-decadal variability of the LPB. The AMO is identified as a coherent pattern of variability in the North Atlantic SSTs with a period around 60 to 80 years (Schlesinger and Ramankutty, 1994). A warm period occurred from 1930 to 1970, while cooler phases occurred during 1902-1925 and 1970-1994. Since 1995, the AMO index has been positive and increasing (Trenberth and Shea 2006). The LPB variability contains a 64-year periodicity due to AMO, with precipitation increase (decrease) during its negative (positive) phase (Zanin and Satyamurty, 2020b). There are other modes of variability that can affect the LPB in different time scales (e.g., SACZ variability [SAV], Atlantic Dipole, and NAO, among others), as reviewed by Zanin and Satyamurty (2020b). For example, Labat et al. (2005) verified that the Paraná River discharges have a 13-14-year periodicity attributed to the SAV and a periodicity of 8.5 years with stronger positive relation with the NAO.

To further study the interdecadal variability of precipitation and streamflow, Figures 4 and 5 present

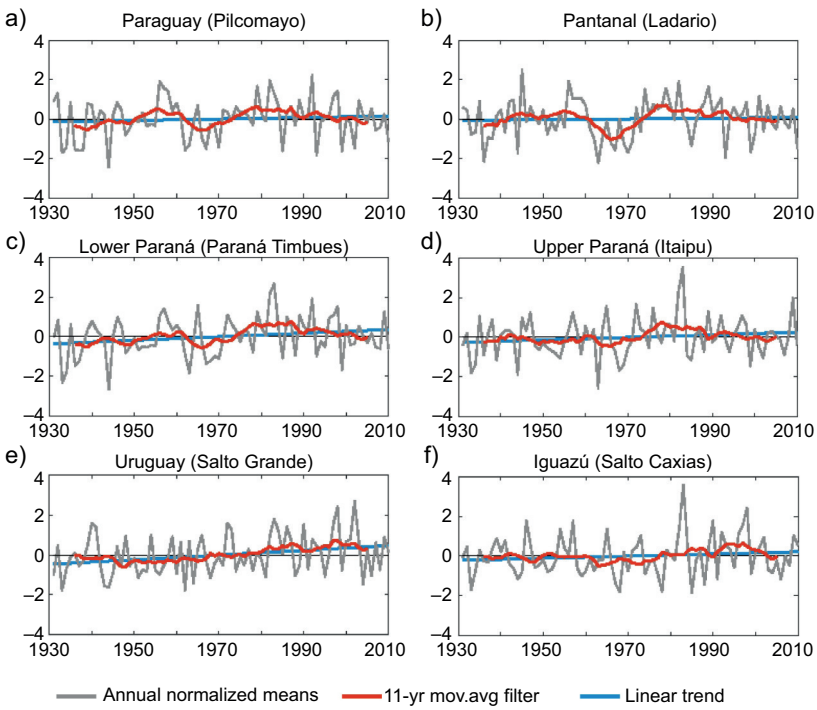

Fig. 4. Normalized annual rainfall together with the linear trend and the 11 year-running mean filtered time series for the period 1931-2010 for each sub-basin, respectively: (a) Paraguay, (b) Pantanal, (c) lower Paraná, (d) upper Paraná, (e) Uruguay, (f) Iguazú. For reference, in parenthesis, the closing points considered for each sub-basin are also indicated.

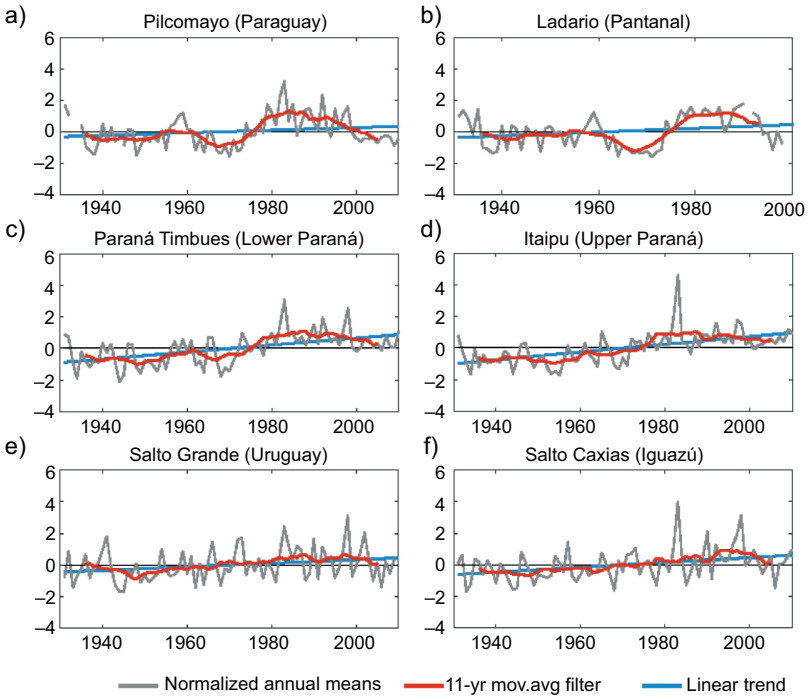

Fig. 5. Normalized annual streamflow together with the linear trend and the 11 year-running mean filtered time series for the period 1931-2010 for each gauging station, respectively: (a) Pilcomayo, (b) Ladario, (c) Paraná Timbúes, (d) Itaipú, (e) Salto Grande, (f) Salto Caxias. The sub-basins to which each gauging station correspond are indicated in parenthesis.

the respective normalized annual time series, together with the linear trend and the corresponding 11-year running means for the period 1931-2010. The trends analysis results exhibit that normalized annual precipitation linear trends are positive in all basins (Fig. 4). Table SII includes the precipitation and streamflow trends for the entire period (1931-2010) expressed in rates (m^3s^{-1}) per decade, respectively. However, only in the Uruguay River and lower Paraná River sub-basins, the linear trends were significant with 95 and 90% of confidence level, respectively. Previous studies have found a significant increase in the total rainfall observed over most of the subtropical South America, east of the Andes, including the middle and south of the LPB during the second half of the twentieth century (Barros et al., 2008; Haylock et al., 2006; Doyle and Barros, 2011; Doyle et al. 2012; Zandonadi et al., 2016; Saurral et al., 2017). The increase in the number of rainfall extremes and the wetter conditions prevailing appear as the main contributors to explain the total precipitation trends in many studies (Penalba and Robledo, 2010; Doyle et al., 2012; Zandonadi et al., 2016; Cerón et al., 2021). The precipitation trends for the period from 1950 to 2000 (not shown), depict significant positive trends in Paraná and Iguazú rivers sub-basins (90% confidence level), and in the Uruguay River basin (95% confidence level). However, the trends in the Paraguay River sub-basins were positive though not significant. The latter may be explained by the use of unique time series for precipitation at the sub-basin scale, calculated as areal averages of the grid points comprising each of them.

Significant positive trends in streamflow are found for the six gauging stations (Fig. 5 and Table SII). Particularly, the flow increase in the Paraná and Paraguay rivers have also been addressed by several authors (e.g., García and Vargas, 1996; Genta et al., 1998; Robertson and Mechoso, 1998; Collischonn et al., 2001; García and Mechoso, 2005). Doyle et

al. (2011) showed that the generalized growth of the river discharges in the LPB could be explained by two main drivers: the increase in precipitation, mainly in the southern sub-basins, and the decrease in evaporation attributable to land-use change, affecting the northern ones to a larger extent. After the 1970s there was a considerable increase in streamflow all over the studied rivers, especially along the Paraguay River (Fig. 5). Changes in precipitation and streamflow based on the annual means between 1931-1970 and 1971-2010 show an increase in precipitation between 4 and 6% in all sub-basins, except for the Uruguay River basin (11%) during the last period (Table IV). However, precipitation increases are amplified when quantifying the changes in the river discharges, which range between 27 and 37%, compared to the period prior to 1970. The election for separating the two sub-periods around 1970 was made to include the same number of years in each period, although this selection is also supported by García and Vargas (1998) results, as they detected an important change of tendency in the LPB discharges between 1970 and 1972.

The 11-year running mean precipitation time series (Fig. 4) display a fraction of medium frequency (inter-decadal) variability in most of the sub-basins. Meanwhile, the streamflows filtered series (Fig. 5) highlight the relevance of the river streamflows medium frequency variability to a higher extent than for precipitation, particularly in the Paraguay River. In addition, Table IV also shows the percentage of total variance explained by the normalized annual precipitation and streamflow time series for each sub-basin, which were filtered by an 11-year moving average. Results evidence the higher relevance of the medium frequency variability (over 40% of total variance explained) in the Paraguay (including Pantanal area) and Paraná rivers discharges compared to those for precipitation. As an example, streamflow at Ladario gauging station exhibits 47% of the explained variance while for precipitation over the Pantanal sub-basin it is only 15%. This difference highlights the importance of the medium frequency variability present in streamflow compared to precipitation. On the contrary, the Uruguay and Iguazú sub-basins show higher frequency variability both in precipitation (11 and 7%, respectively) and streamflow (14 and 20%, respectively) time series. These lower percentages could be related to the size of these sub-basins, in addition to the higher frequency of extratropical transient systems. Several studies found that changes in precipitation are amplified in streamflow over different river basins around the globe (Berbery and Barros, 2002; Chiew, 2006; Fu et al., 2007, among others). This concept is also called in the literature the precipitation elasticity of streamflow, that is, a measure of the sensitivity of streamflow to changes in rainfall (Schaake, 1990; Sankarasubramanian et al., 2001). Regarding the LPB, Berbery and Barros (2002) stated that the variability in precipitation is considerably amplified in the main river streamflows, by a mean factor of 2. The reasons explaining the amplification of precipitation signal in runoff are difficult to assess given that the physics governing

Table IV. Percentages of increase in precipitation and streamflow based on the annual means of the period 1931-1970 and 1971-2010 for each basin ("change between periods"). Also shown are the percentages of the total variance explained by the normalized annual precipitation and streamflow time series for each sub-basin, which were filtered by an 11 year-moving average ("variance explained").

Sub-basin (gauging station)	Change between periods		Variance explained	
	Precipitation (%)	Streamflow (%)	Precipitation (%)	Streamflow (%)
Paraguay (Pilcomayo)	4	36	10	41
Pantanal (Ladario)	4	27	15	47
Lower Paraná (Paraná Timbúes)	6	27	12	46
Upper Paraná (Itaipú)	6	37	8	46
Uruguay (Salto Grande)	11	31	11	14
Iguazú (Salto Caxias)	5	36	7	20

the relationship between interception, infiltration, and runoff are highly nonlinear and also depend on soil and vegetation characteristics (Schaake et al., 1996).

In order to explore the precipitation amplification on streamflow of medium/low frequencies (inter-decadal/multidecadal), Figure 6 presents the MEM power spectrum for the areal average precipitation and the streamflows. A logarithmic *x*-axis scale of the period in years allows focusing on the lower frequencies. As expected, spectral analysis results show that annual precipitation presents a higher spectral power than annual streamflows in most sub-basins while this relation is opposite at the inter-decadal/multi-decadal lower frequencies. Besides, the Uruguay River basin depicts a higher spectral density in the streamflows compared to precipitation, both annually and at the lower frequencies. However, the spectral densities behavior at the annual level is different to the rest of the sub-basins in the LPB. An explanation for this would need additional studies to determine if it could be due to physical features because of the complexity of the basin (e.g,. heterogeneous topography and tributaries of short extension) and/or a methodological artifact. The Iguazú sub-basin exhibits a similar magnitude of the maximum spectral density at the annual scale, for both rainfall and streamflow, probably because of the size of this sub-basin, as in small watersheds the streamflow response to precipitation is faster. Furthermore, a second maximum around 2.5 years can also be noted probably linked to the period of ENSO and its influence on precipitation and streamflow over the different sub-basins.

In summary, in most sub-basins, streamflow variability displays an attenuation with respect to precipitation for the higher frequencies (periods [years] around unity) and an amplification of lower ones. For the Uruguay and Iguazú rivers, though no

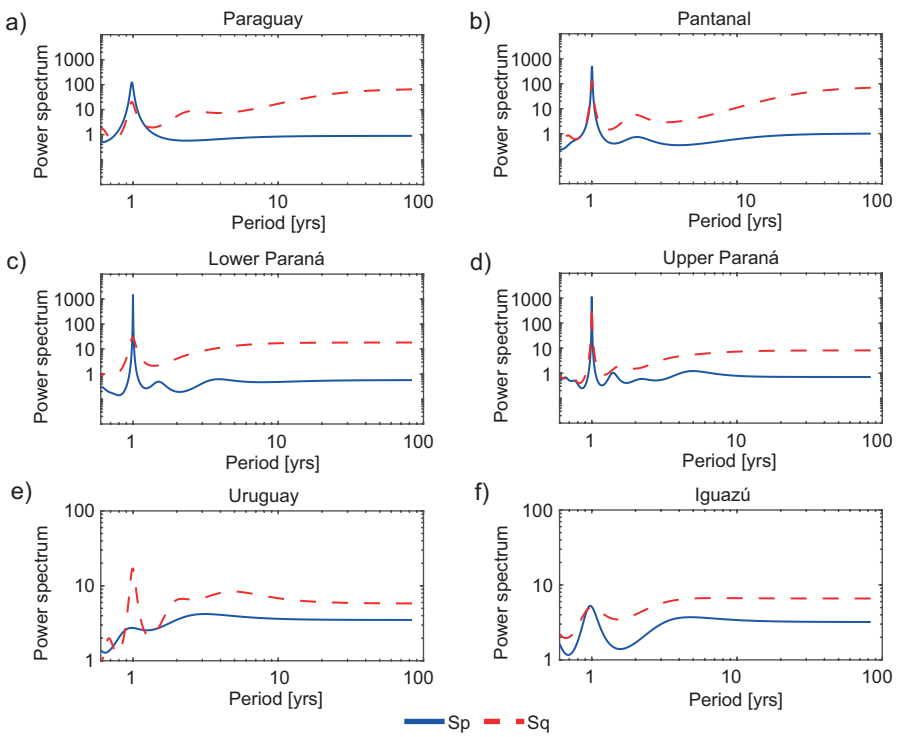

Fig. 6. MEM power spectrums as a function of its period (years), for the areal average precipitation (*Sp*) and for the streamflow at the corresponding closing points (*Sq*) for each sub-basin, respectively: (a) Paraguay, (b) Pantanal, (c) lower Paraná, (d) upper Paraná, (e) Uruguay, (f) Iguazú. Both horizontal and vertical axes are in logarithmic scale.

attenuation is observed in the frequency of the annual cycle of the river discharges, the interdecadal/multidecadal low frequency variability is predominant for the streamflow time series. This result is consistent with the assumption that watersheds in the LPB act on precipitation as spatiotemporal integrators and, therefore, operate as low frequency filters, like moving averages. In this sense, the physical relationship between rainfall and streamflow at the sub-basin scale, in which an amplification of the lower frequencies is observed in the streamflow variability, may be interpreted under an ARMA statistical process. In the following section, this hypothesis will be evaluated.

4. Implications of streamflow temporal variability in applying a transfer function to precipitation time series

The aim of this section is to evaluate if the model, which was estimated through the extension of MEM methodology as explained in section 2, adequately captures, on one hand, the observed delays between rainfall and streamflow; and on the other hand, the streamflow temporal variability through the comparison of both estimated and observed streamflow power spectrums.

Figure 7 shows the simulated lag between rainfall and river discharges for each sub-basin as a function of frequency. In particular, this study focuses on the comparison of observed and simulated seasonal cycles of both variables, that is, when frequency equals unity (see the vertical dotted line in Figure 7). The observed lag for each basin, included in Table SI, is represented by the horizontal line. These lag values were estimated by comparing the occurrence of the maximum in the respective observed precipitation and streamflow annual cycles for the period 1931-2010, as shown in Figures 2 and 3. Results show that the Paraguay and Pantanal sub-basins present a simulated lag between the maximum values of rainfall and streamflow of 4.5 months that is consistent with observations, whilst in the lower Paraná River it is about 2.5 months (slightly higher than the observed lag of two months). The lag in the simulated annual cycles of streamflow in the upper Paraná is around 1.4 months, similar to the observed of one month. As discussed in the previous sections, Uruguay and Iguazú river basins stand out as being the most efficient for surface runoff with lags lower than a month and with good agreement between observations and simulations.

Spectral analysis results of observed and simulated streamflows are presented in Figure 8.

This analysis is performed in order to explore whether the extension of the MEM framework (assuming that the underlying process for streamflow is an ARMA[p,q]), adequately represents the discharges temporal variability. Streamflow spectrums are derived directly from the precipitation spectrum, transformed by a "basin" operator characteristic of the basin itself, which is related to the lags between precipitation and streamflows. The advantage of this novel application is that monthly precipitation time series, areal-averaged at the basin scale, are the only requirement for simulating the streamflow variability at the closing points of each sub-basin. Results indicate that the model captures considerably well the temporal variability of the river discharges in most of the sub-basins since the observed and simulated power spectrums are quite similar. Particularly, a slight underestimation of the observed spectral density for interdecadal/multidecadal frequencies can be noted in the Paraguay River (Fig. 8a, b). In the case of the Uruguay and Iguazú rivers, simulations overestimate the spectral density for these timescales. Additionally, in the Uruguay River sub-basin, maximum spectral density at the annual scale is not well captured. The resulting underestimation can be explained by the fact that the precipitation spectral density at the annual level was considerably smaller than the streamflow one, as shown in the previous section. Overall, the filter ARMA(p,q) effectively fits the observations, particularly to a higher extent in the Paraná River.

5. Summary and conclusions

The purpose of this study was to advance in the understanding of the relationship between precipitation and streamflow variability in the LPB along a wide range of timescales with emphasis on the annual cycles and the interdecadal/multidecadal medium and low frequencies. In this sense, the extent to which precipitation variability can explain streamflow variability was studied through the application of different statistical processes analysis (e.g., AR and

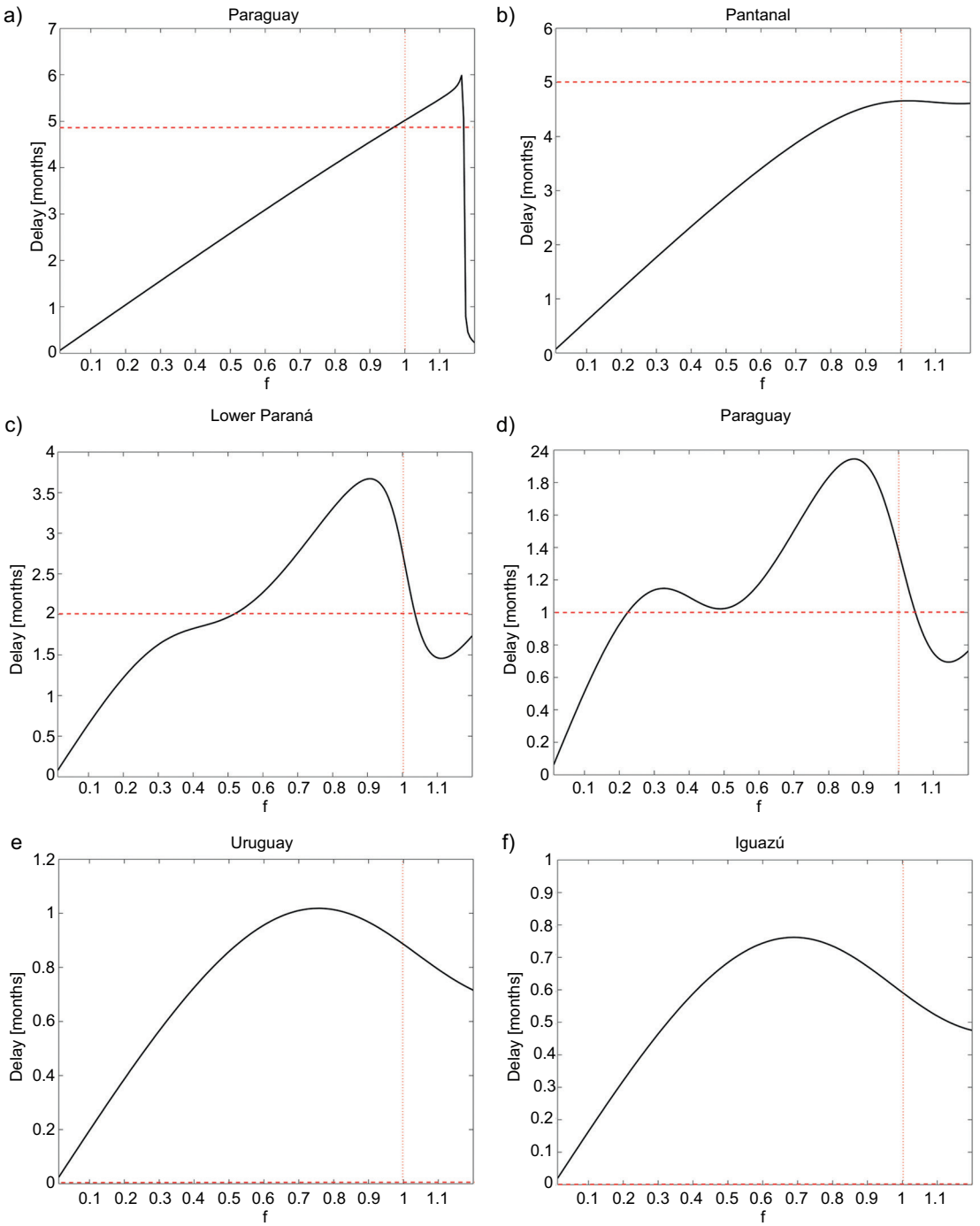

Fig. 7. Simulated lag (months) between rainfall and streamflow as a function of the different frequencies (*f*) for each sub-basin: (a) Paraguay, (b) Pantanal, (c) lower Paraná, (d) upper Paraná, (e) Uruguay, (f) Iguazú. The vertical dotted line indicates where , which corresponds to the annual cycle. The observed lag between the rainfall and streamflow annual cycles is represented by the dotted horizontal line.

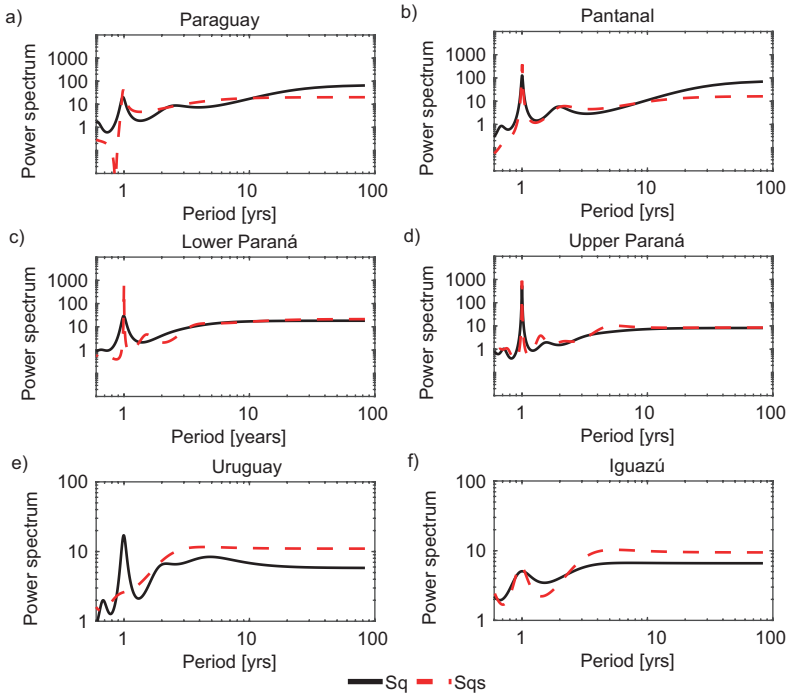

Fig. 8. Observed (*Sq*) and simulated (*Sqs*) streamflow MEM power spectrums as a function of its period (years) at the corresponding closing points for each sub-basin, respectively: (a) Paraguay, (b) Pantanal, (c) lower Paraná, (d) upper Paraná, (e) Uruguay, (f) Iguazú. Both horizontal and vertical axes are in logarithmic scale.

ARMA processes). To carry out this work, the LPB was divided in six sub-basins (Paraguay, Pantanal, Uruguay, Iguazú, lower and upper Paraná), with their corresponding closing points (Pilcomayo, Ladario, Salto Grande, Salto Caxias, Paraná Timbúes and Itaipú), respectively.

A statistical description of both precipitation and streamflow variability for the period 1931-2010 was addressed focusing on the annual cycles and the interdecadal and multidecadal timescales. Different streamflow annual cycle regimes were identified among the six sub-basins as distinct mechanisms affect tropical (Paraná, Paraguay, Pantanal) and subtropical sub-basins (Uruguay and Iguazú). The tropical basins are mainly affected by the SACZ, while subtropical basins are influenced by the SALLJ. Interdecadal variations on the annual cycles were observed in both streamflow and precipitation time series. For the latter, similar annual cycle regimes were observed, with small changes in the rainfall magnitude during certain months, while in the discharges, changes in magnitude throughout the 12 months (and in some cases also in the time of occurrence of the peak) were evident. An enhancement in the monthly discharges was also observed particularly for the recent decades of the study period across the different sub-basins. The observed decadal differences may be partly explained by the higher occurrence of El Niño events in the recent decades associated to the positive phase of the PDO (Deser et al., 2004) between 1977 and 1999. The differences in the strength of ENSO teleconnections for rainfall are related to the PDO, in the sense that these teleconnections act constructively when ENSO and PDO are in the same phase (Kayano and Andreoli, 2007). Other teleconnection patterns also affect the LPB precipitation and streamflow variability in different time scales as reviewed by Zanin and Satyamurty (2020a, b). For example, in addition to the positive PDO phase during 1977-1999, the increase in precipitation was also influenced by

the AMO negative phase between 1970 and 1994. Consistently, positive precipitation and streamflow trends were found in all the sub-basins in the LPB considering the study period 1931-2010, in line with previous studies as discussed in section 3. In addition, a higher percentage of increase in streamflow over the recent decades was observed compared to precipitation, and this was quantified by comparing the annual means of each magnitude, corresponding to the sub-periods 1931-1970 and 1971-2010.

One of the core assumptions considered in the present study was that in the LPB the different sub-basins act on precipitation as spatiotemporal integrators and, therefore, operate as low frequency filters, like moving averages. Therefore, the application of the MEM spectral analysis is proposed to explore extensively the observed amplification of medium and low frequency variability in streamflow compared to precipitation. Results show that streamflow variability displays attenuation with respect to precipitation for the higher frequencies and amplification for longer ones. The interdecadal/multidecadal low frequency variability is predominant for most of the streamflow time series. This feature was consistent with our initial assumption.

In order to understand to what extent rainfall variability can explain river discharges variability, the simulated streamflow power spectrum analysis was addressed assuming that the underlying process was an ARMA instead of an AR, using only areal-averaged precipitation timeseries at the sub-basin scale. Under this statistical framework, results confirm that the application of an ARMA(p,q) filter is able to effectively fit the observed rainfall-streamflow physical relationship, since the observed and estimated temporal variability of the streamflows were similar, particularly at the longer periods, around interdecadal/multidecadal component. Moreover, by applying the described methodology similar lags between rainfall and streamflow annual cycles, consistent with those observed at each sub-basin, were obtained. In this sense, the interpretation of an ARMA statistical process applied to precipitation time series at the sub-basin scale can provide a first insight about the precipitation and streamflow temporal variability which may be useful for further inferences on river discharges future variations. This may be a complementary tool for basins where data are scarce. Finally, the analysis undertaken in this study also suggests a possible pathway to advance in the knowledge of the relationship between the variability of streamflow and precipitation in different catchments of the world.

Acknowledgments

We gratefully acknowledge Alexis Hannart for his valuable contribution to the present study. This work was supported by the Agencia Nacional de Promoción Científica y Tecnológica-Argentina Innovadora 2020 (PICT2016-1422), and the University of Buenos Aires (UBACYT- 20020170100666BA).

References

Akaike H. 1974. New look at the statistical model identification. IEEE Transactions on Automatic Control 19: 716-723. https://doi.org/10.1109/TAC.1974.1100705

Antico A, Schlotthauer G, Torres ME. 2014. Analysis of hydroclimatic variability and trends using a novel empirical mode decomposition: Application to the Paraná River Basin. Journal of Geophysical Research-Atmospheres 119: 1218-1233. https://doi.org/10.1002/2013JD020420

Barros VR, Chamorro L, Coronel G, Báez J. 2004. The major discharge events in the Paraguay River: magnitudes, source regions, and climate forcings. Journal of Hydrometeorology 5: 1161-1170. https://doi.org/10.1175/JHM-378.1

Barros V, Clarke R, Silva Dias P. 2006. El cambio climático en la cuenca del Plata (Climate change in La Plata basin). Buenos Aires, CIMA-CONICET, 232 pp.

Barros VR, Doyle ME, Camilloni IA. 2008. Precipitation trends in southeastern South America: Relationship with ENSO phases and with low-level circulation. Theoretical and Applied Climatology 93: 19-33. https://doi.org/10.1007/s00704-007-0329-x

Berbery EH, Barros VR. 2002. The hydrologic cycle of the La Plata basin in South America. Journal of Hydrometeorology 3: 630-645. https://doi.org/10.1175/1525-7541(2002)003%3C0630:THCOTL%3E2.0.CO;2

Berri GJ, Ghietto MA, García NO. 2002. The influence of ENSO in the flows of the Upper Paraná River of South America over the past 100 years. Journal of Hydrometeorology 3: 57-65. https://doi.org/10.1175/1525-7541(2002)003%3C0057:TIOEIT%3E2.0.CO;2

Bischoff SA, García NO, Vargas WM, Jones PD, Conway D. 2000. Climatic variability and Uruguay River flows. Water International 25: 446-456. https://doi.org/10.1080/02508060008686852

Boulanger J-P, Leloup J, Penalba O, Rusticucci M, Lafon F, Vargas W. 2005. Observed precipitation in the Paraná-Plata hydrological basin: Long-term trends, extreme conditions and ENSO teleconnections. Climate Dynamics 24: 393-413. https://doi.org/10.1007/s00382-004-0514-x

Boulanger J-P, Carril AF, Sánchez E. 2016. CLARIS-La Plata Basin: Regional hydroclimate variability, uncertainties and climate change scenarios. Climate Research 68: 93-94. https://doi.org/10.3354/cr01392

Box GEP, Jenkins GM. 1994. Time Series Analysis: Forecasting and control. Prentice-Hall, 598 pp.

Camilloni I, Barros V. 2000. The Paraná river response to El Niño 1982-83 and 1997-98 events. Journal of Hydrometeorology 1: 412-430. https://doi.org/10.1175/1525-7541(2000)001%3C0412:TPRRTE%3E2.0.CO;2

Camilloni I, Barros V. 2003. Extreme discharge events in the Paraná River and their climate forcing. Journal of Hydrology 278: 94-106. https://doi.org/10.1016/S0022-1694(03)00133-1

Camilloni I. 2005. Extreme flood events in the Uruguay River of South America. VAMOS Newsletter 2: 23-25.

Cerón WL, Kayano MT, Andreoli RV, Ávila-Díaz A, Ayes I, Freitas ED, Martins JA, Souza RAF. 2021. Recent intensification of extreme precipitation events in the La Plata Basin in Southern South America (1981-2018). Atmospheric Research 249: 105299. https://doi.org/10.1016/j.atmosres.2020.105299

Chiew FHS, Whetton PH, McMahon TA, Pittock AB. 1995. Simulation of the impacts of climate change on runoff and soil moisture in Australian catchments. Journal of Hydrology 167: 121-147. https://doi.org/10.1016/0022-1694(94)02649-V

Chiew FHS. 2006. Estimation of rainfall elasticity of streamflow in Australia. Hydrological Sciences Journal 51: 613-625. https://doi.org/10.1623/hysj.51.4.613

Collischonn W, Tucci CEM., Clarke RT. 2001. Further evidence of changes in the hydrological regime of the River Paraguay: Part of a wider phenomenon of climate change? Journal of Hydrology 245: 218-238.

CIC. 2016. Hidroelectricidad y navegación en la Cuenca del Plata. Comité Intergubernamental Coordinador de los Países de la Cuenca del Plata. Editorial Palermo, Buenos Aires, Argentina.

CIC. 2017. Transboundary diagnostic analysis (TDA) for the La Plata River basin. 1st special ed. Comité Intergubernamental Coordinador de los Países de la Cuenca del Plata/Organización de Estados Americanos. Buenos Aires, Argentina.

Coronel G, Menéndez A, Chamorro L. 2006. Fisiografía e hidrología de la cuenca del Plata. In: El cambio climático en la cuenca del Plata (Barros V, Clarke R, Silva Dias P, Eds.). CIMA-CONICET, Buenos Aires, Argentina, 81-92.

Dai A, Qian T, Trenberth KE, Milliman HD. 2009. Changes in continental freshwater discharge from 1948 to 2004. Journal of Climate 22: 2773-2792. https://doi.org/10.1175/2008JCLI2592.1

Deser C, Phillips AS, Hurrell JW. 2004. Pacific interdecadal climate variability: Linkages between the tropics and the North Pacific during boreal winter since 1900. Journal of Climate 17: 3109-3124. https://doi.org/10.1175/1520-0442(2004)017%3C3109:PICVLB%3E2.0.CO;2

Depetris PJ, Kempe S, Latif M, Mook WG. 1996. ENSO-controlled flooding in the Paraná River (1904-1991). Naturwissenschaften 83: 127-129. https://doi.org/10.1007/BF01142177

Doyle ME, Barros VR. 2011. Attribution of the river flow growth in the Plata Basin. International Journal of Climatology 31: 2234-2248. https://doi.org/10.1002/joc.2228

Doyle ME, Saurral I, Barros VR. 2012. Trends in the distributions of aggregated monthly precipitation over the La Plata Basin. International Journal of Climatology 32: 2149-2162. https://doi.org/10.1002/joc.2429

Fu G, Charles SP, Chiew FHS 2007. A two-parameter climate elasticity of streamflow index to assess climate change effects on annual streamflow. Water Resources Research 43: W11419. https://doi.org/10.1029/2007WR005890

Gan MA, Rao VB. 1991. Surface cyclogenesis over South America. Monthly Weather Review 119: 1293-1302. https://doi.org/10.1175/1520-0493(1991)119<1293:SCOSA>2.0.CO;2

García NO, Mechoso CR. 2005 Variability in the discharge of South American rivers and in climate. Hydrological Sciences Journal 50: 459-478, https://doi.org/10.1623/hysj.50.3.459.65030

García NO, Vargas WM. 1996. The spatial variability of runoff and precipitation in the Río de la Plata basin. Hydrological Sciences Journal 41: 279-299. https://doi.org/10.1080/02626669609491503

García NO, Vargas WM. 1998. The temporal climatic variability in the Río de la Plata basin displayed by the river discharges. Climatic Change 38: 359-379. https://doi.org/10.1023/A:1005386530866

Genta JL, Perez-Iribarren G, Mechoso CR. 1998. A recent increasing trend in the streamflow of rivers in southeastern South America. Journal of Climate 11): 2858-2862. https://doi.org/10.1175/1520-0442(1998)011<2858:ARITIT>2.0.CO;2

Ghil M, Allen MR, Dettinger MD, Ide K, Kondrashov D, Mann ME, Robertson AW, Saunders A, Tian Y, Varadi F, Yiou P. 2002. Advanced spectral methods for climatic time series. Reviews of Geophysics 40: 3-1-3-41. https://doi.org/10.1029/2000RG000092

Gulizia C, Camilloni I. 2015. A spatio-temporal comparative study of the representation of precipitation over South America derived by three gridded data sets. International Journal of Climatolology 36: 1549-1559. https://doi.org/10.1002/joc.4416

Haylock MR, Peterson TC, Alves LM, Ambrizzi T, Anunciação YMT, Baez J, Barros VR, Berlato MA, Bidegain M, Coronel G, Corradi V, Garcia VJ, Grimm AM, Karoly D, Marengo JA, Marino MB, Moncunill DF, Nechet D, Quintana J, Rebello E, Rusticucci M, Santos JL, Trebejo I, Vincent LA. 2006. Trends in total and extreme South American rainfall in 1960-2000 and links with sea surface temperature. Journal of Climate 19: 1490-1512. https://doi.org/10.1175/JCLI3695.1

Huntington TG, Billmire M. 2014. Trends in precipitation, runoff, and evapotranspiration for rivers draining to the Gulf of Maine in the United States. Journal of Hydrometeorology 15: 726-743. https://doi.org/10.1175/JHM-D-13-018.1

Jones C, Carvalho LM. 2018. The influence of the Atlantic Multidecadal Oscillation on the eastern Andes low-level jet and precipitation in South America. Nature Climate and Atmospheric Science 40: 1-7. https://doi.org/10.1038/s41612-018-0050-8

Kayano M, Andreoli R. 2007. Relations of South American summer rainfall interannual variations with the Pacific Decadal Oscillation. International Journal of Climatology 27: 531-540. https://doi.org/10.1002/joc.1417

Krepper CM, García NO. 2004. Spatial and temporal structures of trends and interannual variability of precipitation over the La Plata Basin. Quaternary International 114: 11-21. https://doi.org/10.1016/S1040-6182(03)00038-7

Krepper CM, García NO, Jones PD. 2003. Interannual variability in the Uruguay river basin. International Journal of Climatology 23: 103-115. https://doi.org/10.1002/joc.853

Krepper CM, García NO, Jones PD. 2006. Paraguay river basin response to seasonal rainfall. International Journal of Climatology 26: 1267-1278. https://doi.org/10.1002/joc.1313

Krepper CM, García NO, Jones PD. 2008. Low-frequency response of the upper Paraná basin. International Journal of Climatology 28: 351-360. https://doi.org/10.1002/joc.1535

Labat D, Ronchail J, Guyot JL. 2005. Recent advances in wavelet analyses: Part 2—Amazon, Paraná, Orinoco and Congo discharges time scale variability. Journal of Hydrology 314: 289311. https://doi.org/10.1016/j.jhydrol.2005.04.004

Leite SM, JP Peixoto. 1995. Maximum entropy spectral analysis of the Duero basin. International Journal of Climatology 15: 463-472. https://doi.org/10.1002/joc.3370150408

Li Z, Cheng T, Song H, Li Z, Yu J. 2015. Analysis of precipitation cycles based on MEM in the Yellow River basin. Proceedings of the International Association of Hydrological Sciences 368: 114-119. https://doi.org/10.5194/piahs-368-114-2015

Li S, Wu L, Yang Y, Geng T, Cai W, Gan B, Chen Z, Jing Z, Wang G, Ma X. 2020. The Pacific Decadal Oscillation less predictable under greenhouse warming. Nature Climate Change 10: 30-34. https://doi.org/10.1038/s41558-019-0663-x

Lovino MA, Müller OV, Müller GV, Sgroi LC, Baethgen WE. 2018. Interannual-to-multidecadal hydroclimate variability and its sectoral impacts in northeastern Argentina. Hydrology and Earth System Science 22: 3155-3174. https://doi.org/10.5194/hess-22-3155-2018

Maciel F, Díaz A, Terra R. 2013. Multi-annual variability of streamflow in La Plata Basin. Part I: observations and links to global climate. International Journal of River Basin Management 11: 345-360. https://doi.org/10.1080/15715124.2013.847843

Mantua NJ, Hare SR, Zhang Y, Wallace JM, Francis RC. 1997. A Pacific interdecadal climate oscillation with impacts on salmon production. Bulletin of the American Meteorological Society 78: 1069-1079. https://doi.org/10.1175/1520-0477(1997)078%3C1069:APICOW%3E2.0.CO;2

Mechoso CR, Perez-Iribarren G. 1992. Streamflow in southeastern South America and the Southern Oscillation. Journal of Climate 5: 1535-1539. https://doi.org/10.1175/1520-0442(1992)005<1535:SISSAA>2.0.CO;2

Milliman JD, Farnsworth KL, Jones PD, Xu KH, Smith LC. 2008. Climatic and anthropogenic factors affecting river discharge to the global ocean, 1951-2000. Global Planet Change 62: 187-194. https://doi.org/10.1016/j.gloplacha.2008.03.001

Nery JT, Denig EA, Azevedo LC, Martins MLOF. 2006. Estudo da variabilidade inter-anual da precipitaçao na bacia do Rio Uruguai. Revista Brasileira de Recursos Hídricos 11: 149-159. https://doi.org/10.21168/rbrh.v11n2.p149-159

Newman M, Alexander MA, Ault TR, Coob KM, Deser C, Lorenzo E, Mantua NJ, Miller AJ, Minobe S, Nakamura H, Sheneider N, Vimont DJ, Phillips AS, Scott JD, Smith C. 2016. The Pacific Decadal Oscillation, revisited. Journal of Climate 29: 4399-4427. https://doi.org/10.1175/JCLI-D-15-0508.1

Pasquini A, Depetris P. 2007. Discharge trends and low dynamics of South American rivers draining the southern Atlantic seaboard: An overview. Journal of Hydrology 333: 385-399. https://doi.org/10.1016/j.jhydrol.2006.09.005

Penalba OC, Robledo FA. 2010. Spatial and temporal variability of the frequency of extreme daily rainfall regime in the La Plata Basin during the 20th century. Climate Change 98: 531-550. https://doi.org/10.1007/s10584-009-9744-6

Rao VB, Cavalcanti IF, Hada K. 1996. Annual variation of rainfall over Brazil and water vapor characteristics over South America. Journal of Geophysical Research 101: 26539-26551. https://doi.org/10.1029/96JD01936

Robertson AW, Mechoso CR. 1998. Interannual and decadal cycles in river flows of southeastern South America. Journal of Climate 11: 2570-2581. https://doi.org/10.1175/1520-0442(1998)011%3C2570:IADCIR%3E2.0.CO;2

Robertson AW, Mechoso CR. 2000. Interannual and interdecadal variability of the South Atlantic Convergence Zone. Journal of Climate 11: 2947-2957. https://doi.org/10.1175/1520-0493(2000)128<2947:IAIVOT>2.0.CO;2

Ropelewski CF, Halpert MS. 1987. Global and regional scale precipitation associated with El Niño/Southern Oscillation. Monthly Weather Review 115: 1606-1626. https://doi.org/10.1175/1520-0493(1987)115<1606:GARSPP>2.0.CO;2

Ropelewski CF, Halpert MS. 1989. Precipitation patterns associated with the high index phase of the Southern Oscillation. Journal of Climate 2: 268-284. https://doi.org/10.1175/1520-0442(1989)002<0268:PPAWTH>2.0.CO;2

Salio P, Nicolini M, Saulo AC. 2002. Chaco low-level jet events characterization during the austral summer season. Journal of Geophysical Research 107: ACL 32-1-ACL 32-17. https://doi.org/10.1029/2001JD001315

Sankarasubramanian A, Vogel RM, Limburner JF. 2001. Climate elasticity of streamflow in the United States. Water Resources Research 37: 1771-1781. https://doi.org/10.1029/2000WR900330

Saurral R. 2010. The hydrologic cycle of the La Plata Basin in the WCRP-CMIP3 multimodel dataset. Journal of Hydrometeorology 11: 1083-1102. https://doi.org/10.1175/2010JHM1178.1

Saurral R, Montroull, N, Camilloni I. 2013. Development of statistically unbiased 21st century hydrology scenarios over La Plata Basin. International Journal of River Basin Management 11: 329-343. https://doi.org/10.1080/15715124.2014.885440

Saurral RI, Camilloni IA, Barros VR. 2017. Low-frequency variability and trends in centennial precipitation stations in southern South America. International Journal of Climatology 37: 1774-1793. https://doi.org/10.1002/joc.4810

Schaake JC. 1990. From climate to flow. In: Climate Change and US water resources (Waggoner PE, Ed.). John Wiley, New York, 177-206.

Schaake JC, Koren VI, Duan QY, Mitchell K, Chen F. 1996. Simple water balance model for estimating runoff at different spatial and temporal scales. Journal of Geophysical Research 101: 7461-7475. https://doi.org/10.1029/95JD02892

Schlesinger ME, Ramankutty N. 1994. An oscillation in the global climate system of period 65 -70 years. Nature 367, 723-726. https://doi.org/10.1038/367723a0

Schneider U, Becker A, Finger P, Meyer-Christoffer A, Rudolf B, Ziese M. 2015. GPCC full data reanalysis version 7.0 at 0.5: Monthly land-surface precipitation from rain-gauges built on GTS based and historic data. Federal Ministry for Digital and Transport, Germany. https://doi.org/10.5676/DWD_GPCC/FD_M_V7_050

Shannon CE. 1949. Communication in the presence of noise. Proceedings of the IRE 37: 10-21. https://doi.org/10.1109/JRPROC.1949.232969

Shi X, Qin T, Nie H, Weng B, He S. 2019. Changes in major global river discharges directed into the ocean. International Journal of Environmental Research and Public Health 16: 1469. https://doi.org/10.3390/ijerph16081469

Su F., Lettenmaier DP. 2009. Estimation of the surface water budget of the La Plata Basin. Journal of Hydrometeorology 10: 981-998. https://doi.org/10.1175/2009JHM1100.1

Trenberth KE, Shea DJ. 2006. Atlantic hurricanes and natural variability in 2005. Geophysical Research Letters 33: L12704. https://doi.org/10.1029/2006GL026894

Tucci CEM, Clarke RT. 1998. Environmental issues in the La Plata Basin. International Journal of Water Resources D 14: 157-173. https://doi.org/10.1080/07900629849376

Ulrych TJ, Bishop TN. 1975. Maximum entropy spectral analysis and autoregressive decomposition. Reviews of Geophysics and Space Physics 13: 183-200. https://doi.org/10.1029/RG013i001p00183

Walter MT, Wilks DS, Parlange J-Y, Schneider RL. 2004. Increasing evapotranspiration from the conterminous United States. Journal of Hydrometeorology 5: 405-408. https://doi.org/10.1175/1525-7541(2004)005%3C0405:IEFTCU%3E2.0.CO;2

Wilks DS. 2006a. Ch. 8. Time series. In: Statistical methods in the atmospheric sciences (Wilks DS, Ed.). Elsevier, Academic Press, 357-358.

Wilks DS. 2006b. Ch. 8. Time series. In: Statistical methods in the atmospheric sciences, (Wilks DS, Ed.). Elsevier, Academic Press, 362-363.

Wu N. 2012. The maximum entropy method. Springer-Verlag, Heidelberg, Germany, 327 pp. (Springer Series in Information Sciences 32).

Zandonadi L, Acquaotta F, Fratianni S, Zavattini JA. 2016. Changes in precipitation extremes in Brazil (Paraná River Basin). Theoretical and Applied Climatology 123: 741-756. https://doi.org/10.1007/s00704-015-1391-4

Zanin Pr, Satyamurty P. 2020a. Hydrological processes interconnecting the two largest watersheds of South America from seasonal to intra-monthly time scales: A critical review. International Journal of Climatology 40: 3971-4005. https://doi.org/10.1002/joc.6443

Zanin PR, Satyamurty P. 2020b. Hydrological processes interconnecting the two largest watersheds of South America from multi-decadal to inter-annual time scales: A critical review. International Journal of Climatology 40: 4006–4038. https://doi.org/10.1002/joc.6442

Zhang L, Delworth T. 2016. Simulated response of the Pacific Decadal Oscillation to climate change. Journal of Climate 29: 5999-6018. https://doi.org/10.1175/JCLI-D-15-0690.1

SUPPLEMENTARY MATERIAL

Table SI. Months of peak precipitation and streamflows in each sub-basin, and the lag between them, considering the annual cycle for the period 1931-2010.

Sub-basin (gauging station)	Month of peak		Lag (months)
	Precipitation	Streamflow	
Paraguay (Pilcomayo)	January	June	5
Pantanal (Ladario)	January	June	5
Lower Paraná (Paraná Timbúes)	January	March	2
Upper Paraná (Itaipú)	January	February	1
Uruguay (Salto Grande)	October	October	0
Iguazú (Salto Caxias)	October	October	0

Table SII. Precipitation and streamflow trends (m^3s^{-1} $decade^{-1}$) calculated from the annual time series in the period from 1931 to 2010. Significant trends at 90% (*), 95% (**) and 99% (***) significant levels are also indicated.

Sub-basin (gauging station)	Trends (m^3s^{-1}/decade)	
	Precipitation	Streamflow
Paraguay (Pilcomayo)	+1.58	+1.76**
Pantanal (Ladario)	+0.27	+0.68 *
Lower Paraná (Paraná Timbúes)	+9.38*	+8.82***
Upper Paraná (Itaipú)	+3.17	+7.87***
Uruguay (Salto Grande)	+3.13**	+3.00**
Iguazú (Salto Caxias)	+0.26	+1.00***

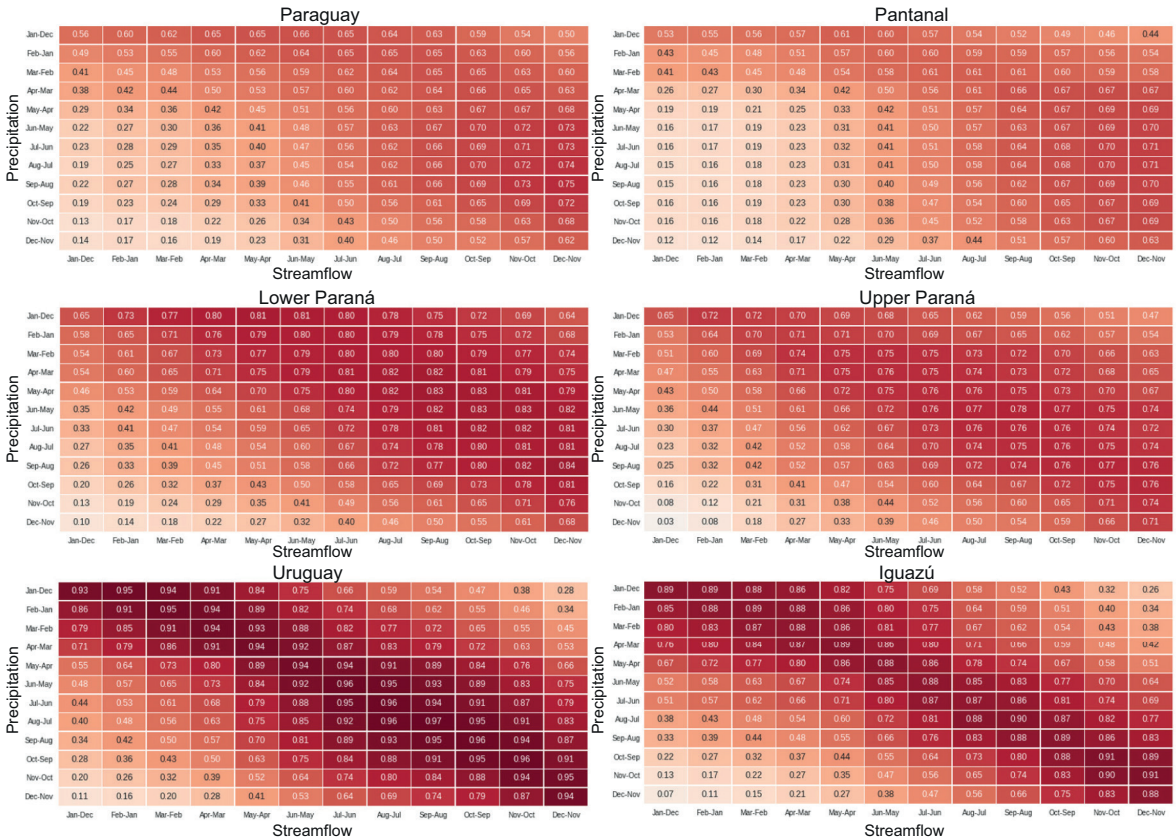

Fig. S1. Heatmaps of the precipitation-streamflow correlations at different lags.

Atmósfera 36(2), 207-224 (2023)
https://doi.org/10.20937/ATM.53093

CONTENTS

Cyclonic circulation and climatology of SST, CHL and wind stress curl in a semi-enclosed bay (Bahía de La Paz, Gulf of California, Mexico)

Emilio BEIER[1], Rubén CASTRO[2]* and Víctor M. GODÍNEZ[3]

[1] *Laboratorio de Macroecología Marina, Centro de Investigación Científica y de Educación Superior de Ensenada, Unidad La Paz, Miraflores 334, 23000 La Paz, Baja California Sur, México.*
[2] *Universidad Autónoma de Baja California (UABC), Facultad de Ciencias Marinas, carretera Tijuana-Ensenada 3917, Fraccionamiento Playitas, 22860 Ensenada, Baja California, México.*
[3] *Departamento de Oceanografía Física, Centro de Investigación Científica y de Educación Superior de Ensenada, Baja California (CICESE), carretera Ensenada-Tijuana 3918, Fraccionamiento Playitas, 22860 Ensenada, Baja California, México.*
*Corresponding author; email: rcastro@uabc.edu.mx

Received: June 4, 2021; accepted: September 24, 2021

RESUMEN

Las primeras observaciones directas de corrientes (con el Perfilador de Corriente Acústico Doppler y derivadores superficiales) en la Bahía de La Paz, situada al suroeste del Golfo de California (GC), coinciden con estudios anteriores en que la principal característica dinámica durante el verano es una circulación ciclónica cerrada. Sin embargo, encontramos que los cálculos goestróficos sobrestiman a las velocidades de rotación: las velocidades observadas (0.20-0.25 m s^{-1}) fueron ~25-40% más bajas que las estimadas a partir del balance geostrófico (0.25-0.35 m s^{-1}). La razón es que en este caso no se puede despreciar la fuerza centrífuga. El período medio de rotación durante las observaciones a bordo del buque en agosto de 2004 fue de ~1.4 días, pero varió durante el tiempo en que los derivadores estuvieron en la bahía, de ~1-2 días en junio-julio a 2.5-3 días en septiembre-octubre. Los análisis satelitales (velocidad del viento, temperatura superficial del mar y clorofila) muestran que, de mayo a septiembre, el rotacional del esfuerzo del viento fue intenso y ciclónico, y la superficie de la bahía fue más fría y más rica que las aguas adyacentes del golfo, lo cual puede atribuirse al rotacional del esfuerzo del viento positivo. Este rotacional del esfuerzo del viento positivo sobre la bahía, es parte de una distribución del rotacional del esfuerzo del viento positivo de mayor escala que rodea la parte sur de la península de Baja California durante el verano, probablemente aumentado en la bahía por las características topográficas locales. Aunque existe un intercambio de agua entre la bahía y el GC, su efecto sobre la dinámica es poco conocido.

ABSTRACT

The first direct current observations (with the Lowering Acoustic Doppler Current Profiler [LADCP] and surface drifters) in Bahía de La Paz, in the southwestern Gulf of California (GC), concur with previous reports that the main dynamical feature during summer is a closed cyclonic circulation. However, we found that geostrophic calculations overestimate the speed of the orbital velocity: actual speeds (0.20-0.25 m s^{-1}) were ~25-40% lower than those estimated from geostrophic balance (0.25-0.35 m s^{-1}). The reason is that the centrifugal force cannot be neglected in this case. The mean rotation period during ship-borne observations in August 2004 was ~1.4 days, but it varied in the time that surface drifters were inside the bay, from ~1-2 days in June-July to ~2.5-3 days in September-October. The analysis of satellite data (wind velocity, sea surface temperature and chlorophyll) shows that from May to September the wind stress curl is strong and cyclonic, and the surface of the bay is cooler and richer than the adjacent Gulf of California waters, which could be attributed to the positive wind stress curl. This positive wind stress curl on the bay is part of a larger-scale

positive wind stress curl distribution that surrounds the southern part of the Baja California Peninsula during summer, probably enhanced in the bay by local topography features. Although there is an exchange of water between the bay and the GC, its effect on the dynamics is poorly known.

Keywords: cyclonic circulation, current observations (LADCP, drifters buoys), wind stress curl, Bahía de La Paz.

1. Introduction

Bahía de La Paz is a semi-enclosed body of water located on the western side of the Gulf of California (GC), near the southern end of the Baja California Peninsula (Fig. 1). It is ~80 km long by ~35 km wide, with a surface area ~2700 km^2. The main bathymetric

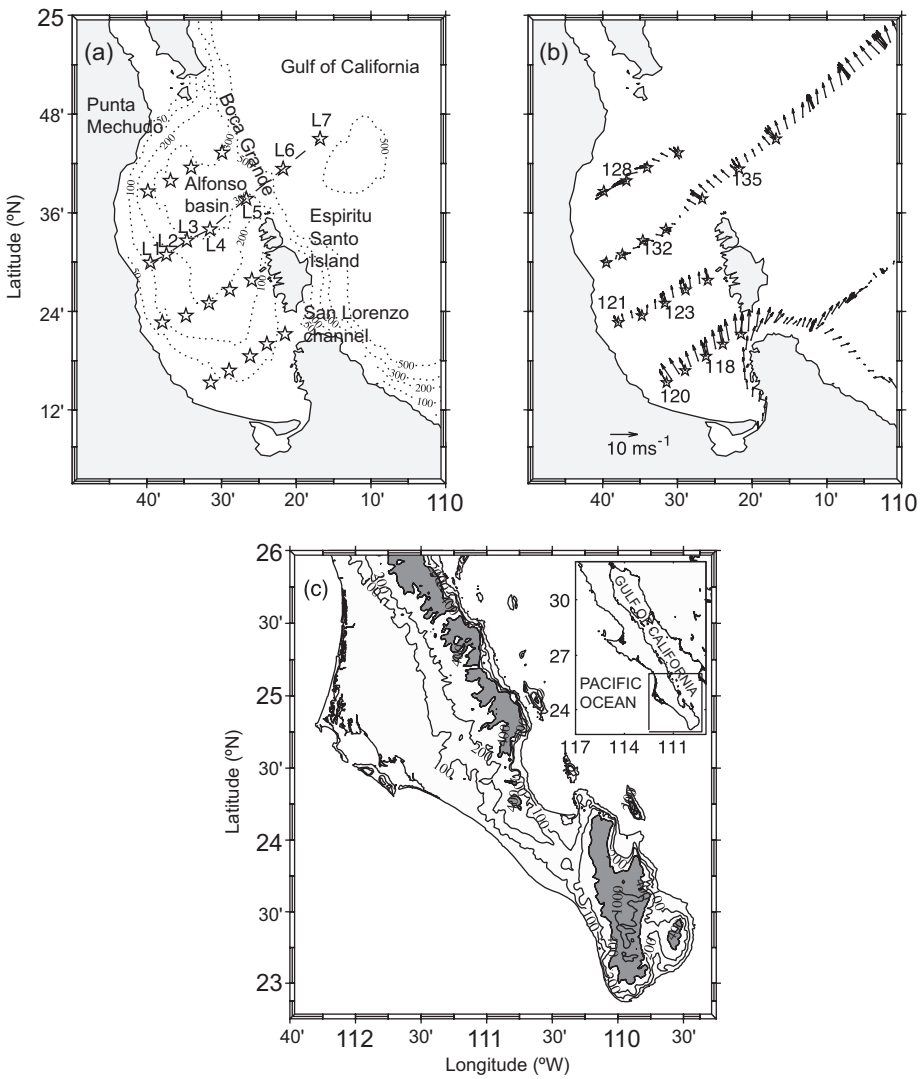

Fig. 1. (a) Bahía de la Paz, with bathymetry (m). Stars indicate the positions of the oceanographic stations made during the NAME-II cruise. Line of stations L1-L7 is referred to in the text. (b) Wind velocity from ship data during cruise. (c) Map of the southern portion of the Baja California Peninsula with topography (m). The insert square shows the location of Bahía de la Paz.

feature is in the northern part: the 450-m deep Alfonso basin, with a sill at 250-300 m at the main connection with the GC, between Punta Mechudo and the north tip of Espíritu Santo Island (Fig. 1). In the south of the bay, the narrow and shallow (< 18 m) San Lorenzo channel probably plays a minor role in the low-frequency interchange of waters between the GC and the bay.

Monreal-Gómez et al. (2001) found that during June 1998 the geostrophic circulation in the northern Bahía de La Paz was cyclonic, and suggested that it could be related to the wind curl. Salinas-González et al. (2003) found evidence of anticyclonic geostrophic circulation in October 1997 and August 1999. Sánchez-Velasco et al. (2006) found cyclonic circulation in July and October 2001, and weak undefined patterns of circulation in May 2001 and February 2002. The latter authors argued, based on fish larvae distributions, that the circulation inside the bay may be connected to the overall seasonal circulation of the GC. During August, 2009, the presence of a cyclonic eddy confined in the northern of Bahía de La Paz was also documented, implying their importance in the distribution of nutrients concentrations and different trophic groups of zooplankton in the region (Coria-Monter et al., 2014; Durán-Campos et al., 2015). The latter authors suggested that the eddy origin could be linked with the topography of the basin and the interaction of the current entering to the bay through Boca Grande (Fig. 1).

These works suggest that cyclonic circulation is a frequent summer feature of the bay's geostrophic circulation, but that the anticyclonic pattern is also possible. However, geostrophic calculations can give only an approximation of the circulation; previous works have taken the bottom, 50 m, or base as level of no motion (to include as many stations as possible), while the distribution of thermohaline variables in vertical cross-sections of the summer hydrography suggest that the cyclonic circulation is at least 150 m deep (Durán-Campos et al., 2015).

The wind field over the GC changes direction with the seasons, blowing from the northwest during winter and from the southeast during summer (see, e.g., review by Lavín and Marinone, 2003; Marinone et al., 2004; Bordoni et al., 2004). The wind over the Bahía de La Paz bay shows the same pattern in winter and fall, however, during late spring and summer winds from southwestern component predominate (Herrera-Cervantes et al., 2017; Muñoz-Barbosa et al., 2020). I has been reported that Bahía de La Paz bay has a strong coupling between wind stress curl, sea surface temperature, and chlorophyll-a (Herrera-Cervantes, 2019).

In this work we present the first directly-observed description of the summer circulation in Bahía de La Paz, based on observations made during the summer of 2004, using satellite-tracked surface drifters and vertical Lowering Acoustic Doppler Current Profiler (LADCP) profiles. The consequences of the bay's circulation on its surface temperature and chlorophyll are described by analysis of satellite data. The role of the wind stress curl on SST and productivity (CHL) of the bay is investigated using a time series of satellite-derived daily wind velocity.

2. Data and methods

Lagrangian surface currents were observed with two Pacific-Gyre SVP ARGOS drifters equipped with holey socks centered at 15 m (see Lavín et al., 2014). The first drifter (s/n 50021) was inside the bay from June 18 to July 18, 2004, and the second (s/n 52083) from August 16 to October 11, 2004. The drifters transmitted their positions in average every 3 h, with accuracy ~300 m, which resulted in an observed velocity error of ~0.028 m s^{-1}. We performed a linear interpolation to obtain positions exactly every 3 h, from which velocities were computed.

One 17-day campaign (NAME-II) was made from the research vessel Francisco de Ulloa, from August 6 to August 22, 2004, covering most of the southern GC (Lavín et al., 2013), but for this paper we use only data collected between August 15 and August 16 inside Bahía de La Paz (Fig. 1). The spacing between stations was ~10 km in the across-bay transects, and the distance between transects was ~25 km. The thermohaline profiles to ~3 m above the bottom were measured with a factory-calibrated conductivity, temperature, and depth (CTD) device (SeaBird SBE-911 plus), with primary and secondary sensors and a 24 Hz sampling rate. The CTD was equipped with fluorescence and dissolved oxygen (DO) sensors. The data were processed and averaged to 1 dbar as documented by Castro et al. (2006). Salinity was calculated with the Practical Salinity Scale (1978).

The potential temperature, θ (ºC), and the potential density anomaly, γ_θ (kg m^{-3}), were calculated according to UNESCO (1991). Prior to geostrophic velocity calculation, the temperature and salinity cross-sections were objectively mapped in order to remove internal waves and other small-scale variability. A standard objective mapping interpolation was used, with a standard Gaussian correlation function with relative errors of 0.1, a 20 km horizontal length scale and a 30 m vertical scale.

The velocity profile was measured with a broadband RDI 300 KHz lowering acoustic doppler current profiler (LADCP) attached to the CTD protection frame. The absolute velocity profiles, in 8-m bins, were obtained with the methods described by Visbeck (2002).

Weekly average satellite images (4 × 4 km) of sea surface temperature (SST) and chlorophyll concentration (CHL) from the MODIS satellites were obtained for the period July 2002 to July 2007 (NASA, 2018). Daily wind data for the period 2002 to 2007 were obtained from the Cross-Calibrated Multi-Platform Ocean Surface Wind Velocity Product for Meteorological and Oceanographic Applications (CCMP), which contains gridded variational analysis method ocean wind vector fields, produced from all available microwave radiometer data, blended with scatterometer data (NSCAT and SeaWinds on QuikSCAT/ADEOS-II) (Atlas et al., 2011). The horizontal resolution of the CCMP winds is 0.25º × 0.25º each 6 h.

To calculate the wind stress (N m^{-2}) we used the bulk aerodynamic formulae given by Trenberth et al. (1989, and references therein):

$$\vec{\tau} = \rho C_D |\vec{W}| \vec{W} \quad (1)$$

where ρ is a constant air density of 1.2 kg m^{-3}, $\vec{W}$ is the wind velocity vector, $|\vec{W}|$ is the wind speed and C_D is the drag coefficient, given by

$$10^3 C_d = 0.49 + 0.065 \left|\vec{W}\right| \; for \quad \left|\vec{W}\right| > 10 \; m\, s^{-1}$$

$$10^3 C_d = 1.14 \; for \quad 3 \le \left|\vec{W}\right| \le 10 \; m\, s^{-1}$$

$$10^3 C_d = 0.62 + 1.56 \left|\vec{W}\right|^{-1} for \left|\vec{W}\right| < 3 \; m\, s^{-1}$$

The wind stress curl (N m^{-3}) was computed by

$$WSC = \left(\frac{\partial \tau_y}{\partial x} - \frac{\partial \tau_x}{\partial y}\right) \quad (2)$$

The seasonal cycles of SST, chlorophyll, wind and wind stress curl were obtained by fitting the time series to mean plus annual and semiannual harmonics:

$$F(t) = A_m + A_a \cos(\omega t - \varphi_a) + A_s \cos(2\omega t - \varphi_s) + R_e(v) \quad (3)$$

where A_m is the temporal mean, A_a and A_s are the annual and semiannual amplitudes, 2π/365 is the annual radian frequency, t is the time, φ_a and φ_s are the phases of the annual and semiannual harmonics, and R_e are the residues containing all the non-seasonal anomalies. The fitting error of amplitudes and phases was calculated as in Beron-Vera and Ripa (2002).

3. Results

3.1 Circulation and hydrography

The drifter velocities observed during 88 days in a time span of four months (June 18 to October 11, 2004) show (Figs. 2 and 3) that the circulation in the bay was always cyclonic, and increased in intensity from June to August and then decreased in September and October. Speeds were ~0.25 m s^{-1}, but it is evident from Figure 2 that the rotation period was not constant. Using power spectra (not showed), we determined that the rotation period varied between 0.92 and 3.15 days, and identified five sampling intervals in which certain rotation periods were dominant, as shown in Table I. The drifter tracks with surface velocity vectors for those five intervals are shown in Figure 3a-e.

In sampling intervals 1, 2 and 5 (Table I) the most important part of the variance was contained in a band of rotation frequencies (column 3), while in intervals 3 and 4, there was a dominant period of rotation, 1.35 and 1.4 days, respectively. The inertial period (1.19 days) appears in the band of the dominant frequencies only in the second interval of observation. The other periods of rotation are far from the inertial and higher frequencies. The longest periods of rotation occurred in the last sampling interval: 2.52-3.15 days, which suggests that by September-October 2004 the cyclonic circulation could be slowing down.

The horizontal distributions of surface density anomaly and the 0-50 m average LADCP velocity measured on

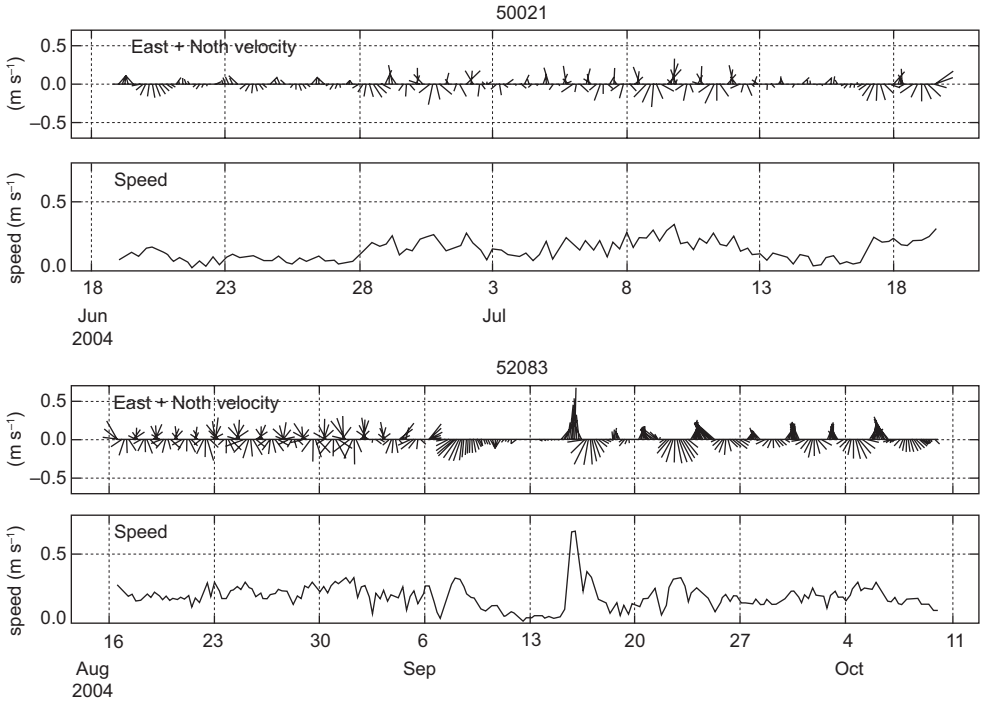

Fig. 2. Time-series of velocity of drifters inside Bahía de la Paz: (a) drifter 50021, from June 18 to July 18, 2004; (b) drifter 52083, from August 16 to October 11, 2004.

August 14-16, 2004, are shown in Figure 4a Observed velocities show the closed cyclonic circulation in the northern sector of Bahía de La Paz, and follow the closed contours of density, which have a maximum in the center of the closed circulation. The contours of temperature, averaged in the first 50 m depth (Fig. 4b) are similar to those of density anomaly (Fig. 4a) and contain a minimum in the center of the circulation. The corresponding spatial distribution of salinity (Fig. 4b) showed less-salty waters in the north of the bay, and the isohalines suggests saltier waters are being advected northward close to Espíritu Santo Island. This suggests that at the time of the observations the interchange of waters between Bahía de La Paz and the surrounding GC was in the cyclonic sense, with fresher waters coming into the bay in the northwest, and exiting the bay in the southeast sector of the mouth, close to the northern tip of Espíritu Santo Island. As Bahía de La Paz is an evaporative basin (Obeso-Nieblas et al., 2014), the water that exits the bay is saltier than that entering.

The vertical structure of observed hydrography and velocity are shown in Figure 5 along transect L1-L7 (see position in Fig. 1), which extends from the western coast of the bay to ~20 km into the GC (Fig 5). The vertical distribution of temperature shows that inside the bay all the isotherms above 14 ºC were displaced upward, relative to their position outside the bay (Fig. 5a). The salinity indicates that most of the bay was filled with water from the Gulf of California ($S \geq 34.9$) (Fig. 5b). The dissolved oxygen distribution (Fig. 5c) shows the decrease with depth previously reported by Monreal-Gómez et al. (2001), a feature of the oxygen minimum zone of the GC and the Eastern Tropical Pacific (Fiedler and Talley, 2006; Cepeda-Morales et al., 2013; Lavín et al., 2013; Castro et al., 2017; Collins et al., 2015; Evans et al., 2020). The DO isolines also show the upward displacement inside the bay, relative to the conditions outside. The fluorescence (Fig. 5d) shows enrichment in the upper 75 m, with maximum on the thermocline, and a slight upward displacement of the isolines inside the bay.

The geostrophic velocity across the same line of stations, relative to the minimum common depth of pairs of stations (Fig. 5e), shows that the cyclonic

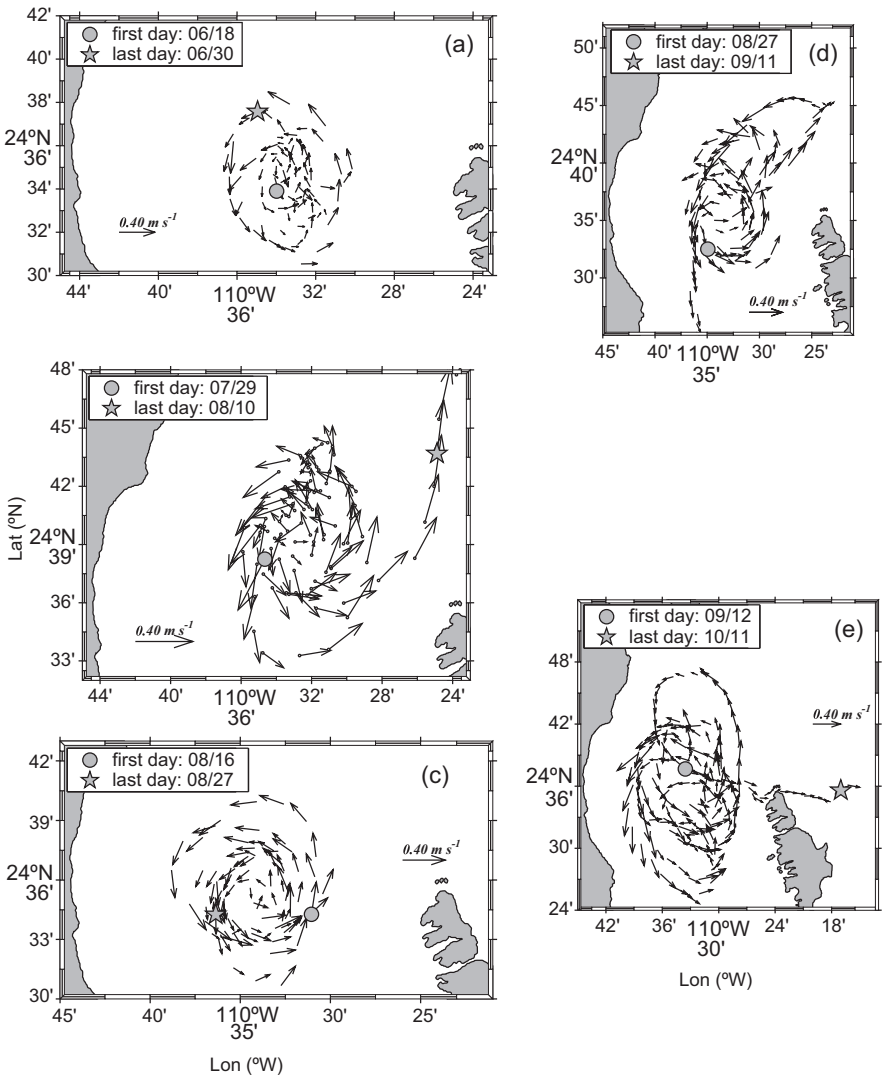

Fig. 3. Drifter tracks and velocity for (Table I): (a) first period, 06/18-06/30; (b) second period, 07/29-08/10; (c) third period, 08/16-08/27; (d) fourth period, 08/27-09/11; (e) fifth period, 09/12-10/11.

Table I. Characteristics of the Bahía de la Paz cyclonic eddy in the summer of 2004, from the satellite-tracked surface drifters.

	Mean velocity ($m\ s^{-1}$)	Rotation period (days)	Explained variance (%)	Interval of sampling	Time spans (days)	Rossby number
Period 1	0.15	1.50 to 2.00	41	06/18-06/30	13	0.81-0.60
Period 2	0.25	0.92 to 1.40	45	07/29-08/10	13	1.32-0.83
Period 3	0.25	1.35	68	08/16-08/27	12	0.89
Period 4	0.20	1.40	50	08/27-09/11	14	0.83
Period 5	0.20	2.52 to 3.15	60	09/12-10/11	29	0.48-0.39

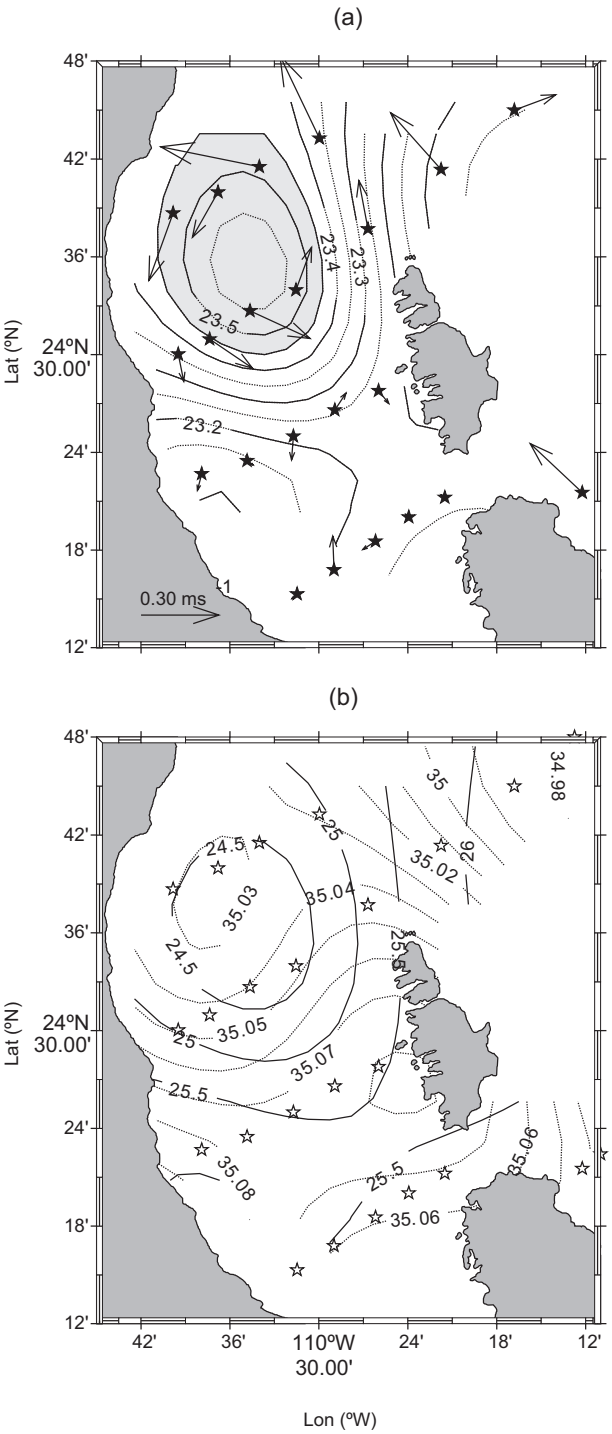

Fig. 4. Spatial distributions (averaged in the top 50 m) of: (a) LADCP current velocities (m s^{-1}) and potential density anomaly (kg m^{-3}) (the shaded area indicates potential density greater than 24.45); (b) potential temperature (°C, full lines) and salinity (dashed lines).

circulation inside the bay (stations L1-L5) was mostly in the top 100 m. The figure also suggests that the flow inside the bay was linked with northward flow outside; that is, to a northward coastal current in the GC (stations L5-L7). This is also suggested by the LADCP data in Figure 4a. The vertical distribution of LADCP velocity (Fig. 5f) shows cyclonic circulation, in a pattern similar to the geostrophic velocity distribution (Fig. 5e), but with lower speeds; it also shows important shears in the deeper levels. Figure 6 is a comparison between LADCP surface velocities (thin arrows) and surface geostrophic velocities (thick arrows), overlaid on contours of geopotential anomaly (the latter two relatives to 100 m). Although cyclonic circulation is present in the two sets of vectors, the geostrophic speed is higher than the LADCP speed. As Table I shows, the mean drifter velocities (0.20-0.25 m s^{-1}) were similar to the LADCP near-surface velocities (0.15-0.20 m s^{-1}, Fig. 5f), and both were lower than the surface geostrophic velocities (0.25-0.35 m s^{-1}, Fig. 5e). Durán-Campos et al. (2015), found geostrophic velocities greater than 0.20 m s^{-1} above 60 m depth, with maxima of 0.70-0.8 m s^{-1} at 20 m depth, while Sánchez-Mejía et al. (2020) reported maximum geostrophic velocities (0.95 m s^{-1}) in the periphery of the cyclonic eddy during August 2017. These values overpassed the speed of this work using surface drifters and geostrophic velocities during the summer (Fig. 2).

3.2 Cyclonic circulation characteristics

The direct observations provided by the surface drifters and the LADCP give support to previous descriptions (based on geostrophic velocity calculations) of an intense surface cyclonic circulation in Bahía de La Paz during the warm period (June to October). However, the fact that the geostrophic speeds are higher than both the LADCP and the drifter speeds suggests that the dynamics of the cyclonic eddy is not simply geostrophic as suggested previously (Monreal-Gómez et al., 2001; Sánchez-Velasco et al., 2006; Durán-Campos et al., 2015). During cruise observations the wind velocities were very small, as can be seen in Figure 1b, and it is possible to assume that the geostrophic balance is the main part of the dynamics.

In cylindrical coordinates, the balance of forces in the radial direction is (Cushman-Roisin, 1994)

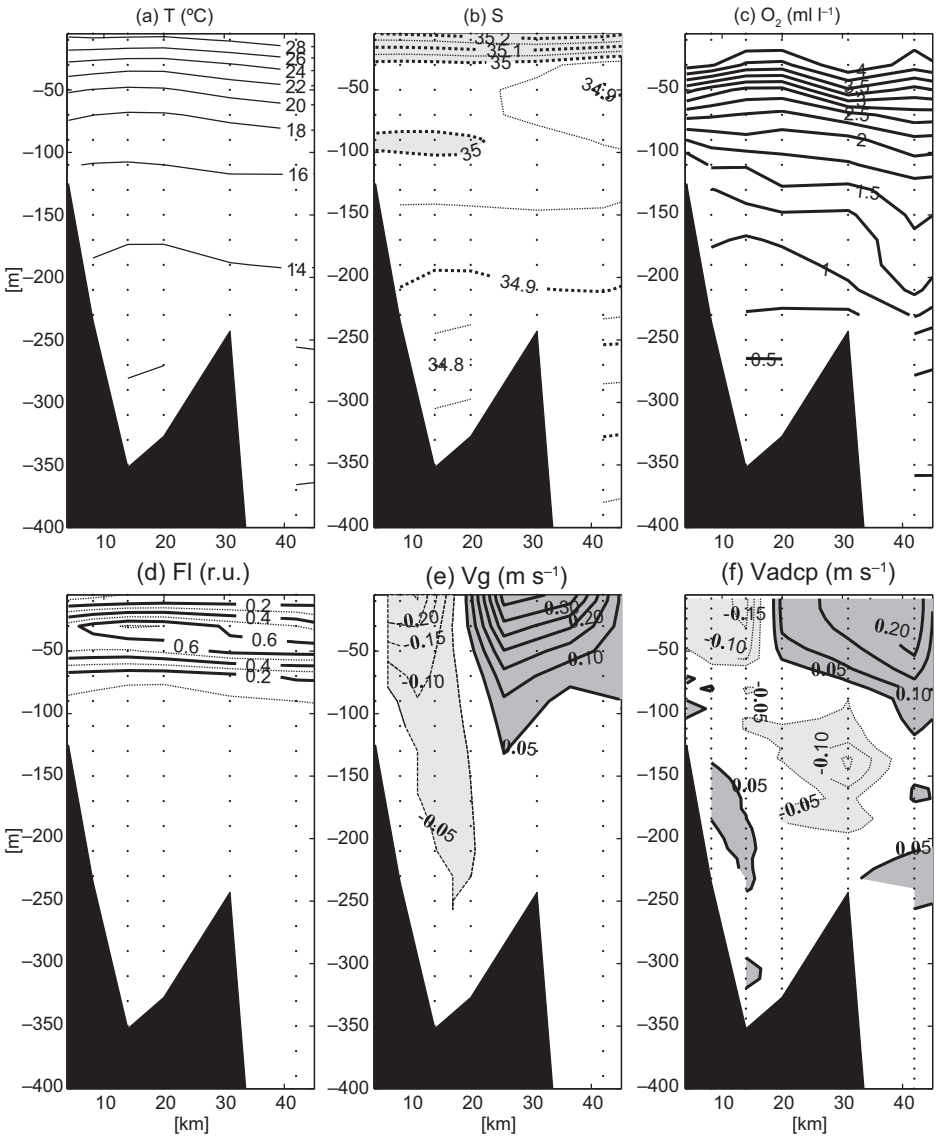

Fig. 5. Vertical distributions on the line of stations L1 to L6 (see Fig. 1): (a) potential temperature (ºC), (b) salinity, (c) dissolved oxygen (ml L^{-1}), (d) fluorescence (relative units), (e) geostrophic velocity (m s^{-1}) relative to minimum common depth between pairs of stations, and (f) normal component to the section of the current velocity (m s^{-1}) from the LADCP.

$$fv + \frac{V^2}{r} = \frac{1}{\rho_o}\frac{\partial P}{\partial r} \quad (4)$$

$$R_o = \frac{V}{fr} = \frac{\omega_b}{f} = 0.83 \quad (5)$$

where r is the distance from the center of the eddy, V the orbital speed (positive counterclockwise), f the Coriolis parameter, ρ_0 a constant water density, and P the water pressure. The ratio of the centrifugal force to the Coriolis force (the Rossby number) in the cyclonic circulation of Bahía de la Paz was

where ω_b is the angular frequency of the eddy, which was obtained from the rotation periods measured during the cruise, but the Rossby numbers for all periods are given in Table I. Therefore, the centrifugal force cannot be neglected; the closed circulation is of "intermediate size", meaning that its radius is

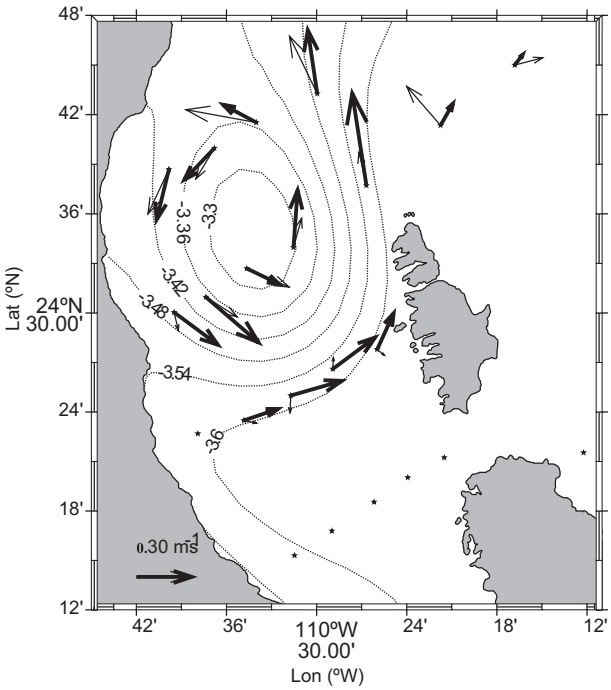

Fig. 6. Comparison between LADCP surface velocities (thin arrows) and surface geostrophic velocities (thick arrows), overlaid on contours of geopotential anomaly (m^{-2} s^{-2}) relative to 100 m depth.

similar to the Rossby radius of deformation (Cushman-Roisin, 1994).

For the purpose of comparing V with the geostrophic velocity, $V_g = \frac{1}{f\rho_o}\frac{\partial P}{\partial r}$, Eq. (1) is rewritten as

$$\frac{V_g}{V} = 1 + \frac{V}{fr} \tag{6}$$

For a cyclonic eddy in the northern hemisphere, like the circulation in Bahía de La Paz (Fig. 4a), $V > 0$ by definition, and therefore $V_g > V$; by contrast, for an anticyclonic eddy, $V_g/V < 1$. The physical explanation is that in cyclonic eddies the pressure gradient is opposed by both, Coriolis and the centrifugal forces, therefore geostrophic calculations will produce faster speeds than observed. In anticyclonic eddies, the Coriolis force balances the combination of centrifugal and pressure gradient forces, and the resulting velocity is higher than geostrophic calculations would produce. This explains why the geostrophic velocities in Figure 5e are stronger than the LADCP currents in Figure 5f, a relation that is also seen in Figure 6. To further quantify this relationship, we show in Figure 7 a plot of V_{LADCP} against V_g corresponding to the top 120 m of transect L1-L6 of Figure 5b, c (in 8 m bins). The origin-crossing regression line in Figure 7 gives $V_{\mathrm{LADCP}} = 0.70\ V_{\mathrm{g}}$, which is close to the 0.55 that is obtained from Eq. (2) and the Ro = 0.8 estimated above from the eddy rotation period given by the drifters. The linear regression in Figure 7 has a standard deviation $(V_{\mathrm{LADCP}} - V_g) = 0.08$ m s^{-1}, and the correlation coefficient is 0.86 at the 99% confidence level.

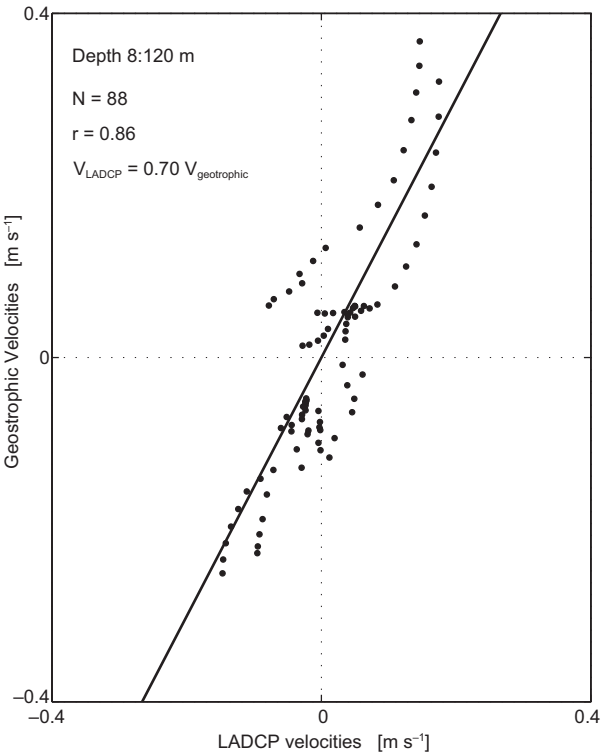

Fig. 7. Relation between observed velocities from LADCP and the observed geostrophic velocities along transect L1 to L7 (see Fig. 1) in the first 120 m depth.

3.3 Surface temperature and chlorophyll

The fluorescence averaged in the top 50 m (Fig. 8) shows a maximum that coincides in position with the SST minimum (Fig. 4b); that is, it is located in the center of the cyclonic circulation. The coolness and chlorophyll (CHL) richness of the bay's surface water (relative to the nearby waters of the Gulf) was apparent in the MODIS satellite images collected throughout the summer of 2004. Figure 9 shows

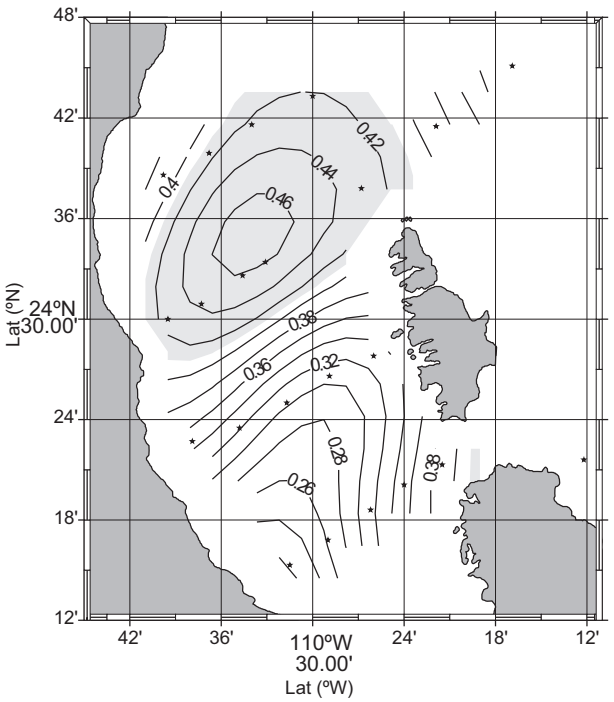

Fig. 8. Fluorescence (relative units) averaged over the top 50 m of the water column during cruise NAME-II.

examples of SST and CHL images for the weeks centered on May 13, June 23 and July 25. To prove that this is a recurrent summer feature of Bahía de La Paz, the weekly-mean SST and CHL MODIS time series of images comprising the bay and surrounding waters was harmonically-analyzed as described in section 2. Then the mean, the annual and the semi-annual harmonics were used to construct the seasonal evolution of SST and CHL, part of which is shown in Figure 10, where the left panels show SST and the right panels show CHL, for the months of June, July, August and February.

While the SST of the GC and the bay increases from June to August (Fig. 10a, c, e), the SST of Bahía de La Paz remains cooler than the surrounding southern GC. Moreover, Figure 10a, c shows an SST relative minimum nucleus inside the bay close to the center of the cyclonic circulation described above. The corresponding distributions of surface CHL for June, July and August (Fig. 10b, d, f) show that although the CHL decreased as the warm seas advanced, throughout the summer the bay maintained higher values of CHL than the adjacent GC. Furthermore, a relative maximum of CHL is found during these months approximately at the center of the cyclonic circulation described above. A contrasting situation is found in February (Fig. 10g, h): the SST inside the bay is slightly warmer than the surroundings, and the CHL is high everywhere in the southern GC, including the bay. Figure 10 also shows that the cool and CHL-rich summer conditions are not constrained to Bahía de La Paz, but are present in a narrow coastal band adjacent to the peninsula, north and south of the bay.

4. Discussion

During the observation period, there were two possible forcing mechanisms for the dynamics inside the bay: (i) the circulation outside the bay and (ii) the curl of the local wind stress. Of course, both mechanisms could act at the same time.

4.1 The outside currents

The hypothesis that the circulation in the GC could force a cyclonic circulation inside the bay was pointed out by Sánchez-Velasco et al. (2006), who supported this idea on the results of previous theoretical and numerical work on the annual circulation in the GC (Beier, 1997; Ripa, 1997). The proposed mechanism is that during the warm period a seasonal internal coastal trapped wave (trapped within ~70 km of the coast) runs cyclonically the whole periphery of the Gulf, and it could force the circulation in a deep basin with a wide entrance, like Bahía de La Paz. Evidence for forcing from outside are the surface distributions of salinity presented here (Fig. 4b) and by Monreal-Gómez et al. (2001), the distribution of fish larvae assemblages of Sánchez-Velasco et al. (2006) and the distribution of zooplankton biomass reported in Durán-Campos et al. (2015).

However, the circulation over the Southern GC during the NAME-II campaign was dominated by a train of cyclonic and anticyclonic eddies with diameters ~70 km (Zamudio et al., 2008; Lavín et al., 2013, 2014). In particular, during August 2004, at northeast of Bahía de La Paz bay a cyclonic eddy occupied much of the Farallón Basin and showed mean swirl speed of ~0.30 m s^{-1} and a mean core radius of 33 km (Lavín et al., 2013). The horizontal scale of the Gulf of California eddies is one order of magnitude larger

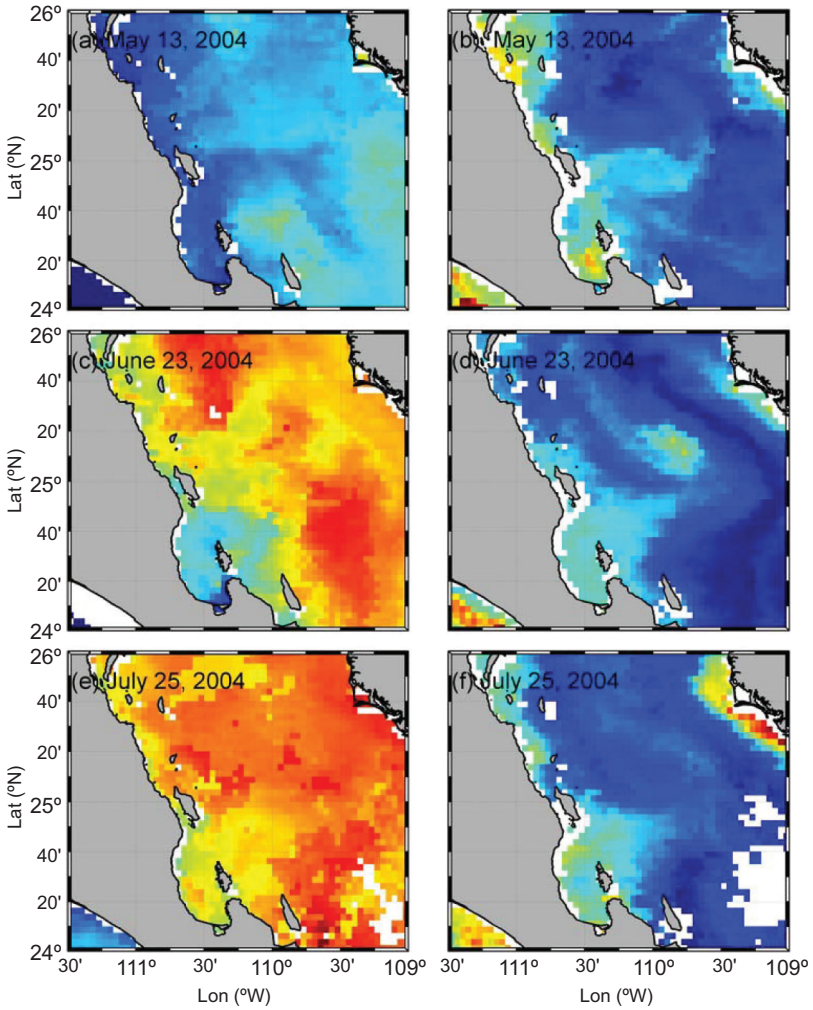

Fig. 9. Examples of SST (a, c, e) and CHL (b, d, f) centered at May 13, June 23 and July 25, 2004.

than that of Bahía de la Paz, therefore in principle they are capable of forcing the circulation inside the bay. If the circulation inside the bay was forced by the Gulf of California eddies, then either sense of rotation could be induced, since they tend to occur with alternate vorticity (Pegau et al., 2002; Zamudio et al., 2008; Lavín et al., 2013, 2014). This would explain the observations of anticyclonic circulation inside the bay by Salinas-González et al. (2003) and could in part also explain the cyclonic circulation described in this work. Durán-Campos et al. (2015) suggested that the origin of the cyclonic eddy could be associated with the shape of the basin, like the interchange of the currents entering to the bay by the north mouth area and the bathymetric effect. No doubt there is a considerable interaction between the bay and the GC, but we consider that more observational and modeling work is necessary to establish this causal relationship for the bay's circulation.

4.2 The wind stress curl

The upward displacement of the higher isothermals, DO and fluorescence isolines (Fig. 5), suggests that Ekman pumping induced by wind stress curl could partially contributed to the forcing mechanism of the cyclonic circulation. The wind stress observed in the bay during the August 2004 NAME-II cruise (not shown) was mainly from the south in the southern

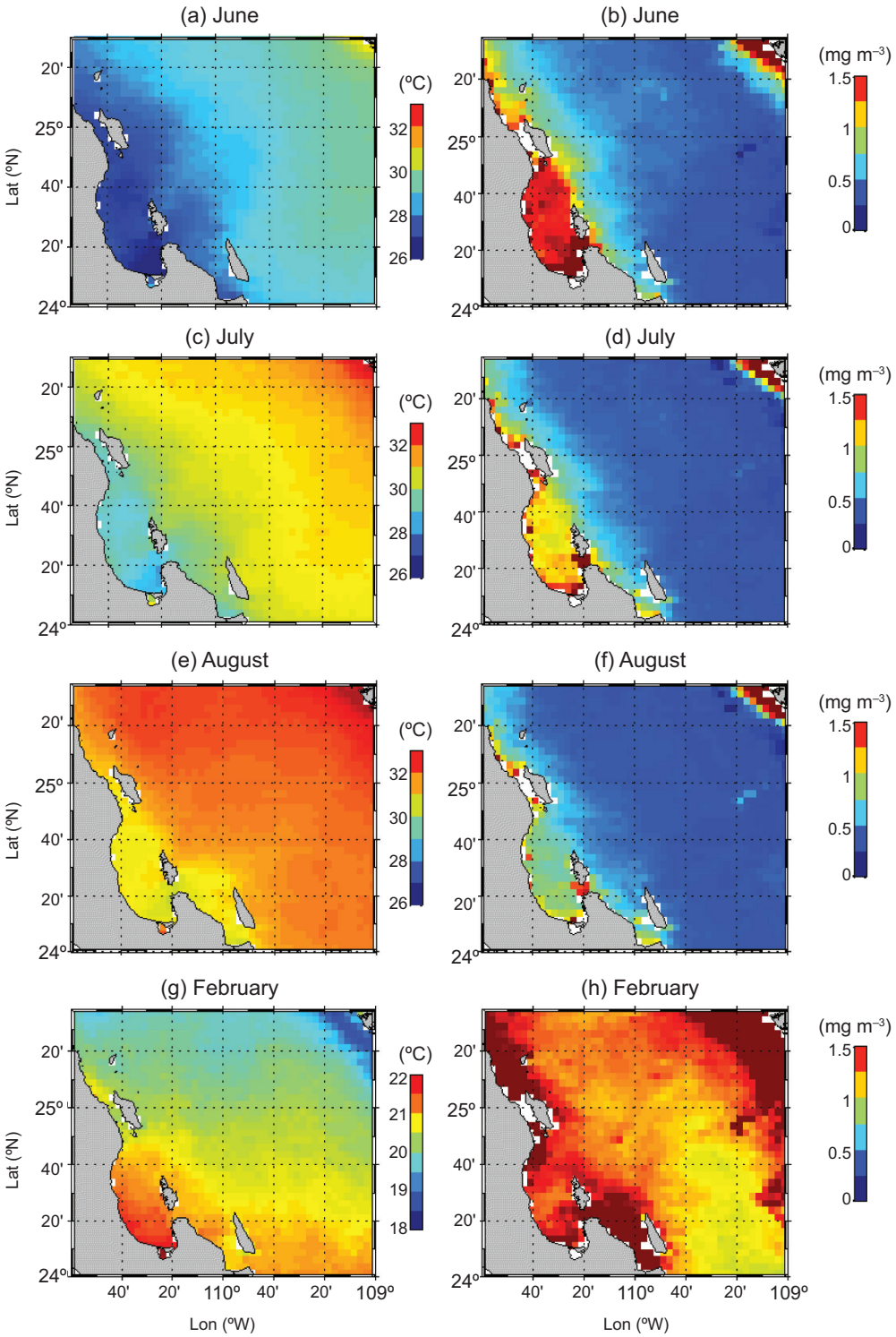

Fig. 10. Seasonal evolution of SST (left panels) and CHL (right panels) from MODIS, during (a) June, (b) July, (c) August, and (d) February.

sector of the bay but rotated westward in the north basin; in principle this positive vorticity could be transferred to the basin. A wind pattern with positive vorticity was also observed in the June 1998 cruise of Monreal-Gómez et al. (2001). Despite this agreement on the sign of the wind stress curl in the two cruises, its effect on the circulation should be estimated with caution, since those measurements took place over only a couple of days, and the winds in Bahía de La Paz are notoriously variable, the breeze system is important (Turrent and Zaitsev, 2014; Cervantes-Duarte et al., 2021). It would be necessary to show that the correct wind stress curl was maintained for a sufficient length of time to generate the upward Ekman pumping, and that the magnitude of the wind stress curl at the appropriate time scale could produce a reasonable (i.e., as observed) displacement of the pycnocline.

The wind stress curl could cause an Ekman pumping, provided that the time scales involved are longer than f^{-1} (~4.6 days in Bahía de La Paz, ~24.5º N). If a positive wind stress curl is present during the summer as a seasonal forcing, it could explain the seasonal behavior of SST and CHL of the bay described above, by virtue of Ekman pumping decreasing the surface temperature and increasing the nutrients and therefore promoting the increase of phytoplankton (Coria-Monter et al., 2014; Herrera-Cervantes, 2019; Sánchez-Mejía et al., 2020). It would also contribute partially to the cyclonic circulation via isopycnal uplift.

The seasonal component of the wind and the wind stress curl in the southern Gulf of California, constructed the same way as the SST and CHL seasonal cycle (Fig. 10), are shown in Figure 11. During the summer period (June, July and August, Fig. 11a-c), the wind blows equatorward on the western side of the peninsula, parallel to the mountain ridge (Fig. 1b) (Castro and Martínez, 2010). However, the wind turns counterclockwise toward the GC over the peninsula and especially when it passes beyond its tip; inside the GC, from May to July it progressively blows to the northeast and then to the north (Bordoni et al., 2004; Lavín et al., 2009, 2014; Herrera-Cervantes et al., 2017). During February the wind blows from the northwest both in the Pacific and in the GC, and the wind stress curl was negative (positive) at east (west) side of the peninsula (Fig. 11d).

The wind distribution causes the peninsula to be surrounded from June to August by a band of positive wind stress curl (Fig. 11a-c). The band on the Pacific side is caused by the lateral wind shear due to the presence of land, while the band that covers from the maximum curl south of the tip of the peninsula and along and off the gulf side of the peninsula is caused by the wind backing (i.e., turning counterclockwise). Bahía de La Paz is located within the coastal band of positive curl inside the GC, and it may present a relative maximum because it is located at the end of a topography gap connecting the Pacific lower atmosphere with that of the GC (Fig. 1b). The wind turning cyclonically at around 24º N is a possible mechanism for transferring positive vorticity to the bay at the seasonal time-scale. Trasviña-Castro et al. (2003) proposed a similar mechanism to explain the upwelling filaments caused by the Santa Ana winds in the Pacific coasts of northern Baja California, in a similar fashion as the generation of eddies in the Gulf of Tehuantepec by the Norte winds (see Willett et al., 2006 and references therein), but their time scales are a few days at most. The wind stress curl values seen in Figure 11a-c are very high (O[10^{-7} N m^{-3}]), only comparable with those found in the Gulf of Tehuantepec (Chelton et al., 2004).

The upward Ekman pumping around the coasts of the peninsula caused by the cyclonic wind stress curl described above is additional to the upwelling generated by offshore Ekman transport caused by the along-shore winds in the presence of the solid coastal boundary. The latter depends only on the component of the wind stress parallel to the coast, while the former depends on the curl strength rather than on the wind stress intensity (Bakun and Nelson, 1991; Castro and Martínez, 2010). Both mechanisms are at work on both sides of the peninsula during summer (Fig. 11a-c), while in winter (Fig. 11d) only coastal upwelling is at work, off the Pacific side of the peninsula and off the mainland side of the GC. The across-gulf tilt of the isotherms noted by Lavín et al. (2009) in the southern GC during June 2004 may have been caused by a wind stress curl similar to that shown in Figure 10a, b, which is seen to change sign across the gulf's axis.

To synthesize the seasonal behavior of wind stress curl, SST and CHL in Bahía de La Paz we space-averaged over the bay the respective weekly

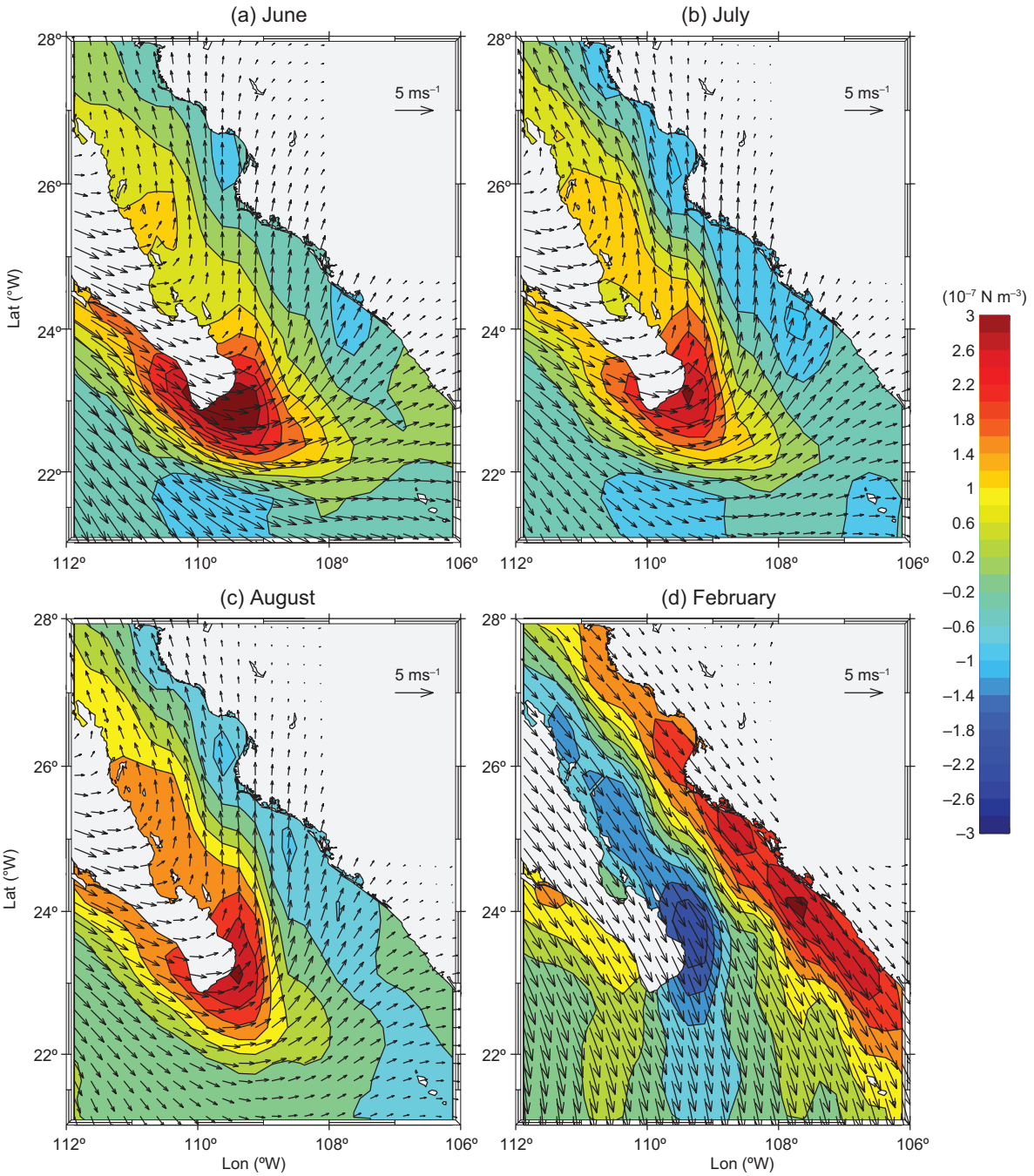

Fig. 11. Seasonal CCMP wind velocity (arrows, m s^{-1}), and wind stress curl (color, 10^7 N m^{-3}), during: (a) June, (b) July, (c) August, and (d) February.

time series for the period 2003-2007; these weekly space-averaged time series are shown in Figure 12, together with their seasonal fits. The parameters of the seasonal fits are listed in Table II. It is noteworthy that the semiannual component is very important; in the case of the wind stress curl and CHL it is as important as the annual one.

The seasonal fit to the wind stress curl (Fig. 12a), which explained 30% of total variance, shows important mean positive values from May to August.

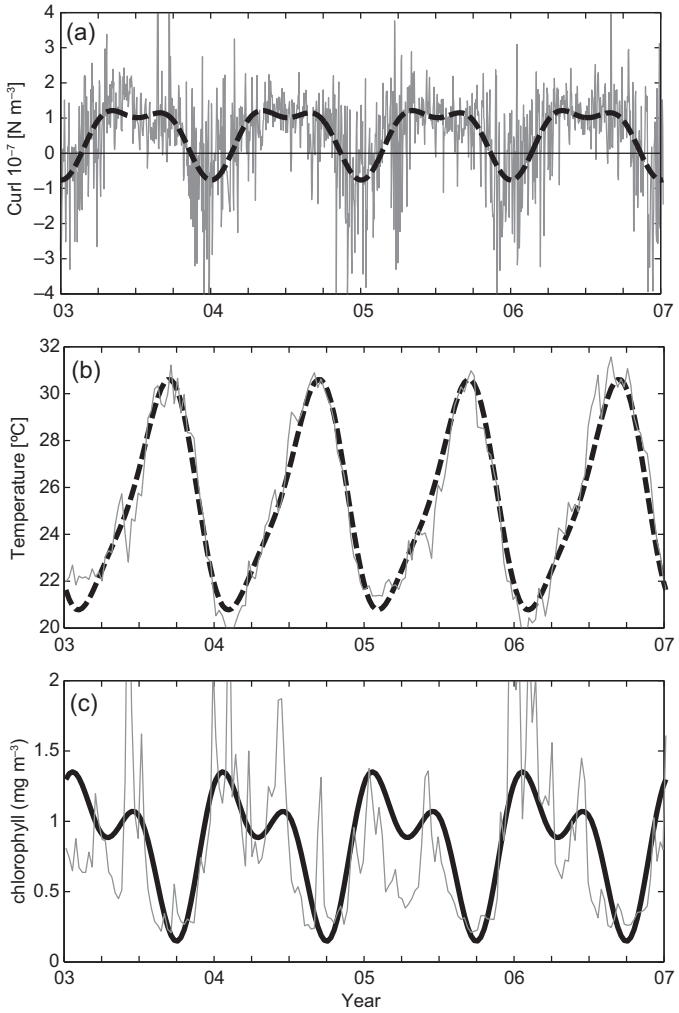

Fig. 12. Time series of the average inside Bahía de La Paz of the CCMP wind stress curl (upper panel), MODIS SST (middle panel), and MODIS CHL (lower panel).

Table II. Parameters of the seasonal fitting.

	Temporal mean	Annual amplitude	Annual phase	Semiannual amplitude	Semiannual phase	Explained variance
SST	25.3 ± 0.06	4.62 ± 0.08	15 August ±1 day	1.10 ± 0.08	15 September ± 2 days	94%
Curl	0.53 ± .02	0.89 ± 0.03	30 June ± 1 day	0.41 ± 0.03	2 October ± 2 days	29%
CHL	0.82 ± 0.03	0.40 ± 0.05	27 February ± 7 days	0.31 ± 0.05	20 June ± 4 days	31%

Superimposed on the seasonal signal there are strong oscillations of the wind stress curl in a period of days (hence the low explained variance), but from May to August of each sampled year the seasonal signal dominated the variance in summer. Although the seasonal fit shows anticyclonic values during winter, the mesoscale variability and shorter time scale cause wide oscillation of the curl, from positive to negative

values; these oscillations dominate the variance in winter. Superimposed to the seasonal fitting, there is strong non-seasonal variability in the wind stress curl which could increase the transfer of positive vorticity to the bay, for example during February-March 2006; or decrease the transfer of positive vorticity as in July-August 2003.

The SST and its seasonal fit are shown in Figure 12b. From May to August, when the wind stress curl is positive, the SST increases in Bahía the La Paz, just like in the entire GC, but as proved above the SST is cooler than in the Gulf, so that it lags its seasonal summer surface heating (Fig. 10). This lag can be related with the transfer of positive vorticity to the bay by the wind stress curl and the associated upward Ekman pumping of cool water. The bay-averaged CHL values and its seasonal fit are shown in Figure 12c. The seasonal variability shows high values of CHL from May to August, which correlates with positive wind stress curl. Note that the plumes of cool, chlorophyll-rich water apparently coming out of Bahía de La Paz seen in Figure 9 do not appear on Figure 10; this indicates that they are mesoscale structures with high variability.

Thus, during summer in Bahía de La Paz, the wind stress curl Ekman pumping brings cool, nutrient-rich subsurface water to the surface, causing chlorophyll enrichment of the upper layers (with maximum at the thermocline). This implies an input of sub-thermocline water into the bay, which by conservation has to be exported in the surface layers; this must be an important feature of the bay-gulf interaction. This could be the explanation of the cool, chlorophyll-rich plumes seen exiting Bahía de La Paz in Figure 9.

5. Conclusions

The first direct current observations (LADCP and surface drifters) in Bahía de la Paz concur with previous reports of cyclonic circulation in the northern sector of the bay. However, it was found that the centrifugal force cannot be neglected. This caused the cyclonic flow to be subgeostrophic; actual speeds (0.20-0.25 m s^{-1}) were ~25-40% lower than those estimated from geostrophic balance (0.25-0.35 m s^{-1}).

The observed hydrographic structure and the analysis of satellite data (wind velocity, SST and CHL), suggest that a partial mechanism on the contribution of the summer (June-August) cyclonic circulation, could be originated by Ekman pumping due to a strong positive wind stress curl. The process raises the thermocline/pycnocline closer to the surface than in the surrounding waters of the Gulf, causing the relative cooling and chlorophyll enrichment of the upper layers of the bay.

Acknowledgments

This study was supported by CONACYT project D41881-F (P.I., M.F. Lavín), by the North American Monsoon Experiment (NOAA contract GC04-219, P.I. Dr. Michael Douglas), and by UABC project P-0324-0352 (P.I. Rubén Castro). Additional support was provided by CICESE. Technical support by: Carlos Cabrera, Karla P. Godínez, Joaquín García and Carlos Flores. We thank the support of the skipper and crew of the RV Francisco de Ulloa.

References

Atlas R, Hoffman RN, Ardizzone J, Leidner SM, Jusem JC, Smith DK, Gombos D. 2011. A cross-calibrated, multiplatform ocean surface wind velocity product for meteorological and oceanographic applications. Bulletin of the American Meteorological Society 92: 157-174. https://doi.org/10.1175/2010BAMS2946.1

Bakun A, Nelson CS. 1991. The seasonal cycle of wind-stress curl in subtropical eastern boundary current regions. Journal of Physical Oceanography 21: 1815-1834. https://doi.org/10.1175/1520-0485(1991)021<1815:TSCOWS>2.0.CO;2

Beier E. 1997. A numerical investigation of the annual variability in the Gulf of California. Journal of Physical Oceanography 27: 615-632. https://doi.org/10.1175/1520-0485(1997)027<0615:ANIOTA>2.0.CO;2

Beron-Vera FJ, Ripa P. 2002. Seasonal salinity balance in the Gulf of California. Journal of Geophysical Research-Oceans 107: 15-1-15-15. https://doi.org/10.1029/2000JC000769

Bordoni S, Ciesielski PE, Johnson RH, McNoldy BD, Stevens B. 2004. The low-level circulation of the North American Monsoon as revealed by QuikScat. Geophysical Research Letters 31: L10109. https://doi.org/10.1029/2004GL020009

Cushman-Roisin B. 1994. Introduction to geophysical fluid dynamics. Prentice-Hall, Englewood Cliffs, NJ.

Castro R, Godínez VM, Lavín MF, Beier E, Cabrera-Ramos C. 2006. Hydrography of the southern Gulf of California during the North American Monsoon Experiment: Campaign NAME-02, August 7-20, 2004. Technical report 46143. Centro de Investigación Científica y de Educación Superior de Ensenada, 221 pp.

Castro R, Martínez JA. 2010. Variabilidad espacial y temporal del campo de viento frente a la península de Baja California. En: Dinámica del ecosistema pelágico frente a Baja California, 1997-2007: diez años de investigaciones mexicanas de la corriente de California (Durazo R. y Gaxiola G., Eds.). Instituto Nacional de Ecología, CICESE, México, 129-147.

Castro R, Collins AC, Rago AT, Margolina T, Navarro-Olache LF. 2017. Corrientes, transportes y variabilidad termohalina en la entrada al Golfo de California (19-21 de abril de 2013). Ciencias Marinas 43: 173-190. https://doi.org/10.1016/j.dsr.2005.09.010

Cepeda-Morales J, Gaxiola-Castro G, Beier E, Godínez VM. 2013. The mechanisms involved in defining the northern boundary of the shallow oxygen minimum zone in the eastern tropical Pacific Ocean off Mexico. Deep-Sea Research Part I-Oceanographic Research Papers 76: 1-12. https://doi.org/10.1016/j.dsr.2013.02.004

Cervantes-Duarte R, González-Rodríguez E, Funes-Rodríguez R, Ramos-Rodríguez A, Torres-Hernández MY, Aguirre-Bahena F. 2021. Variability of net primary productivity and associated biophysical drivers in Bahía de La Paz (Mexico). Remote Sensing 13: 1644. https://doi.org/10.3390/rs13091644

Chelton DB, Schlax MG, Freilich MH, Miliff RF. 2004. Satellite measurements reveal persistent small-scale features in ocean winds. Science 303: 978-983. https://doi.org/10.1126/science.1091901

Collins CA, Castro R, Mascarenhas A. 2015. Properties of upper ocean fronts associated with water mass boundaries at the entrance to the Gulf of California, November 2004. Deep-Sea Research Part II-Topical Studies in Oceanography 119: 48-60. https://doi.org/10.1016/j.dsr2.2014.06.002

Coria-Monter E, Monreal-Gómez, MA, Salas-de-León, DA, Aldeco-Ramírez J, Merino-Ibarra M. 2014. Differential distribution of diatoms and dinoflagellates in a cyclonic eddy confined in the Bay of La Paz, Gulf of California. Journal of Geophysical Research-Oceans 119: 6258-6268, https://doi.org/10.1002/2014jc009916

Durán-Campos E, Salas-de-León DA, Monreal-Gómez MA, Aldeco-Ramírez JA, Coria-Monter E. 2015. Differential zooplankton aggregation due to relative vorticity in a semi-enclosed bay. Estuarine Coastal Shelf Science 164: 10-18. https://doi.org/10.1016/j.ecss.2015.06.030

Evans N, Boles E, Kwiecinski JV, Mullen S, Wolf M, Devol AH, Moriyasu R, Nam S, Babbin RA, Moffett, J.W. 2020. The role of water masses in shaping the distribution of redox active compounds in the eastern Tropical North Pacific oxygen deficient zone and influencing low oxygen concentrations in the eastern Pacific Ocean. Limnology and Oceanography 65: 1688-1705. https://doi.org/10.1002/lno.11412

Fiedler PC, Talley LD. 2006. Hydrography of the eastern tropical Pacific: A review. Progress in Oceanography 69: 143-180. https://doi.org/10.1016/j.pocean.2006.03.008

Herrera-Cervantes H, Beier E, Balart E. 2017. Red de monitoreo ambiental para el estudio de la variabilidad océano-atmósfera en la Bahía de La Paz, B.C.S., México. Recursos Naturales y Sociedad 3: 32-44. https://doi.org/10.18846/renaysoc.2017.03.03.02.0003

Herrera-Cervantes H. 2019. Sea surface temperature, ocean color and wind forcing patterns in the Bay of La Paz, Gulf of California: Seasonal variability. Atmósfera 32: 25-38. https://doi.org/10.20937/ATM.2019.32.01.03

Lavín MF, Marinone SG. 2003. An overview of the physical oceanography of the Gulf of California. In: Nonlinear processes in geophysical fluid dynamics (Velasco-Fuentes OU, Sheinbaum J, Ochoa de la Torre JL, Eds.). Kluwer Academic Publishers, Dordrecht, The Netherlands, 173-204.

Lavín MF, Castro R, Beier E, Godínez VM, Amador-Buenrostro A, Guest P. 2009. SST, thermohaline structure and circulation in the southern Gulf of California in June 2004, during the North American Monsoon Experiment. Journal of Geophysical Research-Oceans 114: C02025. https://doi.org/10.1029/2008jc004896

Lavín MF, Castro R, Beier E, Godínez VM. 2013. Mesoscale eddies in the southern Gulf of California during summer: Characteristics and interaction with the wind stress. Journal of Geophysical Research-Oceans 118: 1367-1381. https://doi.org/doi:10.1002/jgrc.20132

Lavín MF, Castro R, Beier E, Cabrera-Ramos CE, Godínez VM, Amador-Buenrostro A. 2014. Surface circulation

in the Gulf of California in summer from surface drifters and satellite images (2004-2006). Journal of Geophysical Research-Oceans 119: 4278-4290. https://doi.org/10.1002/2013jc009345

Marinone SG. 2003. A three-dimensional model of the mean and seasonal circulation of the Gulf of California. Journal of Geophysical Research-Oceans 108: 3325. https://doi.org/10.1029/2002jc001720

Marinone SG, Parés-Sierra A, Castro R, Mascarenhas Jr A. 2004. Correction to "Temporal and spatial variation of the surface winds in the Gulf of California". Geophysical Research Letters 31. https://doi.org/doi:10.1029/2004GL020064

Monreal-Gómez MA, Molina-Cruz A, Salas-de-León DA. 2001. Water masses and cyclonic circulation in Bay of La Paz, Gulf of California, during June 1998. Journal of Marine Systems 30: 305-315. https://doi.org/10.1016/S0924-7963(01)00064-1

Muñoz-Barbosa A, Delgadillo-Hinojosa F, Torres-Delgado EV, Félix-Bermúdez A, Castro R. 2020. Bajacalifornian dust deposition and atmospheric input of iron to the Gulf of California during the summer. Marine Chemistry 225: 103850. https://doi.org/10.1016/j.marchem.2020.103850

NASA. 2018. Sea-viewing wide field-of-view sensor (SeaWiFS) ocean color data. NASA Goddard Space Flight Center, Ocean Ecology Laboratory, Ocean Biology Processing Group. https://doi.org/10.5067/ORBVIEW-2/SEAWIFS/L2/OC/2018

Obeso-Nieblas M, Gaviño-Rodríguez JH, Obeso-Huerta H, Muñoz-Casillas SI. 2014. Variabilidad espacial termohalina, masas de agua y circulación geostrófica en Bahía de La Paz, Golfo de California. Revista de Biología Marina y Oceanografía 49: 413-426.

Pegau ES, Boss E, Martínez A. 2002. Ocean color observations of eddies during the summer in the Gulf of California. Geophysical Research Letters 29: 1295. https://doi.org/10.1029/2001gl014076

Ripa P. 1997. Toward a physical explanation of the seasonal dynamics and thermodynamics of the Gulf of California. Journal of Physical Oceanography 27: 597-614. https://doi.org/10.1175/1520-0485(1997)027<0597:TAPEOT>2.0.CO;2

Salinas-González, F, Zaytsev O, Makarov V. 2003. Formation of thermohaline structure of water in Bahía de La Paz from summer to autumn. Ciencias Marinas 29: 51-65.

Sánchez-Mejía, J.D., Monreal-Gómez, M.A., Durán-Campos, E., Salas-de-León, A.D., Coria-Monter E, Contreras-Simuta MG, Merino-Ibarra M. 2020. Impact of a mesoscale cyclonic eddy on the phytoplankton biomass of Bay of La Paz in the Southern Gulf of California. Pacific Science 74: 331-344. https://doi.org/10.2984/74.4.2

Sánchez-Velasco L, Beier E, Avalos-García C, Lavín, MF. 2006. Larval fish assemblages and geostrophic circulation in Bahía de La Paz and the surrounding SW region of the Gulf of California. Journal of Plankton Research 28: 1081-1098. https://doi.org/10.1093/plankt/fbl040

Trenberth KE, Olson GJ, Large WH. 1989. A global ocean wind stress climatology based on ECMWF analyses NCAR/TN-338+STR, NCAR. Technical note. Climate and Global Dynamics Division, National Center for Atmospheric Research, Boulder, Colorado.

Trasviña-Castro A, Ortiz-Figueroa M, Herrera H, Cosío MA, González E. 2003. 'Santa Ana' winds and upwelling filaments off Northern Baja California. Dynamics of Atmospheres and Oceans 37: 113-129. https://doi.org/10.1016/S0377-0265(03)00018-6

Turrent C, Zaitsev O. 2014. Seasonal cycle of the near-surface diurnal wind field over the bay of La Paz, Mexico. Boundary-Layer Meteorology 151: 353-371. https://doi.org/10.1007/s10546-014-9908-4

UNESCO. 1991. Processing of oceanographic station data. Technical papers in marine science. United Nations Educational, Scientific and Cultural Organization, Paris.

Visbeck M. 2002. Deep velocity profiling using lowered Acoustic Doppler Current Profiler: Bottom track and inverse solutions. Journal of Atmospheric and Oceanic Technology 19: 794-807. https://doi.org/10.1175/1520-0426(2002)019<0794:DVPULA>2.0.CO;2

Willett CS, Leben RR, Lavín MF. 2006. Eddies and tropical instability waves in the eastern tropical Pacific: A review. Progress in Oceanography 69: 218-238. https://doi.org/10.1016/j.pocean.2006.03.010

Zamudio L, Hogan P, Metzger EJ. 2008. Summer generation of the Southern Gulf of California eddy train. Journal of Geophysical Research 113: C06020. https://doi.org/10.1029/2007jc004467

Atmósfera 36(2), 225-238 (2023)
https://doi.org/10.20937/ATM.53051

CONTENTS

Understanding convective storms in a tropical, high-altitude location with in-situ meteorological observations and GPS-derived water vapor

Alejandro CASALLAS-GARCÍA[1,2]*, Daniel HERNÁNDEZ-DECKERS[1] and Héctor MORA-PÁEZ[3]

[1]*Grupo de Investigación en Ciencias Atmosféricas, Departamento de Geociencias, Universidad Nacional de Colombia, 111321, Bogotá, Colombia.*
[2]*Earth System Physics, The Abdus Salam International Centre for Theoretical Physics, 34151, Trieste, Italy.*
[3]*Grupo de Investigaciones Geodésicas Espaciales, Dirección de Geoamenazas, Servicio Geológico Colombiano, 11121, Bogotá, Colombia.*
*Corresponding author: alcasallasg@unal.edu.co

Received: March 6, 2021; accepted: June 28, 2021

RESUMEN

Con el fin de estudiar tormentas convectivas en la Sabana de Bogotá, un área densamente poblada de gran elevación en los Andes tropicales colombianos, identificamos eventos convectivos con imágenes satelitales y datos de precipitación. Los eventos son más frecuentes en las primeras horas de la tarde de las dos estaciones lluviosas. Estudios previos sugieren que los vientos del oeste de las primeras horas de la tarde y su advección de humedad desde el Valle del Magdalena son la principal causa de estas tormentas. Sin embargo, encontramos que estos vientos están presentes en el 78% de los días en temporada de lluvias, pero las tormentas se desarrollan solo en 26% de ellos. Así, aunque los vientos del oeste parecen necesarios debido a la convergencia que generan, sólo en ocasiones producen tormentas, de manera que no son un buen predictor. Datos de reanálisis indican que el vapor de agua precipitable (PWV, por su sigla en inglés) en el Valle del Magdalena es anormalmente bajo durante días convectivos, lo que sugiere que la humedad converge localmente en lugar de ser transportada desde el oeste. Usando análisis de compuestos, identificamos las señales más prominentes asociadas con la convección: la rapidez del viento es menor que la media en la mañana, mientras que la temperatura es superior al promedio al mediodía, seguida de un anómalo incremento del PWV y la rapidez del viento. Estas características indican que la convección resulta de un fuerte forzamiento diurno, facilitado por convergencia de los vientos del oeste y del vapor de agua, en una escala de tiempo de alrededor de 3 h.

ABSTRACT

We investigate convective storms over the Sabana de Bogotá, a high-altitude and densely populated area in the Colombian tropical Andes. Convective events are identified using infrared satellite images and in-situ precipitation data. As expected, convection shows a strong early-afternoon peak during the two rainy seasons. Previous studies hypothesize that early-afternoon westerly winds and their moisture advection from the warmer Magdalena valley are the main explanatory mechanism for intense storms. We find that early-afternoon westerlies are present in 78% of rainy season days, but convective events develop in only 26% of them. Thus, although westerlies seem necessary for convection due to the convergence they generate, they only occasionally generate storms and are therefore not a good predictor. Furthermore, reanalysis data indicate that precipitable water vapor (PWV) at the Magdalena valley is anomalously low during convective days, suggesting that moisture converges locally instead of being advected from the west. Based on composites of surface wind speed, air temperature, surface pressure, and GPS-derived PWV, we identify the most prominent signals associated with deep convection: a weaker than average wind speed throughout the morning, higher than normal values of surface air temperature towards noon, followed by an anomalous steep increase of PWV and wind speed. These features indicate that convection results from a strong diurnal forcing facilitated

by convergence of westerly winds, combined with sufficient water vapor convergence, with a timescale of about 3 h. This highlights the relevance of high temporal resolution monitoring of PWV offered by Global Navigational Satellite System stations.

Keywords: tropical convection, GPS meteorology, mountain meteorology.

1. Introduction

Atmospheric convection is one of the most important processes in Earth's climate system. It is crucial for the planet's energy balance due to its vertical transport of energy and moisture and is also responsible for most of the tropical precipitation (Sundqvist, 1978; Zeng, 1998; Betts and Jakob, 2002). Atmospheric convection is extremely complex, since it involves many spatio-temporal scales, and incorporates multiple feedbacks between radiation, fluid dynamics, thermodynamics, and microphysics. Its multiple-scale nature makes it hard to parametrize in climate models and constitutes a challenge for observations (Bretherton et al., 2004).

Convection is especially important in the tropics since it produces two-thirds of the precipitation in this area (Zeng, 1998). Thus, to understand precipitation in the tropics, it is essential to study the convective processes and parameters in this area. Unlike regions at higher latitudes, where strong horizontal temperature and pressure gradients are responsible for the main synoptic systems that determine precipitation, at low latitudes precipitation is mainly produced by local moist convective processes. Therefore, convection plays a very important role in the hydrological cycle of the tropics and is also essential for weather prediction (Sherwood et al., 2010).

Column-integrated water vapor, also often referred to as precipitable water vapor (PWV), has been shown to be a key variable in order to understand convective processes in the tropics (e.g., Bretherton et al., 2004; Holloway and Neelin, 2009, 2010; Adams et al., 2013, 2015). The combination of convergence and moisture is crucial for deep convection to develop since these two can interact to allow for convection to organize into deeper cells. Some authors (e.g., Sherwood et al., 2004) describe how dry layers of air in the mid-troposphere can inhibit the development of convection or can change its intensity. Other studies (e.g., Crook, 1996; Tompkins, 2001; Roberts et al., 2008) note that small variations in surface water vapor fields can produce considerable changes in convection development and intensity. For this reason, monitoring the evolution of PWV during the different stages of convective events can provide useful indications of local relevant processes. However, due to the relatively short time scales involved in atmospheric convection, high temporal resolution observations of PWV are required, which cannot be obtained from typical operational soundings, but rather from other methods such as Global Navigation Satellite System (GNSS) stations. GNSS has a subset called GPS which has geodetic applications that require very precise knowledge of the various delays that an electromagnetic wave is subjected to as it travels through Earth's atmosphere, and the total amount of water vapor in the atmospheric column is one of them. Since the 1990s, the methodology proposed by Bevis et al. (1992, 1994) provides a very precise estimate of the PWV from the total tropospheric delay of the GPS signal combined with observations of surface air temperature and pressure. The main advantages of this technology, which are of particular use for convection studies, are that it works under all-weather conditions and that it is capable of providing very high temporal resolution (~5 min).

For example, studies by Adams et al. (2011, 2013) use PWV from GPS stations to study fundamental features of convection in the Amazon. In particular, Adams et al. (2013) found two characteristic time scales for the shallow to deep transition of convection in the Amazon: an 8-h time scale of weak water vapor convergence, followed by a more intense 4-h time scale convergence. These results provide important insights into the typical behavior of convection, which can be used as benchmarks for numerical model evaluation. Furthermore, combining such observations with other meteorological variables can be useful to identify the different roles that different processes have within a convective event.

Here we use this approach to investigate the main characteristics of convective events that occur in the

Sabana de Bogotá, which hosts the Colombian capital with a population of around 7.5 million (Fig. 1). With a latitude of 4.5º N, it is located on a 2550 m high savanna surrounded by mountains in the tropical Andes. Thus, convection at this location is strongly affected by the mountain ranges, and despite its low latitude, the nature of convection events can be very different from other low-elevation locations such as those studied by Adams et al. (2013). Additionally, topography can change the local wind flow, and by doing so, modify the precipitation patterns (e.g., Fairman et al., 2011; Smith et al., 2011). This makes the Sabana de Bogotá a very particular study area, with very specific convective dynamics, but at the same time very relevant due to its tropical location and high population density. Montoya and Eslava (2000) describe how most storms in Bogotá are produced by diurnal convection. They hypothesize that intense storms in Bogotá develop due to humid and warm air masses transported by early-afternoon westerlies from the Magdalena valley (to the west of Bogotá), which produces convergence, instability, and humidifies the environment. However, there is still no direct observational confirmation of this mechanism.

To identify deep convective events in this area we use GOES-13 infrared images, TRMM (3B42) precipitation data every 3, with a spatial resolution of 0.25º (TRMM, 2011), and in situ precipitation data from five automatic rain gauges. To study the atmospheric conditions associated with these events, we use weather information from one of these locations (El Dorado airport), where data has best temporal coverage and quality. The low number of stations with the high temporal resolution required to study isolated convective

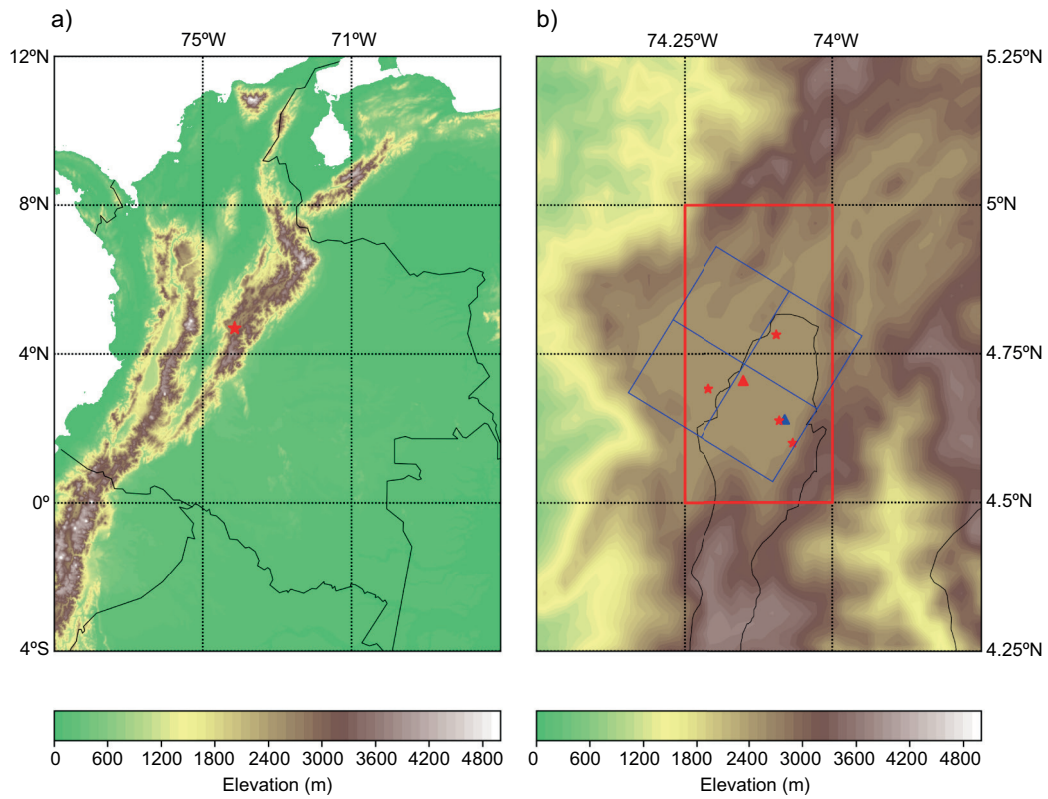

Fig. 1. (a) Location of the Sabana de Bogotá, Colombia (red star) and its surrounding topography. (b) Close-up of the surroundings of the Sabana de Bogotá. Deep convective events are identified in the blue grid, red stars indicate the locations of the rain gauges used to confirm these events, the red triangle denotes the location of the station used for composite analysis (El Dorado), and the blue triangle shows the location of the International GNSS Service (IGS) BOGT station. The red rectangle corresponds to the area where 3-hourly TRMM precipitation data is considered.

events provides a limitation to investigate urbanization effects (e.g., Freitag et al., 2018). However, here we are able to describe the atmospheric conditions that predominate when deep convective events develop, and when convection does not develop. This allows us to better understand which are the most favorable atmospheric conditions for the occurrence of deep convective events in the Sabana de Bogotá, leading to a simple conceptual model that describes the most relevant fundamental processes.

2. Data and method

To identify deep convective events (DCE) in the Sabana de Bogotá, we use infrared images from GOES-13 (NOAA OSPO, 1994) between 2011 and 2017, which is the operational time period of this satellite. These images have a native resolution of about 4 km and are available every 30 min for this region. For each image, brightness temperature (BT) is averaged in the four square 16 × 16 km boxes shown in Figure 1b. Following Adams et al. (2013), we consider a convective event whenever the average BT inside one or more boxes is below 230 K, and there is a decrease in average BT of at least 50 K in 2 h, within a 3-h window. Such a rapid decrease in BT constitutes a clear signature of the deepening of a convective system. Additionally, we use precipitation data from five automatic rain gauges at different locations throughout the Sabana de Bogotá (see Fig. 1b). We only consider cases in which a precipitation intensity of at least 5 mm in 30 min is observed at least at one of these rain gauges within the 3-h window of each event. To investigate the atmospheric conditions associated with convective events, we use meteorological data from the automatic weather station from the Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM), located at Bogotá's international airport (El Dorado). This station provides precipitation and wind data every 10 min, as well as hourly surface air temperature (2 m) and pressure data (IDEAM, 2019). Non-convective days (NCD) are identified as days in which the average BT in the four boxes remains above 250 K between 12:00 and 18:00 LT (UTC-5), with no precipitation in any of the rain gauges nor in the 3-hourly TRMM precipitation data (Huffman et al., 2007) between 11:30 and 17:30 LT in the area indicated in Fig. 1b.

An important asset of this study is the use of precipitable water vapor (PWV) from one GPS station. Tropospheric delay is estimated based on GPS data collected and processed by the Space Geodesy Research Group of the Colombian Geological Survey at the International GNSS Service (IGS) BOGT station (see Fig. 1b for the station's location) with the precise point positioning (PPP) strategy using Gipsy-X software (Bertiger et al., 2020). From the tropospheric delay, we estimate PWV following Bevis et al. (1992, 1994) using air pressure and temperature data from the nearest IDEAM weather station, located 900 m away at the same elevation. Unfortunately, this automatic weather station has long periods of missing data (close to 38 months between 2011 and 2017), so in order to have PWV data for as many DCE as possible, we use hourly climatological values of pressure and temperature to compute the entire PWV time series. This is possible because pressure and temperature at this low latitude location have a very consistent diurnal cycle with very little variability. In fact, comparing the more precise PWV calculations when actual surface temperature and pressure data are available, with those used here, results show a mean error of only 0.05 mm, with RMSE = 0.4 mm.

Finally, hourly ERA5 (Hersbach et al., 2019; Muñoz, 2019) reanalysis data with a 0.1º (surface air temperature, surface pressure, surface wind velocity) and a 0.25º (PWV, divergence, vertically integrated moisture divergence [MD] and vertically integrated water vapor flux [WVF]) resolution are used to investigate regional atmospheric patterns associated with convective events at the Sabana de Bogotá.

3. Results and discussion

3.1 Temporal distribution of convective events

Based on the GOES-13 infrared images we identify 536 convective events, which take place in 467 days. Since the entire period 2011-2017 has 2556 days, this indicates that convection occurs in ~18% of the days. However, convection events are not equally distributed throughout the year, nor throughout the day. Events are concentrated in the two rainy seasons and are most frequent during the early afternoon (Fig. 2a, b). If we limit ourselves to the months of February-May and October-November (see below for justification), we find 372 events taking place in

323 days, which corresponds to 26% of the days (red bars in Fig. 2a, b). Nevertheless, as mentioned in the previous section, we only consider events that have a clear intense precipitation signal at one or more of the five rain gauges, which reduces the number of events to 128, most likely for two reasons: first, precipitation may have not fallen exactly at these locations, and second, precipitation data from the rain gauges are not always complete. However, notice that the monthly and hourly distribution of these filtered events (Fig. 2c, d) is similar to the entire population of events, as is also shown by the hourly distribution of the events during the months of interest (red bars in Fig. 2b, d). Given that we require good quality in situ meteorological information for this study, we consider these 128 events as a starting point.

Convective events occur mainly between February and May, and between October and November, in the early afternoon hours (Fig. 2). Since we are interested in studying the characteristics of the typical or most frequent convective cases, it is convenient to limit this study to the months where most events occur, which closely match the two rainy seasons at this location. These two seasons (February-May, October-November) may have slightly different large-scale circulation patterns (e.g., Wang and Fu, 2002; Vera et al., 2006), but locally are very similar. In fact, the local wind and moisture fields of these two seasons are very similar (not shown here). This allows us to consider both seasons together, which on the other hand is more convenient to have a larger sample of events. Therefore, we only consider cases during the aforementioned months, which leaves us with 105 deep convective events (red bars in Fig. 2c, d) distributed throughout 102 days. On the other hand, 61 of these events have complete meteorological data available at El Dorado, which is the station we use as a reference to study the atmospheric conditions associated with deep convective events. 43 of these (70%) are associated with westerly winds (wind direction between 202.5º and 337.5º), consistent with observations reported by Montoya and Eslava (2000), who also found that the occurrence of intense storms coincides with an early afternoon change in wind direction from the predominant easterlies to westerlies. Based on GOES-13 infrared images and precipitation data from TRMM, we also identify 1284 NCD, out of which 508 correspond to the months February-May and October-November. Of these days, 152 have wind data, with 39 showing westerly winds between 12:00 and 14:00 LT, which we use in order to compare to DCE with early afternoon westerlies. Table I

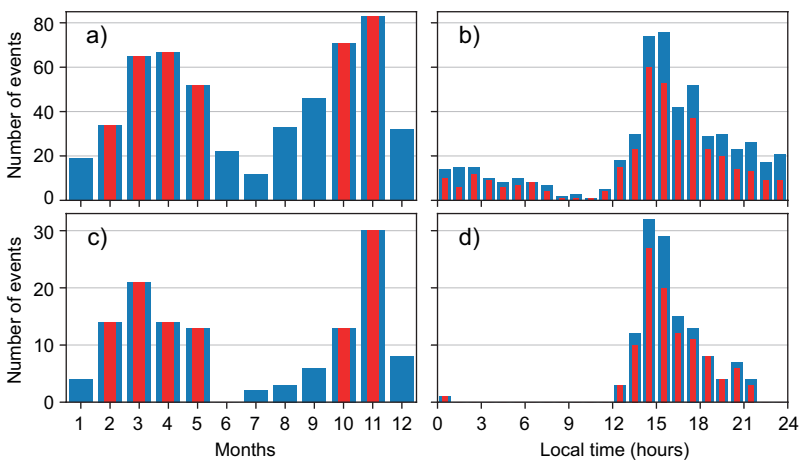

Fig. 2. (a, c) Monthly and (b, d) hourly distribution of deep convective events in the Sabana de Bogotá as identified from GOES-13 infrared images (a, b), and considering only those with a clear precipitation signal in at least one of the rain gauges (c, d). In the four panels, red bars correspond to the number of cases only during the months of interest (February-May and October-November).

Table I. Summary of the relevant numbers and percentages related to the convective and non-convective identified days. Percentages are relative to the periods indicated in the first column, except otherwise stated.

Periods	DCE days	NCD	DCE days with in situ intense precipitation	DCE days with wind data	DCE with westerlies	NCD with wind data	NCD with westerlies
Entire period 2011-2017 (2556 days)	467 (18%)	1284 (50%)	125 (27% of DCE days)	—	—	—	—
Feb-May, Oct-Nov (1262 days)	323 (26%)	508 (40%)	102 (32% of DCE days)	61 (60% of DCE days with precip)	43 (70% of DCE with wind data)	152 (30% of NCD)	39 (26% of NCD with wind data)

DCE: deep convective events; NCD: non-convective days.

summarizes all these numbers, together with the corresponding percentages.

These results indicate that convection at this location is mostly diurnally forced during the two rainy seasons in this region, which are primarily a consequence of the latitudinal migration of the Intertropical Convergence Zone (ITCZ), and thus consistent with previous studies (e.g., Montoya and Eslava, 2000; Aragón, 2015). The fact that convection peaks in the early afternoon (14:00-15:00 LT) suggests that the diurnal forcing near the surface is strong, and events have a short developing timescale in comparison with most tropical continental locations where convection peaks later in the afternoon (e.g., Bowman et al., 2005; Liu and Zipser, 2008).

3.2 Behavior of atmospheric variables

In order to identify the atmospheric conditions associated with deep convective events, we construct hourly composites of air temperature, atmospheric pressure, wind direction, wind speed, precipitable water vapor, and for the vertical integration of water vapor flux and moisture divergence. We do this for three sets of cases: AVG (which corresponds to the entire months of February-May and October-November), DCE (which corresponds to deep convective events during these same months), and NCD (which corresponds to non-convective days). We do not find any significant signal in atmospheric pressure related to the convective events. This is not surprising, since at low latitudes the slight pressure perturbations related to deep convective events cannot be detected with hourly data (e.g., Holton, 2004). Therefore, we do not show atmospheric pressure composites here.

Wind direction is predominantly from the northeast during the months of interest (Fig. 4a) but tends to become westerly most of the days between 14:00 and 17:00 LT (Figs. 3 and 4b). Notice that in the early afternoon, winds from the west (including northwesterlies and southwesterlies) are present in 44% of the available data (Fig. 4b). Furthermore, if we count days with afternoon westerlies as those in which the hourly mean of the wind direction is from the west at least during one hour between 14:00 and 17:00 LT (not shown here), we find that this adds up

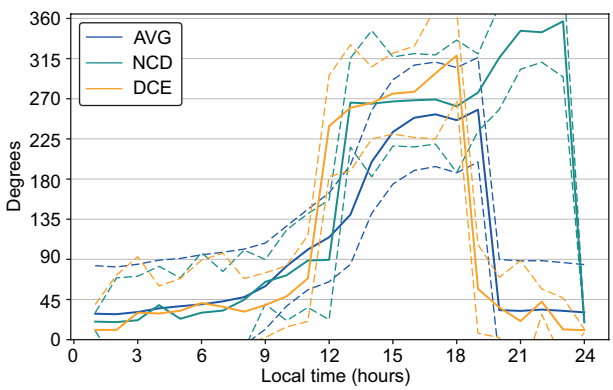

Fig. 3. Wind direction composites in El Dorado. AVG values are in blue, DCE in orange, and NCD in green. Dotted lines represent one standard deviation.

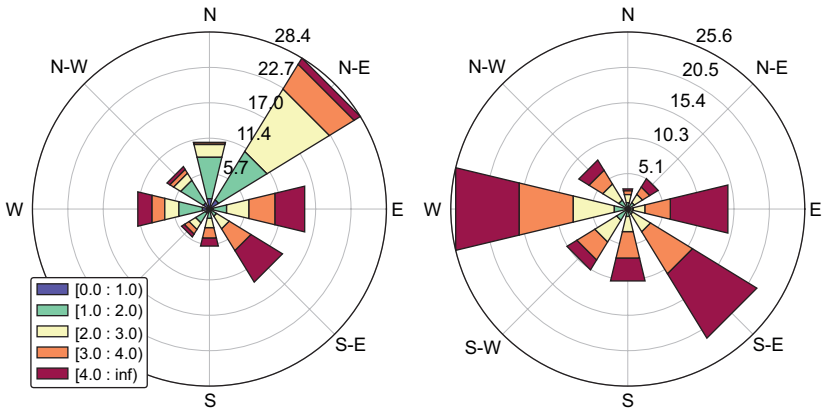

Fig. 4. Average wind direction for El Dorado corresponding to (a) all day and (b) the afternoon hours (14:00 to 17:00 LT), in both cases for February to May and October to November for 2011 to 2017. Colors indicate wind speed (see legend), and numbers at different radii indicate the corresponding percentage of occurrence.

to 78% of the days. On the other hand, the frequency of convective days during the months of interest is 26%, or even less if we only count cases with intense precipitation recorded at any of the five rain gauges (Table I). Therefore, though westerlies are usually present during convective events (in about 70% of cases), most days with westerlies do not necessarily result in DCE. On the other hand, during days with DCE wind direction does turn earlier to the west than in AVG, but in NCD the wind direction is similar to DCE. This indicates that wind direction is probably important but not essential for the generation of convection, and possibly not a reliable predictor.

On the other hand, surface air temperature reaches a higher maximum in DCE compared to NCD and AVG around noon, but decreases faster and to lower values during the afternoon, probably due to evaporative cooling (e.g., Park et al., 2017) during and after the convective event (Fig. 5a, b). This is consistent with convection driven by surface heating (e.g., Emanuel 1994). Figure 5c, d show that wind speed has a particular behavior in DCE compared to average conditions. Before noon on deep convective days, wind speed is considerably slower than in AVG and NCD (almost one standard deviation). This could be important because a slower wind speed can facilitate water vapor convergence and a better organization of convective clouds (e.g., Tian and Parker, 2002; Kirshbaum and Wang, 2014; Nugent et al., 2014). In days with DCE, wind speed has a steeper increase than in AVG from 11:00 to 14:00 LT. Afterward, wind speed decreases very rapidly between 14:00 and 15:00 LT, such that it again becomes slower in the late afternoon and at night compared to average conditions. Similar behavior has been reported by Adams et al. (2013), with a rapid increase in wind speed followed by a steeper decrease before and after deep convective events in the Amazon. This feature is clearly depicted with the wind speed tendency in Figure 5d and seems to be a prominent feature of DCE at this location.

One parameter that seems crucial for DCE is PWV (Fig. 5e, f) (Mora-Páez et al., 2020). Throughout the entire day of deep convective events, PWV is significantly higher than during NCD or the average conditions. It also has a steeper increase between 12:00-16:00 LT, consistent with the time of the day when most convective events develop. The behavior of PWV is very similar to that reported by Adams et al. (2011, 2013, 2015), but differs in magnitude (Bogotá has lower values than Manaus mostly due to its high elevation). In Bogotá, PWV has a slow increase during 2 h (9:00 to 11:00 LT) followed by a steeper increase for about 3 h, so the development of convection seems to occur faster in Bogotá (i.e., 3 h) than in Manaus (i.e., 8 h). Another important difference is that in Bogotá, PWV does not decrease rapidly after the precipitation peaks, as it does in

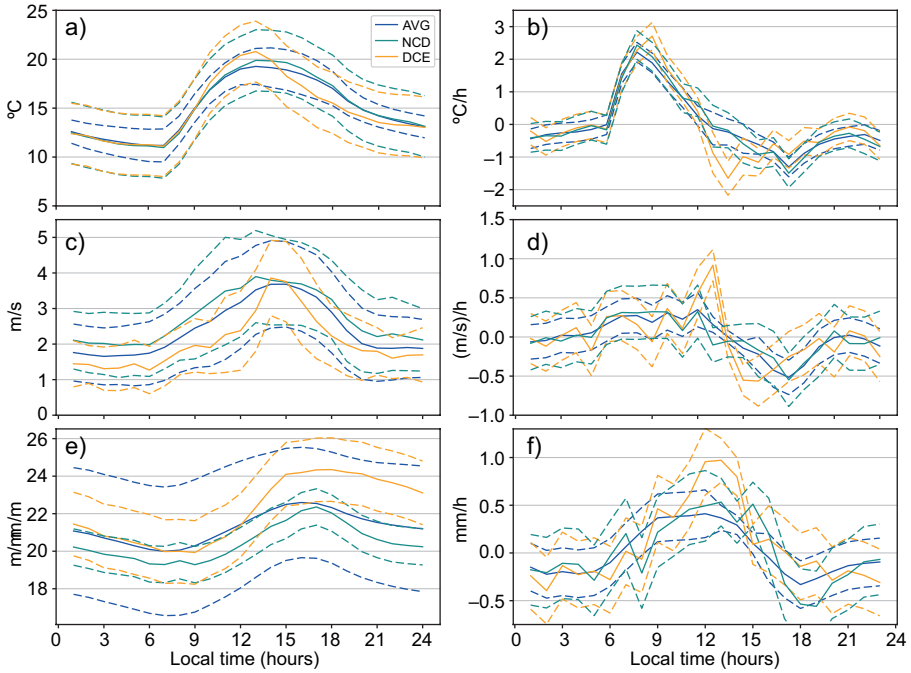

Fig. 5. Composites of surface air temperature, wind speed, and PWV (a, c, e) and their corresponding tendencies (b, d, f) in El Dorado. AVG values are in blue, DCE in orange, and NCD in green. Dotted lines represent one standard deviation. (PWV: precipitable water vapor; AVG: February-May and October-November; DCE: deep convective events; NCD: non-convective days.)

Manaus (Adams et al., 2013). This could be because convective events in Manaus produce more intense precipitation than those at the Sabana de Bogotá.

3.3 Regional behavior of atmospheric variables

We construct composites using reanalysis data from ERA5 to investigate the evolution of regional patterns of the relevant atmospheric variables associated with deep convective events in the Sabana de Bogotá. In general, ERA5 composites are consistent with the results from the individual surface stations described in the previous section. For example, temperature increases to higher values in DCE and then decreases to lower values compared to the overall average. Furthermore, the spatial distribution of surface pressure in the afternoon according to ERA5 (not shown here) shows a dipole consistent with the westerly turn of the winds observed at the surface stations. However, the weaker wind speed detected by the surface stations is not visible in the ERA5 data (not shown here). In fact, ERA5 data suggest higher wind speeds during DCE compared to the AVG. This implies that the decrease in wind speed may be a local effect, which the coarser reanalysis data does not capture. On the other hand, composites of ERA5 divergence at 250 hPa (not shown here) are consistent with the occurrence of convection relative to average conditions (higher than average divergence values at 250 hPa are found for DCE). However, 250 hPa divergence values during NCD are higher than during convective days, suggesting that the dynamical forcing may be similar, or even stronger, during non-convective days than during deep convective days. Thus, other processes may be more determining, such as, for example, moisture content and dynamics. Despite this, in general, we argue that the ERA5 data captures the main signals associated with the occurrence of DCE, so we can use it to further investigate the regional patterns that cannot be otherwise investigated with the available surface station data.

Within ERA5 data, PWV exhibits an important behavior related to DCEs (Fig. 6): throughout the afternoon, PWV has higher values in DCE in comparison to NCD and to AVG, but along the Magdalena

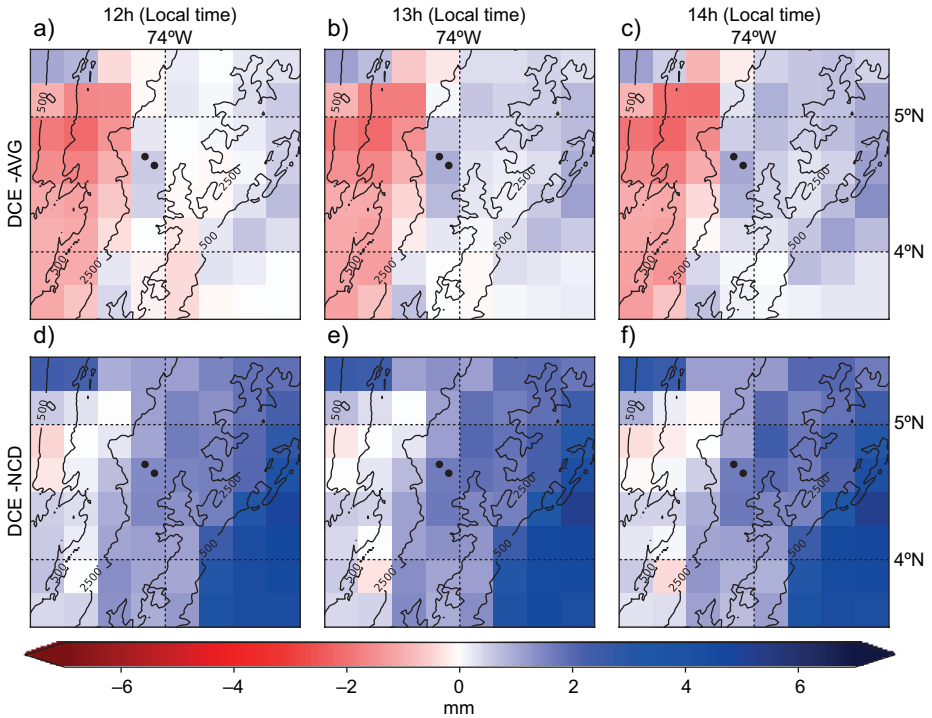

Fig. 6. (a, b, c) composites of DCE-AVG and (d, e, f) DCE-NCD anomalies of precipitable water vapor from ERA-5 reanalysis data at 12:00 (a, d), 13:00 (b, e), and 14:00 (c, f) LT. The dots indicate the location of El Dorado weather station (left) and the GNSS station (right), and contour lines indicate elevations of 500 and 2500 m. (DCE: deep convective events; AVG: February-May and October-November; NCD: non-convective days.)

river valley it clearly has lower values. This, together with higher values of moisture divergence (MD) (Fig. 7) in Bogotá suggests that the high moisture content associated with DCEs is related to local convergence rather than transported from the Magdalena valley, as hypothesized by Montoya and Eslava (2000). Furthermore, the vertically integrated water vapor flux (WVF) from ERA5 (Fig. 8) indicates that moisture is normally advected from the east. During DCE, this flux is lower than in AVG and NCD, suggesting that local convergence plays a more important role. However, higher spatial resolution observations would be needed to confirm this, since reanalysis data is not expected to capture smaller scale circulations, especially with such complex terrain in the deep tropics. All these features suggest that although dynamical processes linked to westerly winds seem necessary, sufficient moisture content is crucial. This also indicates that PWV may be used as an indicator for the formation of deep convection, as shown in other studies (e.g. Crook, 1996; Sherwood et al., 2004; Weckwerth and Parsons, 2006; Keil et al., 2008; Holloway and Neelin, 2009, 2010; Benevides et al., 2015; Liang et al., 2015).

4. Conclusions

Here we study the characteristics of deep convection events (DCE) in the Sabana de Bogotá between 2011 and 2017 to better understand this phenomenon. Since this is a low-latitude location, convection is the main precipitation-producing process, and it is also the main cause for extreme rainfall events that may affect this densely populated area. Furthermore, the region's high elevation (~2550 m) and surrounding topography create specific conditions that determine the local dynamics of convection and make it difficult to study without in situ observations of high temporal resolution. A better understanding of the main convective drivers in this region is extremely valuable to improve short-term weather forecasting, to provide better input for local risk management, and to improve our understanding of tropical convective processes in mountainous areas.

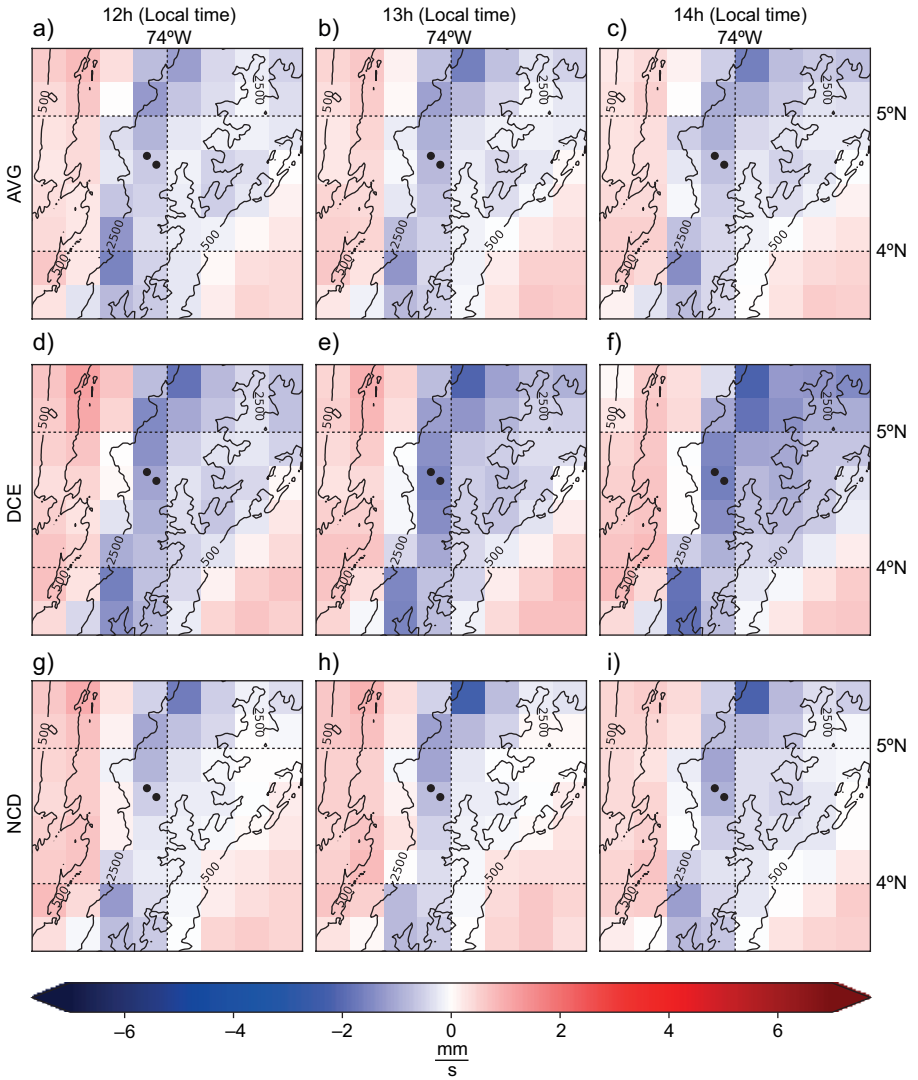

Fig. 7. Vertically integrated moisture divergence (MD) composites of (a, b, c) AVG, (d, e, f) DCE, and (g, h, i) NCD from ERA-5 reanalysis data at 12:00 (a, d, g), 13:00 (b, e, h), and 14:00 (c, f, i) LT. The dots indicate the location of El Dorado weather station (left) and the GNSS station (right), and contour lines indicate elevations of 500 and 2500 m. (AVG: February-May and October-November; DCE: deep convective events; NCD: non-convective days.)

Using GOES-13 infrared images we identify 430 DCE, of which 128 caused precipitation at one or more of the five in situ stations we considered. These events occur mostly during the two rainy seasons and between 14:00 and 16:00 LT, so we limit our study to these periods of time. In order to identify the atmospheric patterns associated with DCE, we carry out a composite analysis of in situ measured atmospheric variables (surface air temperature, atmospheric pressure, and wind velocity), precipitable water vapor (PWV) from a GPS station, and ERA5 reanalysis data. We also construct composites for non-convective days (NCD) and for multi-year hourly mean conditions (AVG) to compare them.

Previous studies of convective activity in this region (e.g., Montoya and Eslava, 2000; Aragón, 2015)

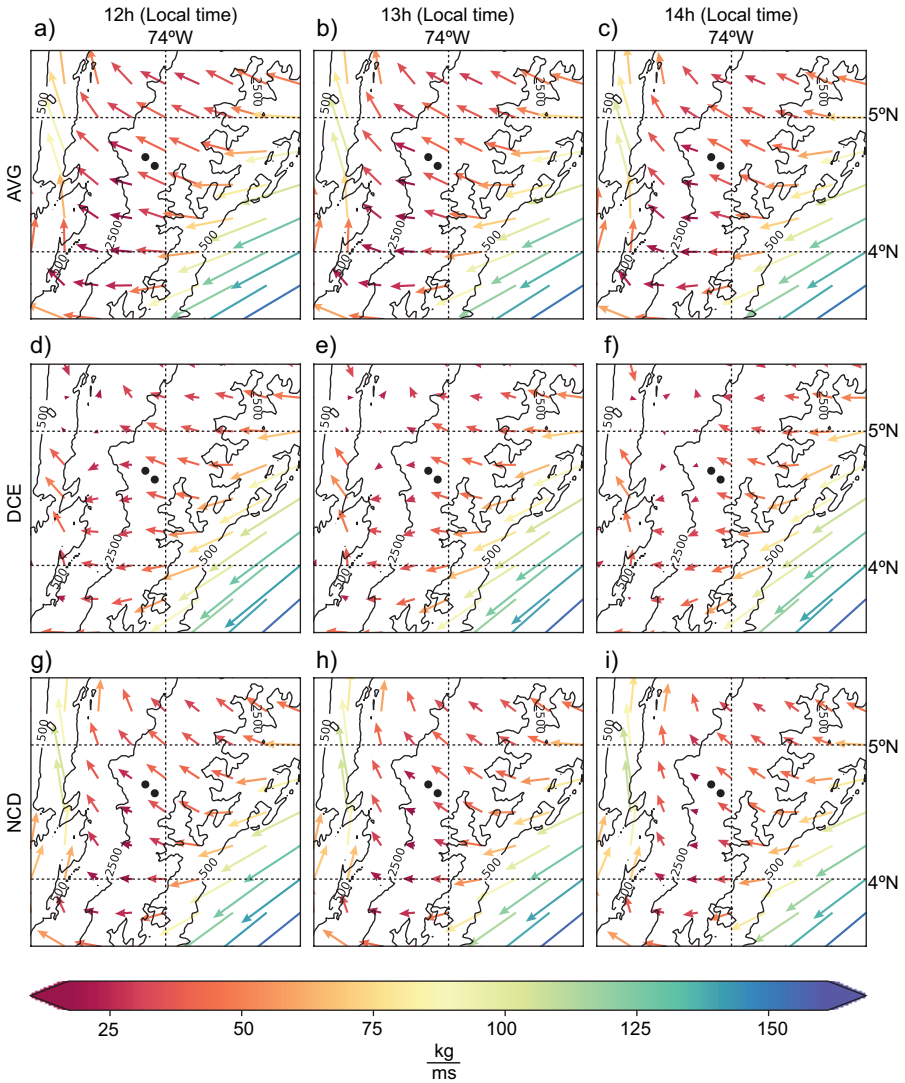

Fig. 8. Vertically integrated water vapor flux (WVF) composites for (a, b, c) AVG, (d, e, f) DCE, and (g, h, i) NCD from ERA-5 reanalysis data at 12:00 (a, d, g), 13:00 (b, e, h), and 14:00 (c, f, i) LT. The dots indicate the location of El Dorado weather station (left) and the GNSS station (right), and contour lines indicate elevations of 500 and 2500 m. (AVG: February-May and October-November; DCE: deep convective events; NCD: non-convective days.)

suggest that the main driver for afternoon storms is a westerly turn of the otherwise predominant easterlies and the possible associated moisture transport from the Magdalena valley that lies west of Bogotá. Our results suggest that westerly winds are an important ingredient of convective events (they are present in 70% of them), consistent with the convergence they generate in the Sabana de Bogotá. However, westerly winds appear in 78% of the afternoons in the two rainy seasons, whereas deep convective events occur in only 26% or less of these days. This implies that although westerly winds and their convergence are important for deep convection to take place, they are not a good indicator of convective initiation. Furthermore, reanalysis data (PWV, and vertically integrated moisture divergence and water vapor flux) suggests that: (a) moisture content over the Magdalena river valley is lower than average during convective days, (b) water

vapor is mainly advected from the east, and (c) local water vapor convergence is more important than advection from nearby regions. Although reanalysis data must be carefully interpreted at these scales and with such complex topography, this does suggest that local processes may be more important than any moisture advection from the Magdalena river valley. However, further research will be necessary to confirm this and determine with precision the relevant moisture sources.

Our results point towards three variables in terms of their relevance for the development of deep convection: wind speed, surface air temperature, and PWV. Wind speed is weaker in the morning of convective days, allowing for convection to organize. After this, a rapid increase in wind speed occurs from 12:00 to 15:00 LT, while surface air temperature also reaches higher than average values. Meanwhile, PWV increases rapidly from 12:00 to 15:00 LT reaching its maximum at the time when most convective events occur. After this, temperature and wind speed decrease to lower-than-average values, while PWV decreases slowly and remains with above-average values. Therefore, we propose that the main short-term precursors for a deep convective event during the rainy seasons in this region are: (a) a weaker than average wind speed throughout the morning, with (b) higher than average surface air temperature towards noon, followed by (c) a rapid increase in PWV (and in wind speed) that leads to the convective event. Based on our PWV and moisture divergence (MD) data, we identify a water vapor convergence time-scale of about 3 h that precedes the deep convective events. Adams et al. (2013) found two water vapor convergence time-scales for deep convection in Manaus, of 8 and 4 h. We hypothesize that the complex topography and high elevation of the Sabana de Bogotá make larger and longer-lasting convection events very difficult to develop due to the multiple perturbations in the dynamic and thermodynamic fields, and only smaller and shorter-lived systems occur.

We find that PWV is a very useful indicator of developing deep convective events in this region. However, for this variable to be useful, it must be measured with high temporal resolution, as is the case with GPS stations. This motivates further efforts for implementing a comprehensive atmospheric water vapor monitoring program using the existing GPS station network in Colombia (Mora-Páez et al., 2020). For example, using this network it would be possible to investigate the possible sources of moisture that feed the deep convective events described here. Furthermore, this network can be used to improve short-term forecasting of extreme events throughout Colombia, and in general to study the dynamics of tropical convection under a wide variety of tropical conditions.

Acknowledgments

We thank two anonymous reviewers for their constructive comments, since they were very important to improve our results and be more precise in the conclusions of this manuscript. The authors also thank Alejandro Jaramillo and José Franklyn Ruiz for providing comments and suggestions during the early stages of this study. The IDEAM was also an important source of information for this research since they provided in-situ data used in this manuscript. Finally, the authors would like to thank the Servicio Geológico Colombiano (Colombian Geological Survey) for sharing their GPS tropospheric delay estimations through GeoRED; this was essential information for this study.

References

Adams DK, Fernandes RMS, Kursinski ER, Maia JM, Sapucci LF, Machado LAT, Vitollero I, Galera-Monico JF, Holub KL, Filizola N, Bennett RA. 2011. A dense GNSS meteorological network for observing deep convection in the Amazon. Atmospheric Science Letters 12: 207-212. https://doi.org/10.1002/asl.312

Adams DK, Seth I, Kirk, Dulcineide S. 2013. GNSS observations of deep convective time scales in the Amazon. Geophysical Research Letters 40: 2818-2823. https://doi.org/10.1002/grl.50573

Adams DK, Fernandes RMS, Holub KL, Gutman SI, Barbosa HMJ, Machado LAT, Calheiros AJP, Bennett RA, Kursinski ER, Sapucci LF, DeMets C, Chagas GFB, Arellano A, Filizola N, Amorim-Rocha AA, Araujo-Silva R, Assuncao LMF, Cirino GG, Pauliquevis T, Portela BTT, Sa A, Sousa JM, Tanaka LMS. 2015. The Amazon dense GNSS meteorological network: A new approach for examining water vapor and deep convection interactions in the tropics. Bulletin of the American Meteorological Society 96: 2151-2165. https://doi.org/10.1175/BAMS-D-13-00171.1

Aragón G. 2015. Análisis de las condiciones meteorológicas asociadas con la presencia de sistemas convectivos en el aeropuerto El Dorado de Bogotá DC. M.Sc. thesis. Universidad Nacional de Colombia. Available at: https://repositorio.unal.edu.co/handle/unal/59169. (accessed on January 30, 2020).

Benevides P, Catalao J, Miranda PMA. 2015. On the inclusion of GPS precipitable water vapour in the nowcasting of rainfall. Natural Hazards and Earth System Sciences 15: 2605-2616. https://doi.org/10.5194/nhess-15-2605-2015

Bertiger W, Bar-Sever Y, Dorsey A, Haines B, Harvey N, Hemberger D, Heflin M, Lu W, Miller M, Moore AW, Murphy D, Ries P, Romans L, Sibois A, Sibthorpe A, Szilagyi, Valisneri M, Willis P. 2020. GipsyX/RTGx, a new tool set for space geodetic operations and research. Advances in Space Research 66: 469-489. https://doi.org/10.1016/j.asr.2020.04.015

Betts A, Jakob C, 2002. Study of diurnal convective precipitation over Amazonia using single column mode. Journal of Geophysical Research 107: ACL25-1-ACL25-13. https://doi.org/10.1029/2002JD002264

Bevis M, Businger S, Herring TA, Rocken C, Anthes RA, Ware RH. 1992. GPS meteorology: Remote sensing of atmospheric water vapor using the global positioning system. Journal of Geophysical Research–Atmospheres 97: 15787-15801. https://doi.org/10.1029/92JD01517

Bevis M, Businger S, Chiswell S, Herring TA, Anthes RA, Rocken C, Ware RH. 1994. GPS meteorology: Mapping zenith wet delays onto precipitable water. Journal of Applied Meteorology 33: 379-386. https://doi.org/10.1175/1520-0450(1994)033<0379:GMMZWD>2.0.CO;2

Bowman KP, Collier JC, North GR, Wu Q, Ha E, Hardin J. 2005. Diurnal cycle of tropical precipitation in Tropical Rainfall Measuring Mission (TRMM) satellite and ocean buoy rain gauge data. Journal of Geophysical Research 110: D21104. https://doi.org/10.1029/2005JD005763

Bretherton CS, Peters ME, Back LE. 2004. Relationships between water vapor path and precipitation over the tropical oceans. Journal of Climate 17: 1517-1528. https://doi.org/10.1175/1520-0442(2004)017<1517:rbwvpa>2.0.co;2

Crook NA. 1996. Sensitivity of moist convection forced by boundary layer processes to low-level thermodynamic fields. Monthly Weather Review 8: 1767-1785. https://doi.org/10.1175/1520-0493(1996)124<1767:SOMCFB>2.0.CO;2

Emanuel KA. 1994. Atmospheric convection. Oxford University Press, New York.

Fairman JG, Nair US, Christopher SA, Mölg T. 2011. Land-use change impacts on regional climate over Kilimanjaro. Journal of Geophysical Research 116: D03110. https://doi.org/10.1029/2010JD014712

Freitag BM, Nair US, Niyogi D. 2018. Urban modification of convection and rainfall in complex terrain. Geophysical Research Letters 45: 2507-2515. https://doi.org/10.1002/2017GL076834

Hersbach H, Bell B, Berrisford P, Biavati G, Horányi A, Muñoz Sabater J, Nicolas J, Peubey C, Radu R, Rozum I, Schepers D, Simmons A, Soci C, Dee D, Thépaut JN. 2019. ERA5 hourly data on pressure levels from 1979 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). https://doi.org/10.24381/cds.bd0915c6

Holloway CE, Neelin JD. 2009. Moisture vertical structure, column water vapor, and tropical deep convection. Journal of the Atmospheric Sciences 66: 1665-1683. https://doi.org/10.1175/2008JAS2806.1

Holloway CE, Neelin JD. 2010. Temporal relations of column water vapor and tropical precipitation. Journal of the Atmospheric Sciences 67: 1091-1105. https://doi.org/10.1175/2009jas3284.1

Holton J. 2004. An introduction to dynamic meteorology. Elsevier Academic Press, USA.

Huffman GJ, Bolvin DT, Nelkin EJ, Wolff DB, Adler RF, Gu G, Hong Y, Bowman KP, Stocker EF. 2007. The TRMM multisatellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at new scales. Journal of Hydrometeorology 8: 38-55. https://doi.org/10.1175/JHM560.1

IDEAM. 2019. Atlas interactivo del IDEAM. Instituto de Hidrología, Meteorología y Estudios Ambientales, Colombia. Available at: http://atlas.ideam.gov.co/presentacion/

Keil C, Rpnack A, Craig GC, Schumann U. 2008. Sensitivity of quantitative precipitation forecast to height dependent changes in humidity. Geophysical Research Letters 35: L09812. https://doi.org/10.1029/2008GL033657

Kirshbaum DJ, Wang CC. 2014. Boundary layer updrafts driven by airflow over heated terrain. Journal of the Atmospheric Sciences 71: 1425-1442. https://doi.org/10.1175/JAS-D-13-0287.1

Liang H, Cao Y, Wan X, Xu Z, Wang H, Hu H. 2015. Meteorological applications of precipitable water vapor measurements retrieved by the national GNSS network of China. Geodesy and Geodynamics 6: 135-142. https://doi.org/10.1016/j.geog.2015.03.001

Liu C, Zipser EJ. 2008. Diurnal cycles of precipitation, clouds, and lightning in the tropics from 9 years of TRMM observations. Geophysical Research Letters 35: L04819. https://doi.org/10.1029/2007GL032437

Montoya G, Eslava J. 2000. Análisis de las condiciones que favorecen el desarrollo de la actividad convectiva de las tormentas eléctricas en la Sabana de Bogotá. Earth Sciences Research Journal 4: 23-30.

Mora-Páez H, Corchuelo Y, Gutiérrez N, Moreno R, Álvarez E, Díaz F. 2020. Coordenadas estaciones geodésicas GeoRED 1.1.2019-ITRF2008. PROD.00740_2020. Servicio Geológico Colombiano, Bogotá, 38 pp.

Muñoz Sabater J. 2019. ERA5-Land hourly data from 1981 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). https://doi.org/10.24381/cds.e2161bac

NOAA OSPO. 1994. NOAA Geostationary Operational Environmental Satellite (GOES) imager data, GVAR IMG band 4. National Oceanic and Atmospheric Administration, Office of Satellite and Product Operations.Available at: https://www.star.nesdis.noaa.gov/GOES/index.php (accessed on June 20, 2018).

Nugent AD, Smith RB, Minder JR. 2014. Wind Speed Control of Tropical Orographic Convection. Journal of the Atmospheric Sciences 71: 2695-2712. https://doi.org/10.1175/JAS-D-13-0399.1

Park SB, Heus T, Gentine P. 2017. Role of convective mixing and evaporative cooling in shallow convection. Journal of Geophysical Research–Atmospheres 122: 5351-5363. https://doi.org/10.1002/2017JD026466

Roberts RD, Fabry F, Kennedy PC, Nelson E, Wilson JW. Rehak N, Fritz J, Chandrasekar V, Braun J, Sun J, Ellis S, Reising S, Crum T, Mooney L, Palmer R, Weckwerth T, Padmanabhan S. 2008. REFRACTT 2006: Real-time retrieval of high resolution, low-level moisture fields from operational NEXRAD and research radars. Bulletin of the American Meteorological Society 89: 1535-1548. https://doi.org/10.1175/2008BAMS2412.1

Sherwood SC, Minnis P, McGill M. 2004. Deep convective cloud-top heights and their thermodynamic control during CRYSTAL-FACE. Geophysical Research Letters 109: D20119. https://doi.org/10.1029/2004JD004811

Sherwood SC, Roca R, Weckwerth TM, Andronova NG. 2010. Tropospheric water vapor, convection, and climate. Reviews of Geophysics 48: RG2001. https://doi.org/10.1029/2009RG000301

Smith JA, Baeck ML, Ntelekos AA, Villarini G, Steiner M. 2011. Extreme rainfall and flooding from orographic thunderstorms in the central Appalachians. Water Resources Research 47: W04514. https://doi.org/10.1029/2010WR010190

Sundqvist H. 1978. Parameterization scheme for non-convective condensation including prediction of clouds. Quarterly Journal of the Royal Meteorological Society 104: 677-690. https://doi.org/10.1002/qj.49710444110

Tian WS, Parker DJ. 2002. Two-dimensional simulation of orographic effects on mesoscale boundary-layer convection. Quarterly Journal of the Royal Meteorological Society 128: 1929-1952. https://doi.org/10.1256/003590002320603476

Tompkins AM. 2001. Organization of tropical convection in low vertical wind shears: The role of water vapor. Journal of the Atmospheric Sciences 58: 529-545. https://doi.org/10.1175/1520-0469(2001)058<0529:OOTCIL>2.0.CO;2

TRMM. 2011. TRMM (TMPA) rainfall estimate L3 3 hour 0.25-degree x 0.25 degree V7. Tropical Rainfall Measuring Mission Project, Goddard Earth Sciences Data and Information Services Center, Greenbelt, MD. Available at: https://doi.org/10.5067/TRMM/TMPA/3H/7 (accessed on June 20, 2018).

Vera CC, Silvestri G, Liebmann B, González P. 2006. Climate change scenarios for seasonal precipitation in South America from IPCC-AR4 models. Geophysical Research Letters 33: L13707. https://doi.org/10.1029/2006GL025759

Wang HH, Fu R. 2002. Cross-equatorial flow and seasonal cycle of precipitation over South America. Journal of Climate 15: 1591-1608. https://doi.org/10.1175/1520-0442(2002)015<1591:CEFASC>2.0.CO;2

Weckwerth T, Parsons DB. 2006. A review of convection initiation and motivation for IHOP_2002. Bulletin of the American Meteorological Society 134: 5-22. https://doi.org/10.1175/MWR3067.1

Zeng X. 1998. The relationship among precipitation, cloud-top temperature, and precipitable water over the tropics. Journal of Climate 12: 2503-2514. https://doi.org/10.1175/1520-0442(1999)012<2503:TRAPCT>2.0.CO;2

Atmósfera 36(2), 239-252 (2023)
https://doi.org/10.20937/ATM.53038

CONTENTS

Regional flow climatology for central Mexico (Querétaro): A first case study

Daniel Atreyu ROZANES-VALENZUELA[1], Adolfo Vicente MAGALDI[2]* and Dara SALCEDO[1]

[1]*Unidad Multidisciplinaria de Docencia e Investigación Juriquilla, Facultad de Ciencias, Universidad Nacional Autónoma de México, Blvd. Juriquilla 3001, 76230 Querétaro, Querétaro, México.*
[2]*ENES-Juriquilla, Universidad Nacional Autónoma de México, Blvd. Juriquilla 3001, 76230 Querétaro, Querétaro, México.*
*Corresponding author: adolfo.magaldi@gmail.com

Received: February 16, 2021; accepted: August 4, 2021

RESUMEN

Se estableció una climatología de flujo para el Área Metropolitana de Querétaro (MAQ), en el centro de México, mediante el análisis de cuatro años (2014-2017) de retrotrayectorias generadas mediante el modelo HYSPLIT. Se encontraron dos regímenes de flujo: uno de junio a septiembre (régimen lluvioso) y el otro de diciembre a mayo (régimen seco). Octubre y noviembre se consideraron meses de transición. Los flujos del noreste estuvieron presentes durante todo el año; en contraste, las trayectorias desde el suroeste fueron mucho menos frecuentes y se observaron principalmente durante el régimen seco. Un análisis de los campos de viento de la base de datos del North American Regional Reanalysis (NARR) durante un periodo de tiempo más largo (1979-2019) sugiere que estos resultados son representativos de las condiciones promedio de la atmósfera en el sitio de estudio. Algunas de las trayectorias observadas hacia el noreste se originan dentro de una región desértica del estado de Querétaro, donde se ubican varias minas de piedra caliza. Durante el régimen seco y los meses de transición, algunos clústeres se originan en el área industrial de Guanajuato, que incluye la refinería de Salamanca. Dado que el transporte aéreo de contaminantes sigue estos caminos, este análisis podría ser útil para identificar fuentes regionales que afecten al MAQ y posiblemente aumenten su carga de contaminación del aire. De hecho, la variabilidad de las concentraciones de contaminantes criterio coincidió con los regímenes de flujo descritos anteriormente.

ABSTRACT

A flow climatology was established for the Metropolitan Area of Querétaro (MAQ) in central Mexico, by analyzing four years (2014-2017) of back-trajectories generated using the HYSPLIT model. Two flow regimes were found: one from June until September (rainy regime); the other from December to May (dry regime). October and November were considered transition months. Northeasterly flows were present throughout the year; in contrast, trajectories from the southwest were much less frequent and observed mainly during the dry regime. An analysis of the wind fields from the North American Regional Reanalysis (NARR) database for a longer period of time (1979-2019) suggests that these results are representative of the average conditions of the atmosphere at the study site. Some of the northeasterly trajectories observed originate within a dessert region of the state of Querétaro, where several limestone mines are located. During the dry regime and transition months some clusters originate at the industrial area in Guanajuato, which includes the Salamanca refinery. As air transport of pollutants follow these paths, this analysis could be useful for identifying regional sources that affect the MAQ and possibly increase its air pollution load. In fact, the variability of criteria pollutants concentrations matched the flow regimes described above.

Keywords: HYSPLIT, back-trajectories, cluster analysis, transport pathways, wind regimes.

1. Introduction

Inefficient use of fuels and energy processes in growing urban areas are associated with poor air quality, which has important implications on health and climate change at local, regional, and global scales. The region in central Mexico known as El Bajío, with a population of approximately 20 million, is a major center of economic activity that includes large, industrialized cities such as Guadalajara, León, Querétaro, and Aguascalientes, as well as the Salamanca refinery. Manufacturing and transportation are probably the most important anthropogenic activities with emissions to the atmosphere in this region (INEGI, 2015). Air quality in the main cities of El Bajío is often reported as bad; however, there is a lack of understanding of the origin (emissions and transport) of air pollutants in this region despite its large population and economic relevance (INECC, 2019). Without such knowledge, policy makers cannot develop environmental and economic policies leading to a more orderly planning and sustainable growth, and to improvements in air quality, protection of public health, and mitigation of the effects of climate change.

The Metropolitan Area of Querétaro (MAQ) located 200 km north of Mexico City on the southeast border of El Bajío, is one of the fastest growing urban areas in Mexico, with a population of 1.3 million in 2015. In addition to local industry and traffic, the MAQ is surrounded by important area sources such as the industrial corridor León-Irapuato-Salamanca, where the Salamanca refinery is located. Little is understood about the interactions between emissions, air quality, and climate in the MAQ and El Bajío, and it is unclear how regional sources of emissions might be affecting the concentrations of pollutants measured at the MAQ in contrast to local human activities.

A flow climatology identifies the main pathways of atmospheric transport, based on the classification of multiple simulated Lagrangian back-trajectories arriving at a certain place of interest (Katsoulis, 1999; Markou and Kassomenos, 2010; Stein et al., 2016), and it is valuable for a regional source contribution assessment because it can show the underlying wind patterns that drive atmospheric transport in a given region and may point to source-receptor relationships for pollutants.

In this paper we analyze and classify four years (2014-2017) of back-trajectories to the MAQ, generated using the Hybrid Single Particle Lagrangian Integrated Trajectory (HYSPLIT) model (Stein et al., 2016), in order to generate a flow climatology for the region. Climatological surface data was used to validate the back-trajectory classification, and to identify a connection between the seasonality of flow regimes and variables such as precipitation and temperature. The North American Regional Reanalysis' (NARR) derived wind fields in a larger time frame (1979-2019) were used to validate the representativeness of the proposed flow climatology.

2. Methodology

2.1 Study region

The MAQ (Fig. 1) is the capital city of the state of Querétaro, which is surrounded by the states of

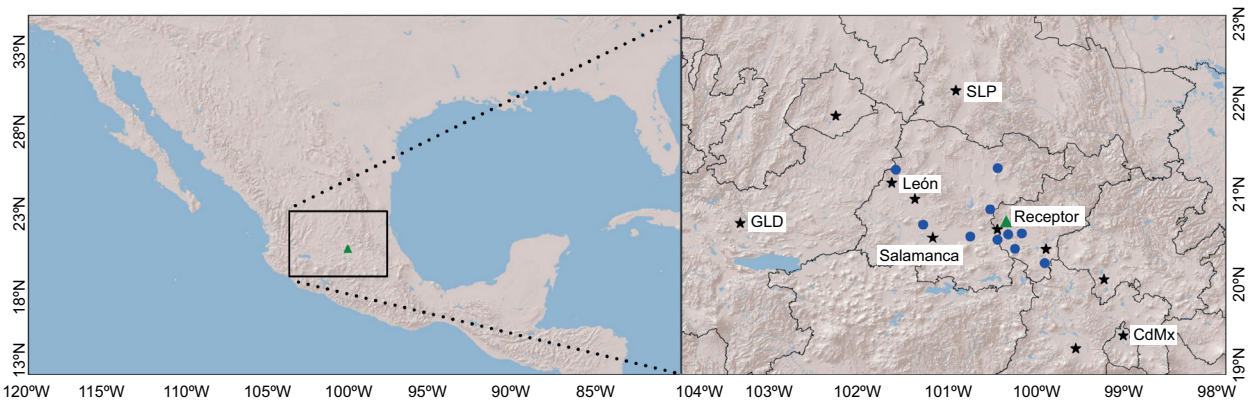

Fig 1. Study region. The green triangle represents the receptor site; black stars represent main urban areas; blue circles represent the climatological stations from the Servicio Meteorológico Nacional (National Weather Service).

Guanajuato, Hidalgo, San Luis Potosí, Michoacán, and the State of Mexico. Querétaro and Guanajuato, together with Jalisco, Zacatecas and San Luis Potosí comprise El Bajío.

The orography of the region is defined by three key features: (1) the Trans-Mexican Volcanic Belt, which runs northwest to southeast, and where the MAQ is located; (2) the Central Mexican Plateau, which lies northwest and west of the MAQ, and (3) the Sierra Madre Oriental, a mountain range running northeast and east of the MAQ, beyond the limits of the Central Plateau. The Sierra Madre Oriental greatly affects precipitation as it forces orographic lifting, promoting rain on the windward side and inhibiting it on its leeward side.

The receptor site for this study was located at the National University Campus Juriquilla (JUR) on the northern border of the MAQ (20° 35' 35" N, 100° 23' 32" W, 1945 masl) because it is the host of a monitoring station, part of a University Network of Atmospheric Observatories. The station collects atmospheric data such as meteorology parameters and concentration of criteria pollutants.

2.2 Trajectory calculations

Four years (January 2014 to December 2017) of three-dimensional 12-h back trajectories reaching the JUR site, were calculated using the HYSPLIT model release 4.9 (Draxler and Hess, 1998; HYSPLIT, 2021). Trajectories were calculated at 100 and 500 m above ground level (magl) every 6 h (05:00, 11:00, 17:00, and 23:00 UTC). UTC time corresponds to local time minus 6 h in general, or minus 5 h during daylight saving time. Twelve-hour trajectories were chosen in order to keep the uncertainty results as small as possible, while encompassing important regional sources such as industrial parks and the Salamanca refinery. The HYSPLIT toolbox for Python (PySPLIT, 2021) was used to automatize trajectory generation and analysis (Warner, 2018).

HYSPLIT was driven by the wind fields from the North American Mesoscale (NAM) model, which is maintained by the National Centers for Environmental Prediction (NCEP) of the National Ocean and Atmospheric Agency (NOAA) with a horizontal resolution of 12 km (NAM, 2021).

As a control mechanism, a verification process was carried out in order to detect and estimate the HYSPLIT integration error. This process consisted in generating the inverse (forward trajectory) for all the trajectories. A relative error was calculated as the distance between the start of the back trajectory and the end of the forward trajectory, divided by the total distance of both trajectories. Figure S1 in the supplementary material shows the relative errors for all trajectories calculated each month.

Homogenous groups of transport patterns were identified using Ward's agglomerative hierarchical clustering algorithm, which is included in the HYSPLIT desktop package used to run the model. We looked for the minimum number of clusters that represent all possible flow patterns, using the total spatial variance (TSV) as an aid in this process.

2.3 Surface data

Criteria pollutants (SO_2, O_3, NO, CO, PM_{10}, and $PM_{2.5}$) concentration, and surface wind direction and speed data at the JUR monitoring station was obtained from the network website (RUOA, 2021). Wind data goes from June 2014 to December 2017, and pollutants data starts on august in August 2014. CO data ends in November 2017, and NO_2 in September 2017.

Additionally, temperature and precipitation data at 11 climatological stations close to the receptor site was obtained from the Servicio Meteorológico Nacional (National Weather Service, SMN) (CONAGUA, 2021). All stations are randomly scattered across the region surrounding the MAQ in the states of Querétaro and Guanajuato (see Table SI and Fig. 1).

3. Results and discussion

3.1 Trajectories

Some files from the NAM database were corrupted or not available for downloading. Therefore, 2% of the trajectories could not be simulated. In addition, trajectory integration errors greater than two standard deviations of the average relative error of the month (approximately 5% of the total) were discarded. All the remaining trajectories (a total of 5490) were grouped by month, independently of the year. Next, the clustering algorithm was applied to obtain the main transport pathways for each month. This procedure was performed separately for both

the 100 and 500 magl trajectories, and the results are shown in Figures 2 and S2, respectively. The comparison of both figures shows that flow regimes at both heights are similar; albeit slightly faster at 500 magl, as the wind above is less affected by the surface. The similarity between heights is in agreement with other authors (Eneroth et al., 2003; Markou and Kassomenos, 2010). Because of this, and the fact that emission sources lie on the first 100 m above the surface, we will focus on the results for 100 magl in the rest of the manuscript.

We also calculated the mean direction of all trajectories considering the angle of the vector difference between their endpoint and starting point, and classified them as northeast (NE, 1-90º), southeast (SE, 91-180º), southwest (SW, 181-270º) and northwest (NW, 270-360º). Figure 3 shows the percent of trajectories with a given direction in each month at 100 magl. The figure clearly shows that from June to October, more than 80% of the trajectories originated NE from the MAQ, in contrast to December to April with less than 50% of trajectories with this direction.

3.2 Surface wind and climatology data analysis

Figure 4 shows the monthly frequency of wind direction, as measured at the JUR station from June 2014 until December 2017. In general, winds were observed predominantly from the NE from June to October. However, starting November, northerly winds increased their frequency becoming predominant in January and February. From December until May, strong winds from the southwest were also observed. Wind roses for these months show a wider and more homogenous distribution of wind direction. Previous studies show that wind direction measured at JUR is coherent with wind direction measured in other sites within the MAQ (Camacho-Díaz, 2013; Olivares-Salazar, 2016), which suggests that the transport pathways found in this work might apply to the whole urban area.

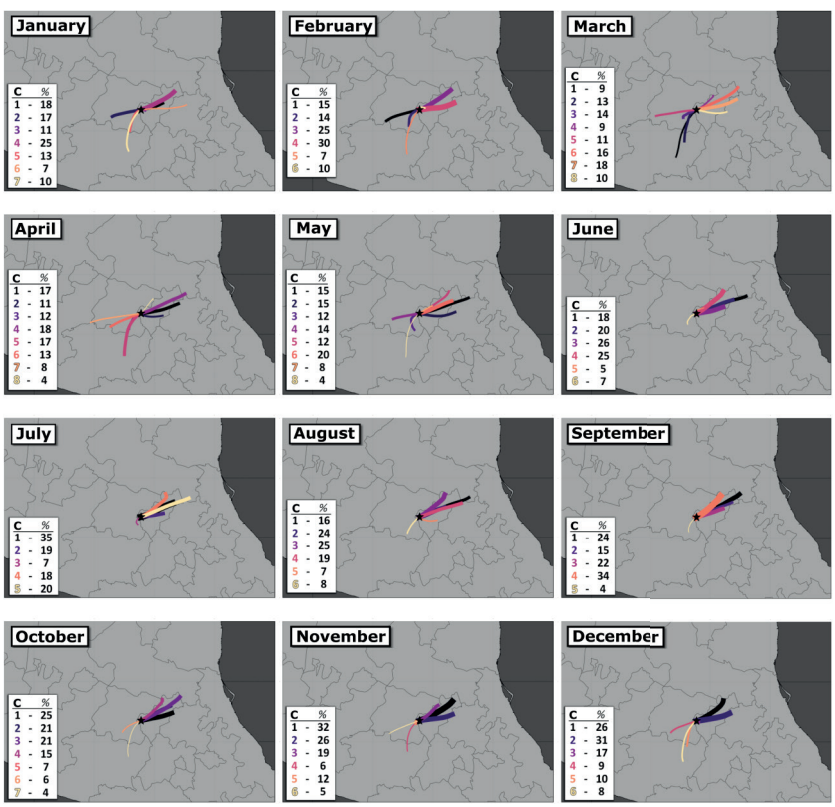

Fig 2. Monthly trajectory clusters calculated for the period 2014-2017, at 100 magl. The thickness of each line is proportional to the percent of trajectories in each cluster, which is specified for each month.

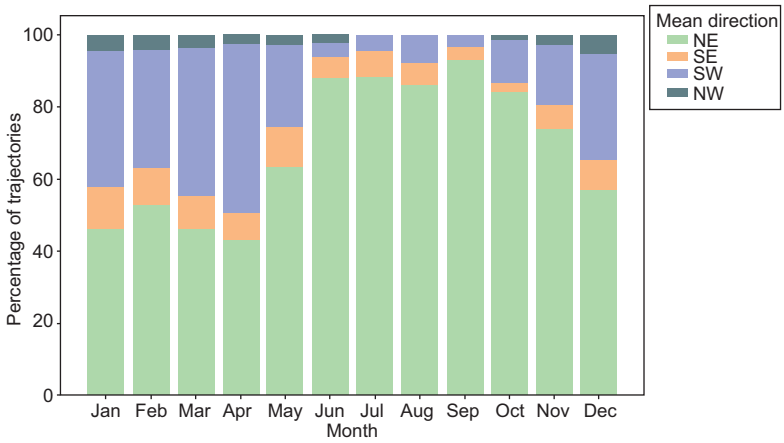

Fig 3. Monthly percent of trajectories with a given mean direction for the period 2014-2017, at 100 magl.

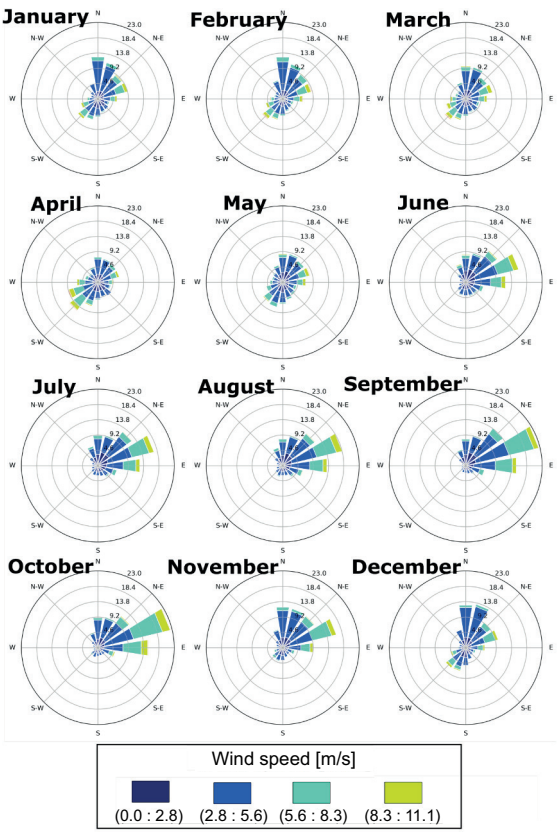

Fig 4. Monthly frequency of wind direction at the RUOA Juriquilla station (receptor site), from June 2014 until December 2017.

The climatological data of the selected 11 stations in the region is shown in Figure 5. The months with highest precipitation were June to September, with very little precipitation from November to April. Although the variability on precipitation reported among stations increased during the rainy months, the trend was the same for all stations. On the other hand, temperature increased from January until May, when it reached its maximum. After the rainy months began, temperature dropped a couple of degrees and remained relatively constant until October, when it started to drop again, reaching a minimum in December.

3.3 Seasonal transport pathways

The comparison of trajectories direction variability with local and regional wind and rain conditions shows that they all were consistent among two regimes. The temperature did not seem to have a correlation with those regimes.

The rainy regime, from June until September, was characterized by the largest amount of precipitation during the year. In these months, the flow regime was overwhelmingly from the east and northeast, which was clear in both the trajectories (with as few as 5% of the simulated trajectories coming from any other direction) and the wind roses at the receptor sites.

From December to May, very little precipitation was observed, so we defined this period as the dry regime. During these months, the flow regime had an important fraction of trajectories from the NE; however, many were also from the south and west. For March and April, this flow makes up approximately 50% of the total flow. This agrees with wind roses at JUR, which show winds coming from the north and NE, as well as SW

Surface measurements show an important northerly component of the wind for January and February, which is not present in the trajectory results. This situation could be explained by a microscale flow that is not captured by the 12-km NAM model. Other studies which include wind surface measurements at different points at the MAQ do not show this northerly component (Camacho-Díaz, 2013; Olivares-Salazar, 2016), which supports this explanation.

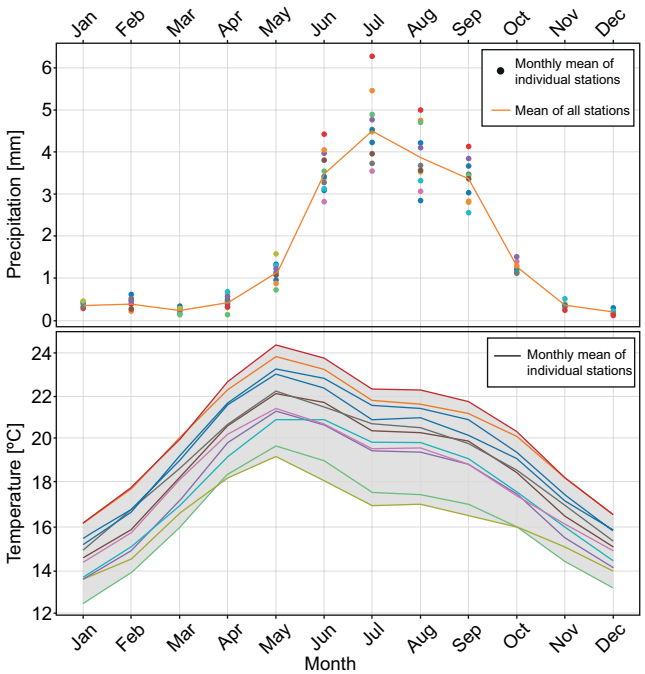

Fig 5. Temperature and rain data from the 11 climatological stations.

October and November did not exhibit consistent behavior among all the variables being considered; hence, we classified them as "transition months". For example, October exhibited some rain, and had a wind rose at JUR similar to the one during the rainy regime; but the trajectories from the SW were more frequent, exhibiting a similar pattern to the one seen in November.

Based on this classification, trajectories were grouped by regime and the clustering algorithm was applied again, obtaining the main transport pathways for each regime, shown in Figure 6. This figure constitutes the flow climatology of the MAQ and surrounding region for the period 2014-2017.

In order to validate the proposed flow climatology, we employed the NARR database (Mesinger et al., 2006; NARR, 2021), which uses the high resolution NCEP Eta Model (32 km/45 layers) together with the Regional Data Assimilation System (RDAS) to generate outputs every 3 h. Figure 7 shows wind field data in two different time ranges: 2014-2017, corresponding to the same period as the trajectories calculated in this work; and 1979-2019, corresponding to the climatology of the region. The wind directions observed during 2014-2017 are similar to those during the climatology (1979-2019); however, winds seem to be somewhat weaker during 2014-2017 in December-May and October-November, but stronger in June-September. These results indicate that the years studied in this work were not atypical in comparison with a longer timeframe.

In the comparison between the trajectories produced and the NARR (2014-2017) (Figs. 6 and 7) there is a coincidence between the direction of the most frequent trajectories and the average wind during the three regimes. During December-May, it is also observed that the receptor site is located in a convergence zone, which is why the dominant components of the trajectories are located in the NE and SW. Finally, it is evident that the trajectories and the winds derived from NARR have shorter paths that follow the surface topography. This is also observed in a closer look at the trajectory's paths: they never cross the highest topographic features (> 2200 masl) such as the Trans-Mexican Volcanic Belt and the Sierra Madre Oriental.

3.4 Criteria pollutants

Figure 8 shows the monthly mean and standard deviation concentration of SO_2, CO, NO, O_3, PM_{10}, and $PM_{2.5}$ at the receptor site during the study period. All these pollutants displayed two maxima in December and April or May, with the exception of O_3, which only displayed the April maximum. The lowest concentration of O_3 was observed in December; for the rest of the pollutants, the minimum was observed in September.

SO_2, CO, NO, and PM_{10} are primary pollutants whose concentrations are strongly related to emissions and transport. The main emission source of SO_2 is industry which, for the MAQ, represents a regional source (for example, the industrial corridor León-Irapuato-Salamanca). CO and NO are of industrial origin; however, they are also locally emitted by vehicles. PM_{10} is mainly regionally emitted by industry, and by resuspension in a semi-arid climate. Hence, it is unsurprising that the highest concentrations of these pollutants were observed during the dry season, when winds from the west were observed. The two maxima observed in the concentrations during this regime were probably related with meteorological factors such as temperature inversions during the coldest months.

In order to explore the relationship between the estimated trajectories, regional emissions, and pollutant concentrations at MAQ, we determined the frequency of the mean direction (see section 3.1) of the trajectories in days with high and low concentrations of SO_2 at the receptor site during the dry regime. To do so, we calculated the media (C) and standard deviation (SD) of the daily averages of SO_2 concentrations. Days with daily average $SO_2 > C + 2SD$ were classified as high SO_2 concentration level; days with daily average $SO_2 < (C - 2SD)$ were classified as low SO_2 concentration level. Figure 9 shows the percent of trajectories with a given mean direction for days with high and low SO_2 concentrations at 100 magl during the dry regime. Results indicate that during high SO_2 days, 70% of the trajectories originate on the SW direction. In low SO_2 days, only 40% of the trajectories originate on this direction and 50% on the NE direction. These results suggest that emissions on the SW quadrant might be affecting the air quality at MAQ.

For O_3, being a secondary pollutant, its concentration might be more dependent on meteorological conditions that affect reaction rates, such as temperature and radiation, rather than on wind direction. In the beginning of the year, high temperatures and

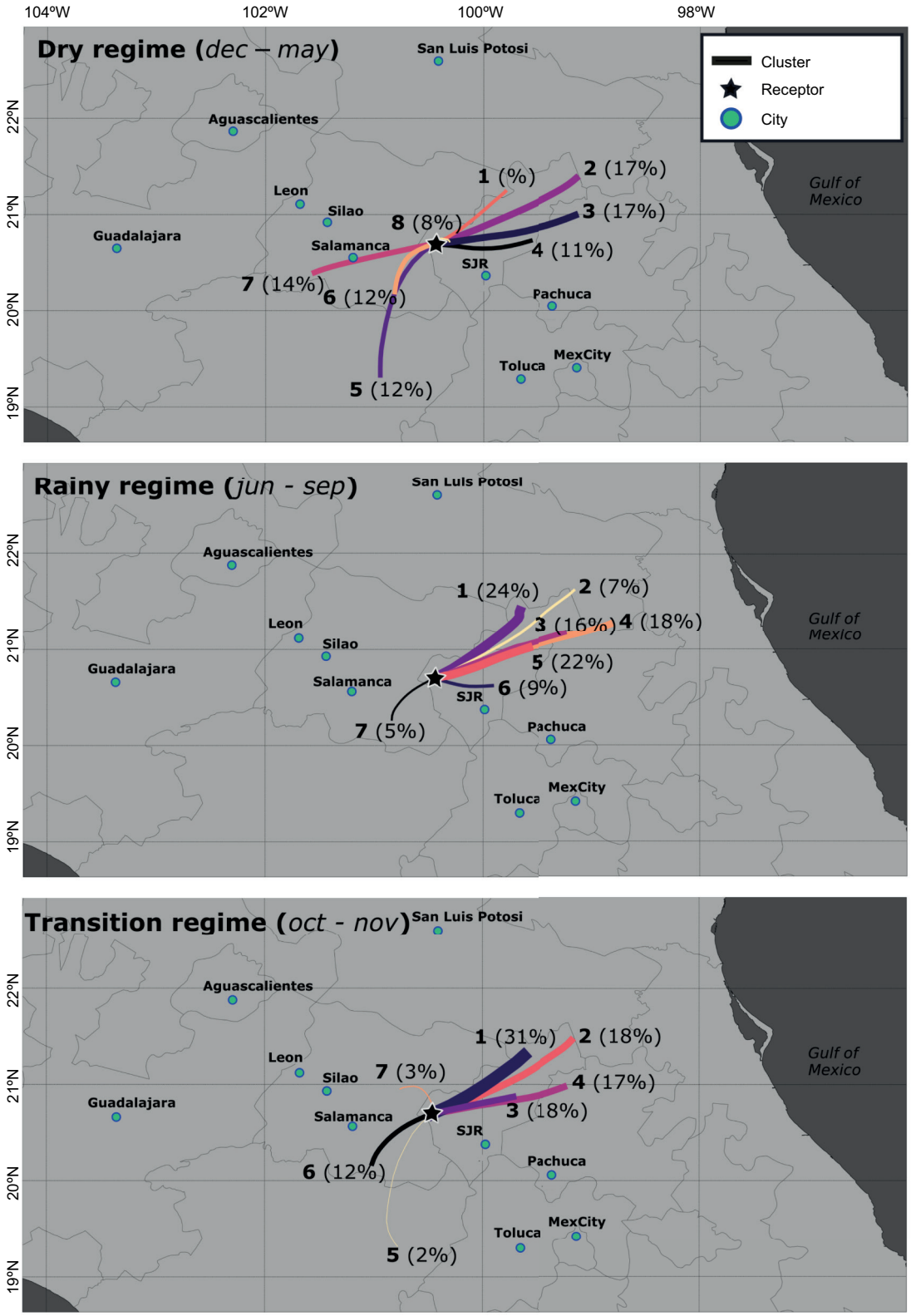

Fig 6. Flow climatology of the MAQ and surrounding region for the period 2014-2017, at 100 magl. The thickness of each line is proportional to the percent of trajectories in each cluster.

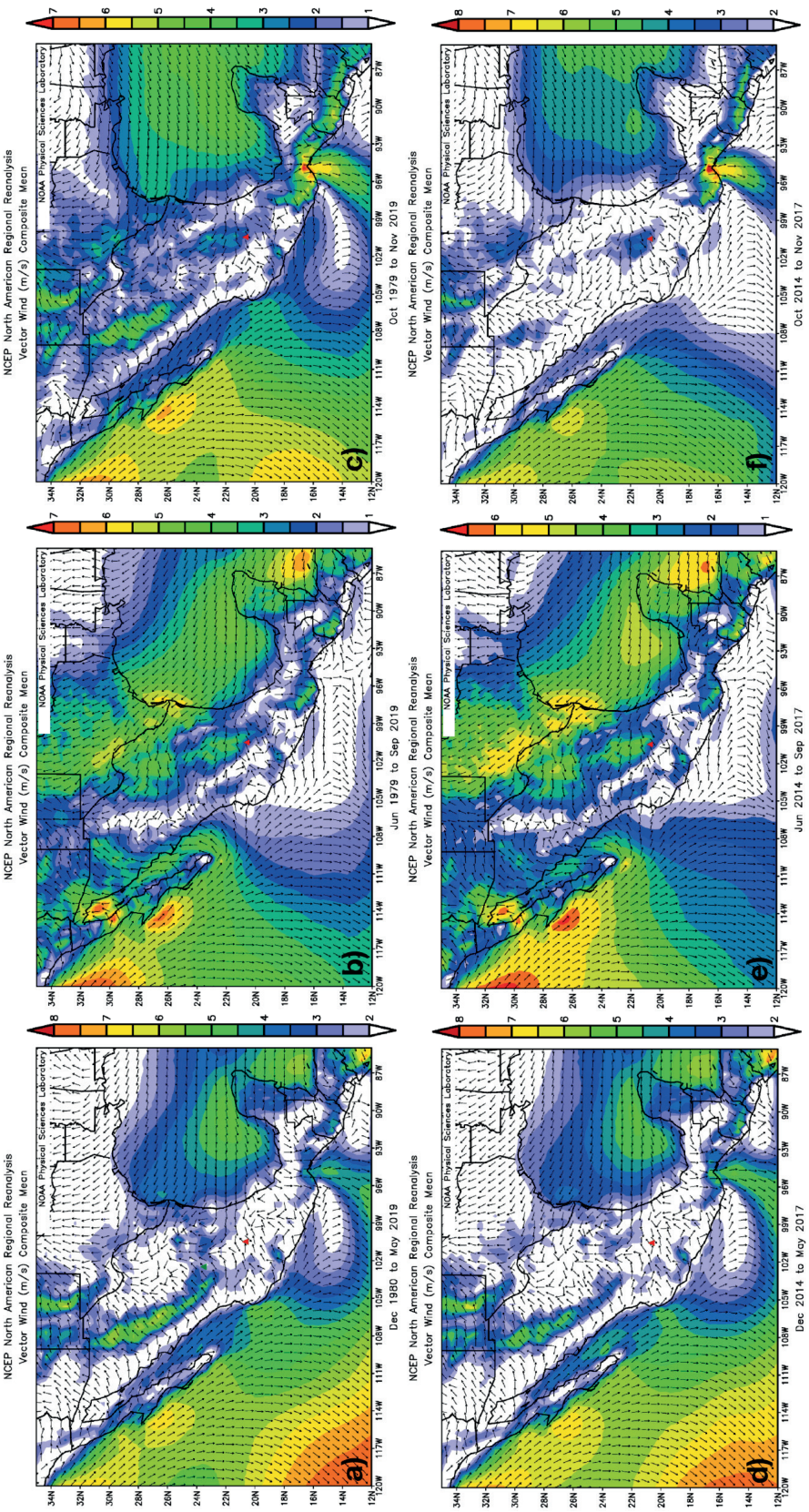

Fig 7. NARR derived wind fields during 1979-2019 (top) and 2014-2017 (bottom) periods, for (a and d) dry regime, (b and e) rainy regime, and (c and f) transition months. The red triangle represents the receptor site (JUR) used in the trajectory analysis.

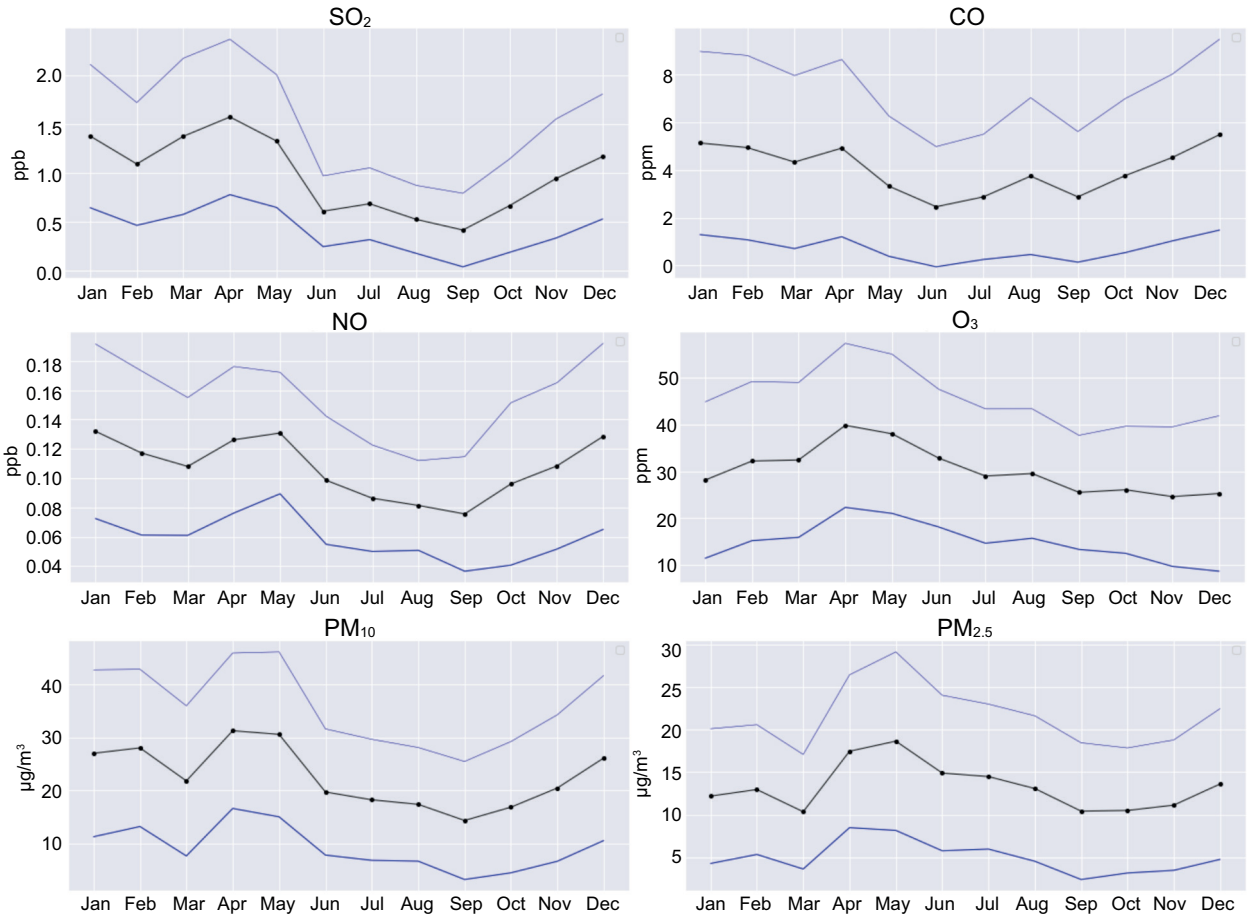

Fig 8. Monthly mean (black dots) of criteria pollutants at the receptor site during the study period. Blue lines represent mean ± standard deviation.

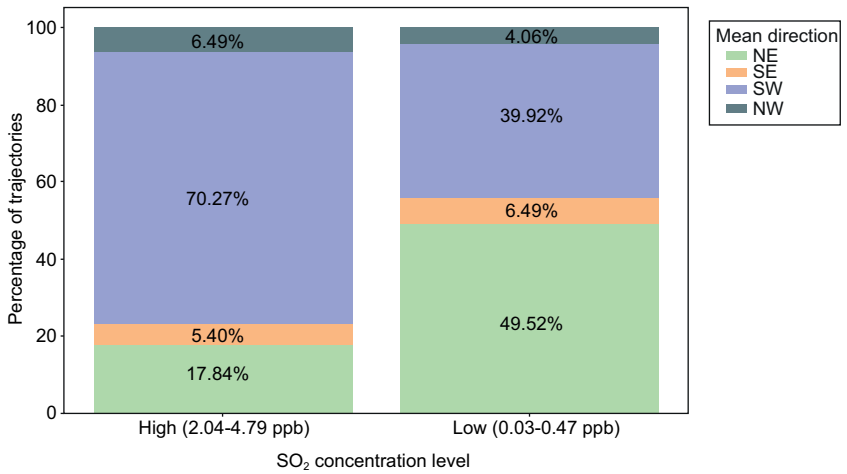

Fig 9. Percent of trajectories with a given mean direction for days with high and low SO_2 concentrations, during the dry regime, at 100 magl.

radiation combined with weak winds (Fig. 4) might generate the large O_3 concentrations observed. The same is observed in Mexico City, where the O_3 season spans from mid-February to mid-June (Zavala et al., 2020). For the MAQ, transport of ozone precursors from the industrial areas during the dry regime, might also contribute to the high concentrations of this gas. During the rainy regime, with no winds from the west, as well as faster local winds and rain, pollutants concentrations decreased.

4. Conclusions

Each cluster in Figure 6 represents a main transport pathway of the flow regime towards the MAQ. The percentage of trajectories assigned to a cluster (and the thickness of the line) indicates the expected frequency of the air following the path of that cluster during that regime. It is evident that a NE flow is present throughout the year. Trajectories from the SW also occur all year long, but they are much less frequent during the rainy regime and the transition months. The clustering algorithm groups trajectories not only by mean direction, but also by longitude and curvature. Hence, short clusters are composed of slow-moving trajectories or trajectories with a high degree of curvature that remain close to the receptor site. The meteorological conditions that give rise to this kind of flow are calm winds or recirculation. The result of both is that air might remain stagnant above the same area facilitating the concentration of pollutants. These clusters are observed during the dry regime and transition months.

It is important to remark that the analyzed period (2014-2017) is consistent with a longer period of time (1979-2019), which implies that the results adequately represent the average conditions of the atmosphere at the study site. Furthermore, the wind patterns of El Bajío are strongly related to the orography of the region.

Some of the trajectories from the NE originate within a desertic region of the state of Querétaro (NE from the MAQ), where several limestone mines are located. Dust resuspension might be common there, especially during the dry months. During the dry regime and transition months, it is clear that some clusters originate in the industrial area of Guanajuato, which includes the Salamanca refinery. Because air transport of pollutants follows these paths, regional sources might be affecting the MAQ air quality.

It is noticeable that no clusters show any transport from the SE of the MAQ, so it is unlikely that there is a contribution of pollutants from Mexico City and its surroundings probably due to the presence of the Trans-Mexican Volcanic Belt south of the MAQ.

This study describes for the first time a flow climatology in central Mexico, which might be useful for finding both geographical and temporal patterns of the wind driven transport of the region. The results raise important questions regarding regional sources of pollutants that might be affecting the air quality. Future studies should address the influence that industrial areas around the MAQ, in El Bajío, have on the concentration of air pollutants such as ozone and particulate matter. They should also weigh the effect of local emissions, regional transport, and meteorological conditions on the air quality of the region.

Acknowledgments

This work was supported by UNAM-PAPIIT 2019-IN115719. NCEP Reanalysis data was provided by the NOAA/OAR/ESRL PSL, Boulder, Colorado, USA, from their website (https://psl.noaa.gov/).

References

Camacho-Díaz JG. 2013. Mapeo de la calidad ambiental urbana mediante bioindicadores en la ciudad de Santiago de Querétaro. Ph.D. thesis. Universidad Autónoma de Querétaro, Querétaro, Mexico.

CONAGUA. 2021. Información estadística climatológica. Comisión Nacional del Agua, Mexic. Available at: https://smn.conagua.gob.mx/es/climatologia/informacion-climatologica/informacion-estadistica-climatologica (accessed on January 10, 2021).

Draxler R, Hess G. 1998. An overview of the HYSPLIT_4 modeling system for trajectories, dis-persion, and deposition. Australian Meteorological Magazine 47: 295-308.

Eneroth K, Kjellström E, Holmén K. 2003. A trajectory climatology for Svalbard; investigating how atmospheric flow patterns influence observed tracer concentrations. Physics and Chemistry of the Earth, Parts A/B/C 28: 1191-1203. https://doi.org/10.1016/j.pce.2003.08.051

HYSPLIT. 2021. HYSPLIT web site. National Oceanic and Atmospheric Administration, College Park, MD, USA. Available at: https://ready.arl.noaa.gov/HYSPLIT.php (accessed on January 10, 2021).

INECC, 2019. Informe nacional de calidad del aire 2018. Instituto Nacional de Ecología y Cambio Climático, Mexico, 343 pp.

INEGI, 2015. Encuesta intercensal 2015. Instituto Nacional de Estadística y Geografía, Mexico. Available at: https://www.inegi.org.mx/programas/intercensal/2015/ (accesed on September 14, 2022).

Katsoulis BD. 1999. The potential for long-range transport of air-pollutants into Greece: A clima-tological analysis. Science of The Total Environment 231: 101-113. https://doi.org/10.1016/S0048-9697(99)00100-X

NAM. 2021. NCEP North American Regional Reanalysis archive. National Oceanic and Atmospheric Administration, Boulder, CO, USA. Available at: ftp://arlftp.arlhq.noaa.gov/pub/archives/nam12 (accessed on January 10, 2021).

Markou MT, Kassomenos P. 2010. Cluster analysis of five years of back trajectories arriving in Athens, Greece. Atmospheric Research 98: 438-457. https://doi.org/10.1016/j.atmosres.2010.08.006

Mesinger F, DiMego G, Kalnay E, Mitchell K, Shafran PC, Ebisuzaki W, Jović D, Woollen J, Rogers E, Berbery EH, Ek MB, Fan Y, Grumbine R, Higgins W, Li H, Lin Y, Manikin G, Parrish D, Shi W. 2006. North American Regional Reanalysis. Bulletin of the American Meteorolog-ical Society 87: 343-360. https://doi.org/10.1175/bams-87-3-343

NARR. 2021. NCEP North American Regional Reanalysis archive. National Oceanic and Atmospheric Administration, Boulder, CO, USA. Available at: https://psl.noaa.gov/cgi-bin/data/narr/plotday.pl (accessed on January 10, 2021).

Olivares-Salazar SE. 2016. Estudio de la calidad del aire de la ciudad de Querétaro en el año 2013. B.Sc. thesis. Universidad Nacional Autónoma de México, Mexico.

PySPLIT. 2021. PySPLIT repository. Python Software Foundation, Fredericksburg, VA, USA. Available at: https://pypi.org (accessed on January 10, 2021).

RUOA, 2021. RUOA's data archive. Red Universitaria de Observatorios Atmosféricos, Univerasidad Nacional Autónoma de México, Mexico. Available at: https://www.ruoa.unam.mx (accessed on January 10, 2021).

Stein A F, Draxler RR, Rolph GD, Stunder BJB, Cohen MD, Ngan F. 2016. NOAA's HYSPLIT Atmospheric Transport and Dispersion Modeling System. Bulletin of the American Meteorological Society 96: 2059-2077. https://doi.org/10.1175/BAMS-D-14-00110.1

Warner MSC. 2018. Introduction to PySPLIT: A Python toolkit for NOAA ARL's HYSPLIT Model. Computing in Science & Engineering 20: 47-62. https://doi.org/10.1109/MCSE.2017.3301549

Zavala M, Brune WH, Velasco E, Retama A, Cruz-Alavez LA, Molina LT. 2020. Changes in ozone production and VOC reactivity in the atmosphere of the Mexico City Metropolitan Area. Atmospheric Environment 238: 117747. https://doi.org/10.1016/j.atmosenv.2020.117747

SUPPLEMENTARY MATERIAL

Table SI. Climatological stations from the National Weather Service, and the time range of data for each one.

Station Name	Data period
El Batan	06/1965-07/2016
El Pueblito	01/1961-06/2016
San Juan del Rio	02/1973-06/2015
E.T.A. 128 Carrillo	03/1987-04/2015
La Joya	08/1975-04/2017
Juriquilla	11/1979-01/2015
Querétaro (DGE)	06/1982-07/2016
Celaya	02/1922-10/2016
Irapuato	05/1903-01/2016
La Gavia	07/1979-07/2016
San Juan Temascatio	01/1993-01/2016

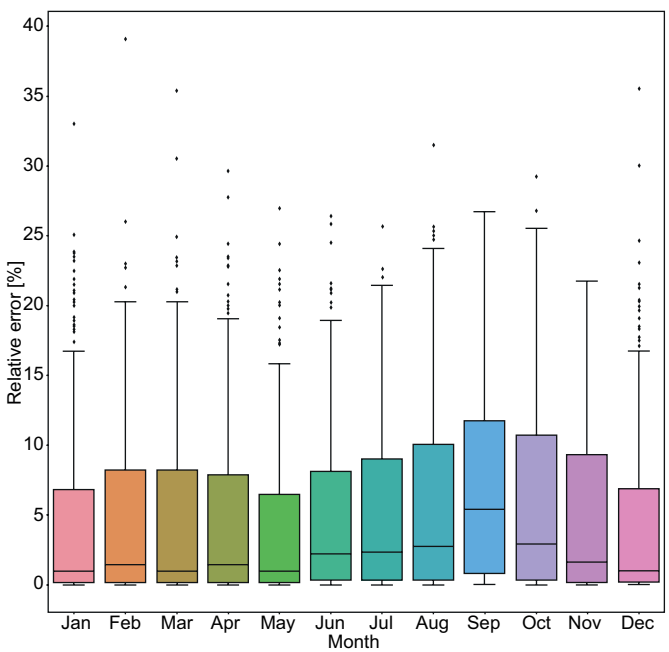

Fig. S1. Relative error of all trajectories calculated each month. 10th, 25th, 50th, 75th, and 90th percentiles are represented with the whiskers and box.

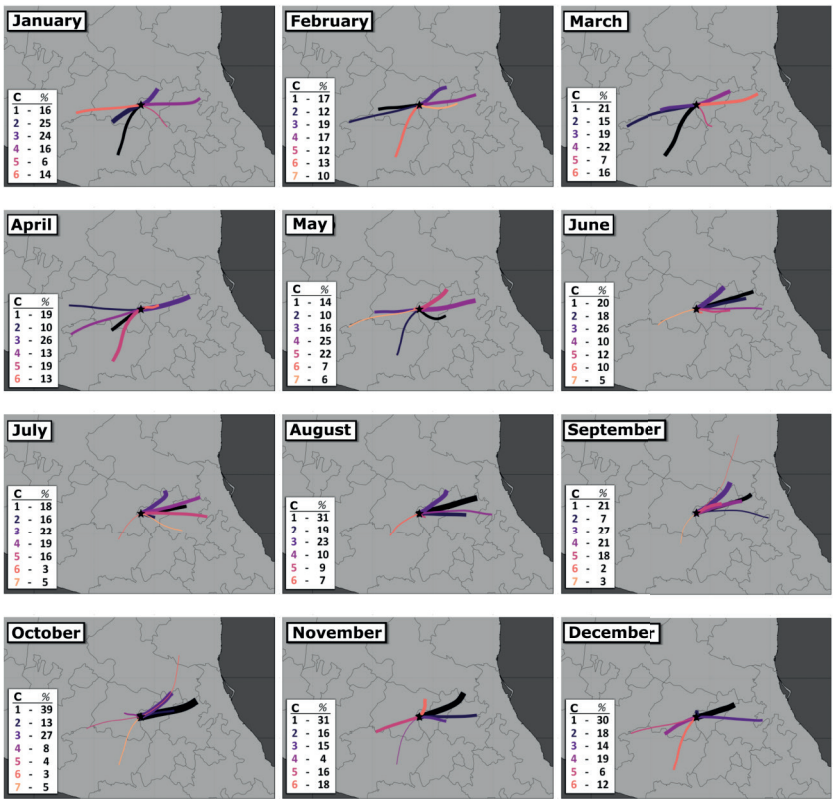

Fig S2. Monthly trajectory clusters calculated for the period 2014-2017, at 500 magl. The thickness of each line is proportional to the percent of trajectories in each cluster, which is specified for each month.

Atmósfera 36(2), 253-277 (2023)
https://doi.org/10.20937/ATM.52968

CONTENTS

Performance evaluation of the WRF model in a tropical region: Wind speed analysis at different sites

Noéle B. P. SOUZA[1], Erick G. S. NASCIMENTO[2] and Davidson M. MOREIRA[1,2]*

[1]*Universidade Federal do Espírito Santo, Vitória, 29075-910, Espírito Santo, Brasil.*
[2]*Centro Integrado de Manufatura e Tecnologia, Salvador, 41650-010, Bahia, Brasil.*
*Corresponding author: davidson.moreira@fieb.org.br

Received: August 28, 2020; accepted: June 29, 2021

RESUMEN

En este estudio se evalúa el rendimiento del modelo de pronóstico e investigación meteorológica de mesoescala (WRF) utilizando combinaciones de tres capas de límite planetario (PBL, por su sigla en inglés) (YSU, ACM2 y MYJ) y tres esquemas de modelo de superficie terrestre (LSM) (UC, Noah y Noah-MP), con el fin de identificar los parámetros óptimos para la determinación de la velocidad del viento en una región tropical. El estado de Bahía en Brasil se selecciona como la ubicación para el estudio de caso y las simulaciones se realizan durante un periodo de ocho meses entre 2015 y 2016 (estaciones secas y lluviosas). Los resultados de las simulaciones se comparan con los datos de observación obtenidos de tres torres equipadas con anemómetros a alturas de 80, 100, 120 y 150 m, colocadas estratégicamente en cada sitio y evaluadas con los índices estadísticos MB, RMSE, MAGE, IOA, R, Fac2 y desviación estándar. Se observa una sobreestimación de la velocidad del viento en las simulaciones, a pesar de las similitudes entre las direcciones del viento simuladas y observadas. Además, se observa que las precisiones de las simulaciones correspondientes a sitios que están más cerca del océano son menores. Las estimaciones de velocidad del viento más precisas fueron las correspondientes a Mucugê, que se encuentra más alejado del océano. Finalmente, el análisis de los resultados obtenidos de cada torre, teniendo en cuenta los periodos con mayor y menor precipitación, revela que la combinación del esquema PBL-YSU con el esquema LSM-RUC arroja los mejores resultados.

ABSTRACT

In this study, the performance of the mesoscale Weather Research and Forecasting (WRF) model is evaluated using combinations of three planetary boundary layers (PBL) (YSU, ACM2, and MYJ) and three land surface model (LSM) schemes (RUC, Noah and Noah-MP) in order to identify the optimal parameters for the determination of wind speed in a tropical region. The state of Bahia in Brazil is selected as the location for the case study and simulations are performed over a period of eight months between 2015 and 2016 (dry and rainy seasons). The results of the simulations are compared with observational data obtained from three towers equipped with anemometers at heights of 80, 100, 120 and 150 m, strategically placed at each site and evaluated with statistical indices: MB, RMSE, MAGE, IOA, R, Fac2 and standard deviation. Overestimation of wind speed is observed in the simulations, despite similarities between the simulated and observed wind directions. In addition, the accuracies of simulations corresponding to sites that are closer to the ocean are observed to be lower. The most accurate wind speed estimates are obtained corresponding to Mucugê, which is located farthest from the ocean. Finally, analysis of the results obtained from each tower accounting for periods with higher and lower precipitation reveals that the combination of the PBL-YSU scheme with the LSM-RUC scheme yields the best results.

Keywords: WRF, parametrization, Bahia, wind energy, tropical climate, renewable energy.

1. Introduction

The utilization of wind energy has increased exponentially over the recent decades and it has been well-established as a sector of energy. However, several challenges remain regarding its growth and development (Draper et al., 2018). For example, simulation of relevant physical phenomena need to be improved to ensure better operational practices by enabling the integration of larger fractions of wind energy into energy networks, scheduling maintenance in wind power installations, and defining design criteria for next-generation turbines and installations. Current numerical models used by wind-farm operators to obtain wind estimates utilize series with typical horizontal grid resolution less than 2 km. However, estimation accuracy is critical for wind power operators as improvements by even 1 ms^{-1} in wind speed prediction can induce millions of dollars in savings due to the stiff financial penalties for overestimation and underestimation of energy production (Marjanovic et al., 2014). Therefore, the choice of physical parameterization and the computational grid to be utilized in wind energy problems is crucial to improve the representation of wind energy in the region under analysis.

The northeast region of Brazil generates the highest proportion of wind energy in the country, accounting for 84 % of the national wind production (ABEEólica, 2017). In particular, the northeastern state of Bahia, located in the tropics, is the second largest producer of wind energy in Brazil and receives the highest number of wind projects in the country. Weather forecasting and simulation in the tropics are challenging as tropical climate is liable to rapid variation induced by prevalent convection and sea breeze, and is dominated by local, meso and macroscale effects. Winds in the tropics are generally light and variable, and observations of surface and upper layers of air that are required for numerical weather models are also scarce (Surussavadee, 2017a). For these reasons, despite the considerable investments that have already been made, the region needs further in-depth studies based on data obtained using computational tools at various altitudes to obtain better estimates of wind speed.

The WRF mesoscale model (Skamarock et al., 2008) is a numerical weather forecasting model that has been widely used in research and evaluation of wind and solar energy production. The physical parameterizations required by the model can be classified into several categories: planetary boundary layer (PBL), land surface model (LSM), surface layer, microphysics, cumulus, and longwave and shortwave radiation, each one containing multiple parameterizations available for modeling. Physical parameterizations are used to describe physical processes using simplified equations in order to reduce the number of unknown factors that influence atmospheric behaviors. Thus, the choice of the optimal physical parameterization based on comparisons with data measured in different parts of the world is essential to maximize the diversity of applications of wind energy (Kumar et al., 2010; Mohan and Bahati, 2011; Soni et al., 2014; Islam et al., 2015; Imran et al., 2017). The analysis of these schemes is not trivial and requires multiple simulations. Most previous studies have focused on evaluating the performance of forecasting models in tropical regions (Salvador et al., 2016a). This is an inherently challenging task due to the intense convective characteristics of these regions, which rapidly change the state of the atmosphere, and the influence of sea breezes and local and mesoscale movements (Hariprasad et al., 2014; Surussavadee, 2017b). In the said context, this study examines the effectiveness of different combinations of PBL and LSM schemes in terms of wind speed estimation.

PBL is parameterized to represent vertical flows within subgrade scales caused by turbulent transports throughout the boundary layer, and not merely in the superficial layer. When a PBL scheme is activated, explicit vertical diffusion is disabled under the assumption that the process is controlled by the scheme. The WRF model offers several types of PBL parameterization, each of which corresponds to a distinct parameterization of the flow divergence terms (turbulent part) of the governing atmospheric flow equations to capture the trends of the variables induced by these terms. Several previous studies have proposed optimal PBL parameterization schemes (Hariprasad et al., 2014; Banks and Baldasano, 2016; Banks et al., 2016; Boadh et al., 2016; Salvador et al., 2016b; Avolio et al., 2017; Imran et al., 2017; Kitagawa et al., 2017). Further, the WRF model provides certain LSM schemes to calculate soil temperature profiles, surface temperature profiles, soil moisture profiles, snow covers, and canopy properties. LSM is utilized

in forecasting models because of the importance of soil surface processes in wind energy applications. Multiple important processes of the Earth's surface are addressed by this parameterization, such as evapotranspiration from canopy water, evapotranspiration from snow, runoffs and melting of snow, depending on the complexity of the scheme used. Several comparative studies have demonstrated that meteorological models are sensitive to the choice of LSM (Pei et al., 2014; Wharton et al., 2015; Lee et al., 2016; Jain et al., 2017; Salamanca et al., 2018; Liu et al., 2019).

Unlike most studies that use 10 m-high surface data to validate results, this study uses data from towers equipped with anemometers at 80, 100, 120, and 150 m located at three different sites (Esplanada, Mucuri, and Mucugê). This ensures that the data are representative of the true values of wind energy. In this sense, the primary objective of this study is the identification of optimal parameterizations of PBL and LSM schemes using the WRF model to improve the evaluation of wind speeds in different regions in the state of Bahia, that is characterized by the influence of dry and rainy seasonality on energy behavior in the northeast region of Brazil (Ramos et al., 2013). The rest of the study is organized as follows. Certain aspects about the parameterizations used in the study are discussed in section 2. The methodology used in the study is introduced in detail in section 3. The numerical results are presented in section 4. Finally, the study is concluded in section 5.

2. Parameterizations

2.1 PBL schemes

Accurate model representation of PBL processes and interactions via parameterization of models is essential to realistic simulations. In particular, turbulence is parameterized to infer about the variation of atmospheric parameters that are not resolved at all grid points. In this study, the performances of three PBL schemes are investigated, two nonlocal first order closures (YSU and ACM2) and one local closure (MYJ).

The YSU parameterization is called a first-order closure model, as it does not require additional equations to express the effect of turbulence on the primary variables. To solve the equation, the term is added to the mixed layer to express the turbulent diffusion, as follows:

$$\frac{\partial C}{\partial t} = \frac{\partial C}{\partial z}\left[K_c\left(\frac{\partial C}{\partial z} - \gamma_c\right) - (\overline{w'c'})_h\left(\frac{z}{h}\right)^3\right] \quad (1)$$

where C denotes heat or momentum, K_c denotes the eddy diffusivity coefficient, γ_c denotes an adjustment term for the local gradient, z denotes the elevation above the ground, h denotes the PBL height, and $(\overline{w'c'})_h$ denotes the flux at the inversion layer (Xie et al., 2012).

The ACM2 parameterization is a mixed local and nonlocal closure model that addresses nonlocal flows using a transient matrix. The governing equation is as follows:

$$\frac{\partial C_i}{\partial t} = f_{conv} Mu C_1 - f_{conv} Md_i C_i + f_{conv} Md_{i+1} C_{i+1} \frac{\Delta z_{i+1}}{\Delta z_i} + \frac{\partial}{\partial z}\left[K_c(1-f_{conv})\frac{\partial C_i}{\partial z}\right] \quad (2)$$

where Mu denotes the nonlocal upward convective mixing rate, Md_i denotes the downward mixing rate from the layer i to the layer i-1, C_i denotes the scalar at the layer i, Δz_i denotes the thickness of the layer i, and f_{conv} determines the contribution of nonlocal mixing compared to that of local mixing. The partitioning factor f_{conv} is derived from the ratio of the nonlocal heat flux to the total heat flux at the top of the surface layer (0.1 h) (Holtslag and Boville, 1993). f_{conv} increases rapidly from zero to a stable or neutral atmosphere and is a near asymptote 0.5 corresponding to unstable atmospheres. f_{conv} is defined as follows:

$$f_{conv} = \left[1 + \frac{k^{-2/3}}{0.1a}\left(-\frac{h}{L}\right)^{-1/3}\right]^{-1} \quad (3)$$

where a denotes a constant equal to 7.2, k denotes the von Karman constant, h denotes the PBL height, and L denotes the Monin-Obukhov length scale (Xie et al., 2012).

The MYJ parameterization is a local closure scheme, also called the Turbulent Kinetic Energy (TKE) closure scheme because it determines eddy diffusion coefficients from prognostic TKE and provides the turbulent flows corresponding to each point based on the mean values of atmospheric variables (Xie et al., 2012). The model expresses the diffusivity as follows:

$$K_c = S_c l e^{1/2} \quad (4)$$

where S_c denotes a numerical coefficient, l denotes the master length scale, and e denotes the TKE.

Each of the three PBL schemes employs a distinct technique to diagnose PBL heights (h), which is used in other physics parameterizations. The YSU scheme determines h as the first neutral level determined by the bulk Richardson number for unstable conditions (Hong et al., 2006). The ACM2 scheme follows a similar method, determining h as the height of the level where the bulk Richardson number exceeds a critical value (Pleim, 2007). In contrast, the MYJ scheme determines h as the height where TKE assumes the minimum value.

2.2 Surface Layer Parameterization

The precise formulation of the surface layer is essential to determine the interaction between soil and atmosphere, which is required to deduce a more accurate evolution of the atmosphere for numerical models. The surface layer determines friction velocity and surface stress, used in the PBL scheme, as well as the coefficients used in the LSM scheme to compute heat and moisture flows. In the presence of water on the surface, the surface fluxes are calculated within the surface layer scheme itself. Turbulent surface fluxes are often calculated using the similarity theory proposed by Monin-Obukhov (Monin and Obukhov, 1954; Obukhov, 1971), which provides information about the profiles in the surface layer that aid the determination of the temperature and humidity at a height of 2 m and wind characteristics at a height of 10 m, among other variables. The surface layer scheme operates in combination with a particular PBL scheme.

The YSU and ACM2 schemes can be coupled with the MM5-similarity surface layer scheme (Zhang and Anthes, 1982) to determine the surface exchange coefficients of momentum, heat, and moisture. The coupled schemes utilize the stability functions proposed by Paulson (1970) and consider four stability schemes. Over the ocean, the Charnock parameter is used to relate roughness length to friction velocity and to improve determination of surface fluxes of heat and moisture. It uses the convective velocity proposed by Beljaars (1994).

The MYJ scheme can be coupled with only the Eta-similarity surface layer scheme, which is also based on the similarity theory proposed by Monin-Obukhov and includes parameterizations of a viscous sublayer. It calculates surface fluxes using an interactive method. To estimate temperature and humidity over land, the effects of the viscous sublayer are related to the variable roughness height proposed by Zilitinkevich (1995), whereas to do the same over the ocean, the viscous sublayer is parameterized following Janjic (1994). In the case of an unstable surface layer, the Beljaars (1994) correction is applied to avoid singularities.

2.3 Land Surface Models

The WRF model provides a few LSMs that compute heat and moisture fluxes over land, which vary with respect to complexity and methodology. To assess their role in the simulations in the Bahia area, the RUC model, Noah, and Noah-MP LSM schemes are used.

The RUC model (Smirnova et al., 1997, 2000) solves equations of energy balance and moisture on the surface by employing an implicit scheme to calculate surface flows. It solves the heat diffusion equation and Richard's moisture transfer equation by considering six layers of soil and the phase changes of water contained in the soil during cold periods.

The Noah model (Chen and Dudhia, 2001) uses a single linear equation of the surface energy balance to calculate skin temperature by considering the ground-vegetation surface. The equations of the model consider the type of vegetation and the soil texture. Soil temperature and moisture, as well as canopy moisture, are predicted corresponding to four layers. Soil moisture is obtained via Richard's equation and the temperature is given by the thermal diffusion equation. As in the case of the RUC model, the effect of plant transpiration is considered, but the resistance of the canopy is determined to be similar to that obtained via the scheme proposed by Xiu and Pleim (2001).

Noah-MP is a land surface model that uses multiple options to estimate crucial land-atmosphere interaction processes (Niu et al., 2011). It considers a separate vegetation canopy defined by the canopy top and bottom, crown radius, and leaves with prescribed dimensions, orientation, density, and radiometric properties. The Noah-MP model can be executed by prescribing both the horizontal and vertical densities

Table I. Geographical information of the measurement sites.

Measurement sites	Latitude S	Longitude W
Esplanada Tower	11º 50' 55.22953"	37º 55' 44.31164"
Mucuri Tower	18º 1' 31.52"	39º 30' 51.69"
Mucugê Tower	13º 21' 01.9289"	41º 31' 53.76975"

of vegetation using either ground- or satellite-based observations. The model is capable of distinguishing between different photosynthesis pathways and can define vegetation-specific parameters related to plant photosynthesis and respiration.

3. Methodology

3.1 Study area

This study focuses on the state of Bahia, located in northeastern Brazil, between the parallels 08º 31' 58" S and 18º 20' 55" S, and the meridians 46º 37' 02" W and 37º 20' 28" W. The state of Bahia was the second largest producer of wind energy in Brazil in 2017 (7.79 TWh), behind only the state of Rio Grande do Norte (13.24 TWh). It also accounted for the second-highest factor average capacity in 2017 (48.5%), behind the state of Maranhão (68%) (ABEEólica, 2017). The interior of the state experiences the most intense winds concentrated in the dry period, unlike the conditions on the coast. Wind direction is observed to have little variation, with the east-west direction being predominant. Winds in the opposite direction are rarely recorded, and when recorded they exhibit very low speeds. The predominant climate is tropical, with high average temperatures and maximum annual temperatures above 30 ºC. In the hinterland, the climate is semi-arid, with annual rainfall below 800 mm. The rainy season is irregular, with prolonged drought events in the interior. The humidity on the coastal strip is higher than that in the interior and the annual accumulated precipitation exceeds 1600 mm in some regions (Camargo-Schubert, 2013). The generation of wind energy can vary in periods with more and less rain. In general, during periods of greatest drought (when it is not very windy), wind energy production suffers a small decrease. During the rainy periods (when there are more winds) there is a greater production of energy. Ramos et al. (2013) showed the influence of dry and rainy seasonality on energy behavior in the Northeast region of Brazil. Thus, analyses corresponding to the dry and rainy periods in each region are interesting, as the sites are very different from each other in terms of geographical position and seasonality.

The towers with anemometers are located at three different sites: the cities of Esplanada, Mucuri, and Mucugê. Table I show the summarized geographical information of the measurement sites.

The city of Esplanada is located in the micro-region of the northern coast of Bahia at an altitude of 158 masl. Its anemometric tower is located 40 km from the sea. Analysis of the climatological normals (Fig. 1) available for the Alagoinhas station (closest to Esplanada) reveals that the months of December, January, February, September, and October are less rainy, while May and June are the rainiest months in this region.

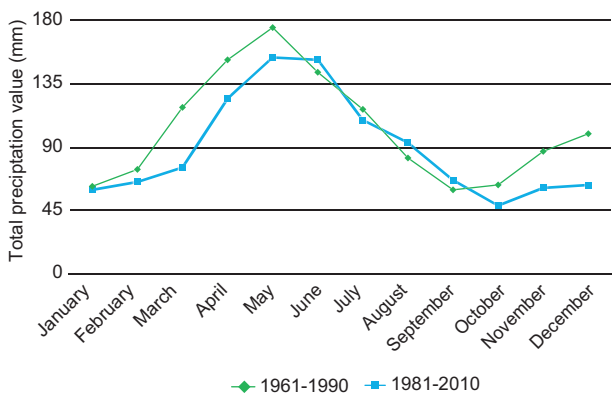

Fig. 1. Comparative accumulated precipitation (mm) at the Alagoinhas station.

The city of Mucugê is situated at an altitude of 983 masl and is one of the municipalities belonging to Chapada Diamantina, the central region of the state of Bahia, characterized by its mountainous terrain. The

Mucugê anemometric tower is located approximately 280 km from the coast of Bahia. Analysis of the climatological normals (Fig. 2) available for the Ituaçu station (closest to Mucugê) reveals that the months of May, June, July, August, and September are relatively dry, while November, December, January, and February are the rainiest months in this region.

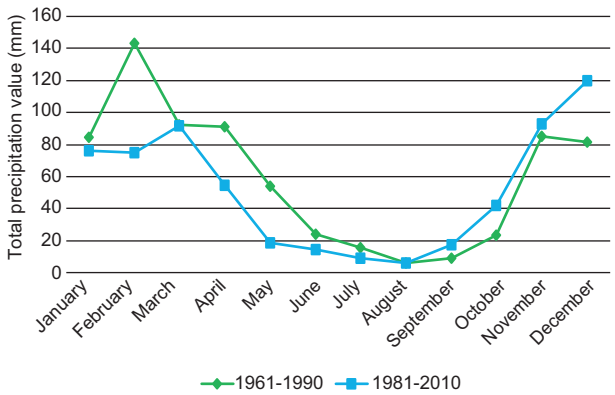

Fig. 2. Comparative accumulated precipitation (mm) at the Ituaçu station.

Finally, the city of Mucuri is located at an altitude of 7 masl. The Mucuri anemometric tower is located on a coastal plain, at a distance of 340 m from the sea. Analysis of the climatological normals (Fig. 3) available for the Caravelas station (closest to Mucuri) reveals that November, December, March, April, and May are the rainiest months in this region, while January, February, and August represent the driest

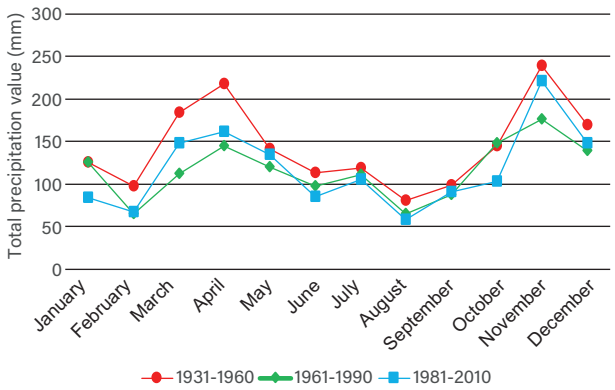

Fig. 3. Comparative accumulated precipitation (mm) at the Caravelas station.

period, considering the most current climatological normal (1981-2010).

3.2 Configuration of the WRF model

This study was conducted using the WRF meteorological model, version 3.9.1.1. The WRF model was configured using two nested domains with grid resolutions of 9 and 3 km, respectively. Inside the second domain, which covers the entire state of Bahia, three domains were designed with 1 km grid resolutions, centered on the three anemometric towers: Mucugê (Domain 3), Esplanada (Domain 4), and Mucuri (Domain 5). Figure 4 depicts the locations and distribution of domains in the WRF model. The domains were designed with horizontal dimensions of 223 × 223 and 420 × 420 grid cells corresponding to domains 1 and 2, respectively, and 60 × 60 grid cells corresponding to domains 3 to 5. During the initialization of WRF, data from the National Centers for Environmental Prediction (NCEP) Final Analysis (FNL), with a spatial resolution of 0.25º prepared operationally every 6 h were used (NCAR, 2015). Land use and occupation data were provided by the United States Geological Survey (USGS) with a resolution of 2' corresponding to the largest domain

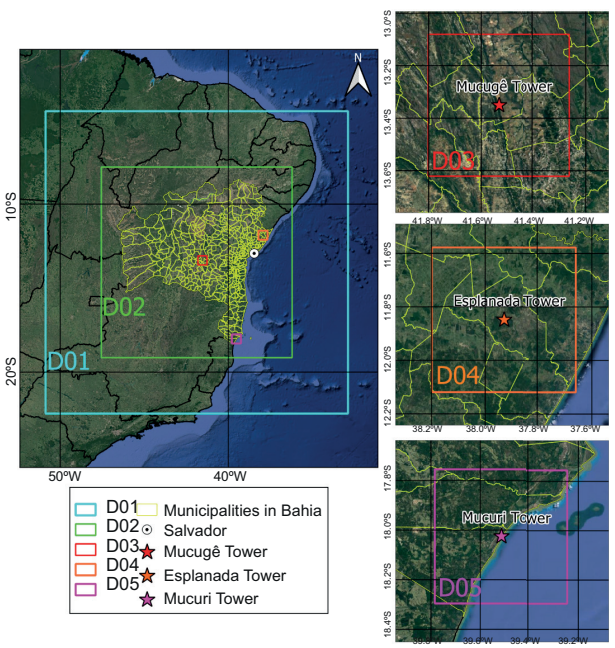

Fig. 4. Locations and distribution of anemometric towers and domains in the WRF model.

and 30'' corresponding to the others, which was made available to the default installation of the WRF model (Skamarock et al., 2008). The simulated winds were interpolated for heights of 80, 100, 120 and 150 m depending on the vertical profile of the WRF, based on hydrostatic pressure (ETA levels).

The simulations were carried out during December, 2015, and January, February, May, June, July, September, and October, 2016 (which are the months that have a representative percentage of observed data [> 75%] every hour). To obtain realistic initial conditions, a 24-hh spin-up was used, i.e., the simulation for each month was initialized from 00:00 UTC on the last day of the previous month. WRF outputs have been stored with hourly frequency, and the main model output parameters used were wind speed and direction.

The wind speed data acquisition was performed at three monitoring towers in the Bahia cities of Esplanada, Mucuri and Mucugê, as shown in Table I. The respective geographic locations are shown in Figure 4. Wind velocities are acquired by four anemometers located at different heights in the weather towers. The anemometers record a measurement at each level every 10 min and record an average every 60 min. The simulated data were also filtered in a similar way. The data was acquired using the Elipse E3 Supervisory Control and Data Acquisition (SCADA), which is a realtime SCADA platform for mission critical applications, being a well established SCADA platform, offering scalability and constant evolution for several types of applications, from simple HMI interfaces to complex operating centers in real time. It was developed to meet current and future connectivity the mean hourly values for the monitored data were computed. Suspicious or bad data (like negative values or a sequence of 20 or more equal-valued data) was treated as invalid data, which was discarded in order to produce valid samples. The anemometers are all Thies First Class.

3.3 Combinations of PBL and LSM schemes

The simulations tested the performances of all nine scenarios formed by the various combinations of three PBL schemes (MYJ, YSU, and ACM2) and three LSM schemes (Noah Land Surface Model [NLSM], RUC, and Noah-MP [multi-physics]), as presented in Table II. These PBL and LSM parameterization schemes were selected since they are the most common ones, as in the applications of Pei et al. (2014), Wharton et al. (2015), Lee et al. (2016), Salamanca et al. (2018), and Liu et al. (2019).

The other parameterization schemes were maintained constant during the simulations. The 5-class WRF Single-Moment (WSM5) microphysics scheme (Hong et al., 2004) was chosen following Kitagawa et al. (2017), who studied the Metropolitan Region of Salvador (RMS) in the state of Bahia and obtained good estimates for the variables of wind speed and direction. The Radiative Rapid Transfer Model (RRTM) and the Dudhia scheme were selected to estimate long-wave and short-wave radiation due to their efficiency and satisfactory performance in studies of wind resources in various regions of the world (Amjad et al., 2015; Mattar and Borvoran, 2016; Giannaros et al., 2017; Argüeso and Businger, 2018). Convective processes were represented using the Kain-Fritsch cumulus scheme in the external

Table II. Details of the test scenarios specifying the physical options.

Scenarios	1	2	3	4	5	6	7	8	9
PBL	MYJ	MYJ	MYJ	YSU	YSU	YSU	ACM2	ACM2	ACM2
Land surface model	NLSM	RUC	NOAH-MP	NLSM	RUC	NOAH-MP	NLSM	RUC	NOAH-MP
Surface layer	Eta	Eta	Eta	MM5	MM5	MM5	MM5	MM5	MM5
Cumulus	Kain-Fritsch								
Microphysics	WSM5								
Shortwave radiation	Dudhia								
Longwave radiation	RRTM								

PBL: planetary boundary layer; MYJ: Mellor-Yamada-Janjic PBL scheme; YSU: Yonsei University PBL scheme; ACM2: Asymmetric Convective Model 2; NLSM: Noah Land Surface Model; RUC: rapid update cycle; WSM5: the 5-class WRF single-moment; RRTM: Radiative Rapid Transfer Model.

domain, while it was disabled corresponding to other domains on the basis of the assumption that the majority of convective circulation is explicitly resolved. The physical options used are summarized in Table II, depicting the different scenarios.

3.4 Statistical evaluation

The performance of the model was evaluated using the statistical metrics of mean bias (MB), root mean squared error (RMSE), mean absolute gross error (MAGE), agreement index (IOA), and Pearson's correlation coefficient (R) (Carvalho et al., 2012, 2014; Cheng et al., 2013; Zempila et al., 2016; Gunwani and Mohan, 2017; Surussavadee, 2017a, b). In addition, Factor of 2 (Fac2) (Zucatelli et al., 2019) and standard deviation (SD) (Penchah et al., 2017) were also used. The average error, also called the average bias, is defined as follows:

$$MB = \bar{M} - \bar{O} \tag{5}$$

where $\bar{M}$ denotes the modeled average and $\bar{O}$ denotes the observed average.

Positive values of MB indicate an overestimation of the simulated data, while negative values imply underestimation. Positive/negative MB values corresponding to wind direction indicate that the modeled wind direction exhibits clockwise/counterclockwise overcorrection compared to the observations. It should be noted that, in the case of wind direction, the averages were calculated using scalar values and not vector values. The difference between each pair of simulated and observed wind directions (Δd) is defined as follows (Jiménez and Dudhia, 2013):

$$\text{if, } d_{WRF} - d_{obs} \leq \left|180\right|$$

$$\Delta d = d_{WRF} - d_{obs}$$

$$\text{if, } d_{WRF} - d_{obs} > 180$$

$$\Delta d = d_{WRF} - d_{obs} - 360$$

$$\text{if, } d_{WRF} - d_{obs} < -180$$

$$\Delta d = d_{WRF} - d_{obs} + 360$$

The RMSE expresses the total error of the model and exhibits a value of zero in ideal cases.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(M_i - O_i\right)^2} \tag{6}$$

In both of the aforementioned cases, lower values indicate better agreement between observed and modeled data. MAGE calculates the absolute mean error between simulated and observed values, expressing the average magnitude of the errors in the simulations.

$$MAGE = \frac{1}{n}\sum_{i=1}^{n}\left|M_i - O_i\right| \tag{7}$$

As in the case of RMSE, lower values of MAGE indicate greater similarity between the observed and modeled data series. IOA evaluates how close the simulated value is to that observed. The metric ranges from –1 to 1, with the value "1" meaning perfect agreement. The term c is a constant relative to the model's output frequency, which was assigned a value of 2 (Willmott et al., 2012). The variable P_i corresponds to the simulated ("predicted") value for instant (i), with O_i indicating the value observed at the same time i and $\bar{O}$ representing the average of the observed values.

$$IOA = 1 - \frac{\sum_{i=1}^{n}\left|P_i - O_i\right|}{c\sum_{i=1}^{n}\left|O_i - \bar{O}\right|}, \quad \text{where } \sum_{i=1}^{n}\left|P_i - O_i\right| \leq c\sum_{i=1}^{n}\left|O_i - \bar{O}\right| \tag{8}$$

$$IOA = \frac{c\sum_{i=1}^{n}\left|O_i - \bar{O}\right|}{\sum_{i=1}^{n}\left|P_i - O_i\right|} - 1, \quad \text{when } \sum_{i=1}^{n}\left|P_i - O_i\right| > c\sum_{i=1}^{n}\left|O_i - \bar{O}\right| \tag{9}$$

Pearson's correlation coefficient is a measure of linear association between modeled and observed data. Its value is zero in case of no correlation and the correlation increases as the value of the coefficient approaches –1 or +1. Values close to +1 imply positive correlation between the two variables, while those close to –1 imply negative correlation. It is defined as follows:

$$R = \frac{\sum_{i=1}^{n}[(M_i - \bar{M})(O_i - \bar{O})]}{\sqrt{\sum_{i=1}^{n}(M_i - \bar{M})^2 \sum_{i=1}^{n}(O_i - \bar{O})^2}} \quad (10)$$

Factor of 2 (Fac2) calculates the ratio between the simulated (M_i) and observed (O_i) data, indicating the percentage of data that should lie within the range $0.5 \leq M_i / O_i \leq 2$. Obviously, its value in the optimal case is 1.

4. Numerical results

Figure 5 depicts the average hourly time evolution of the observed and simulated wind speeds at elevations of 80, 100, 120, and 150 m, at the Mucuri, Esplanada, and Mucugê towers, obtained using the different combinations of PBL and LSM (see Table II). For the construction of hourly averages, all observed and simulated data records from the months studied were used.

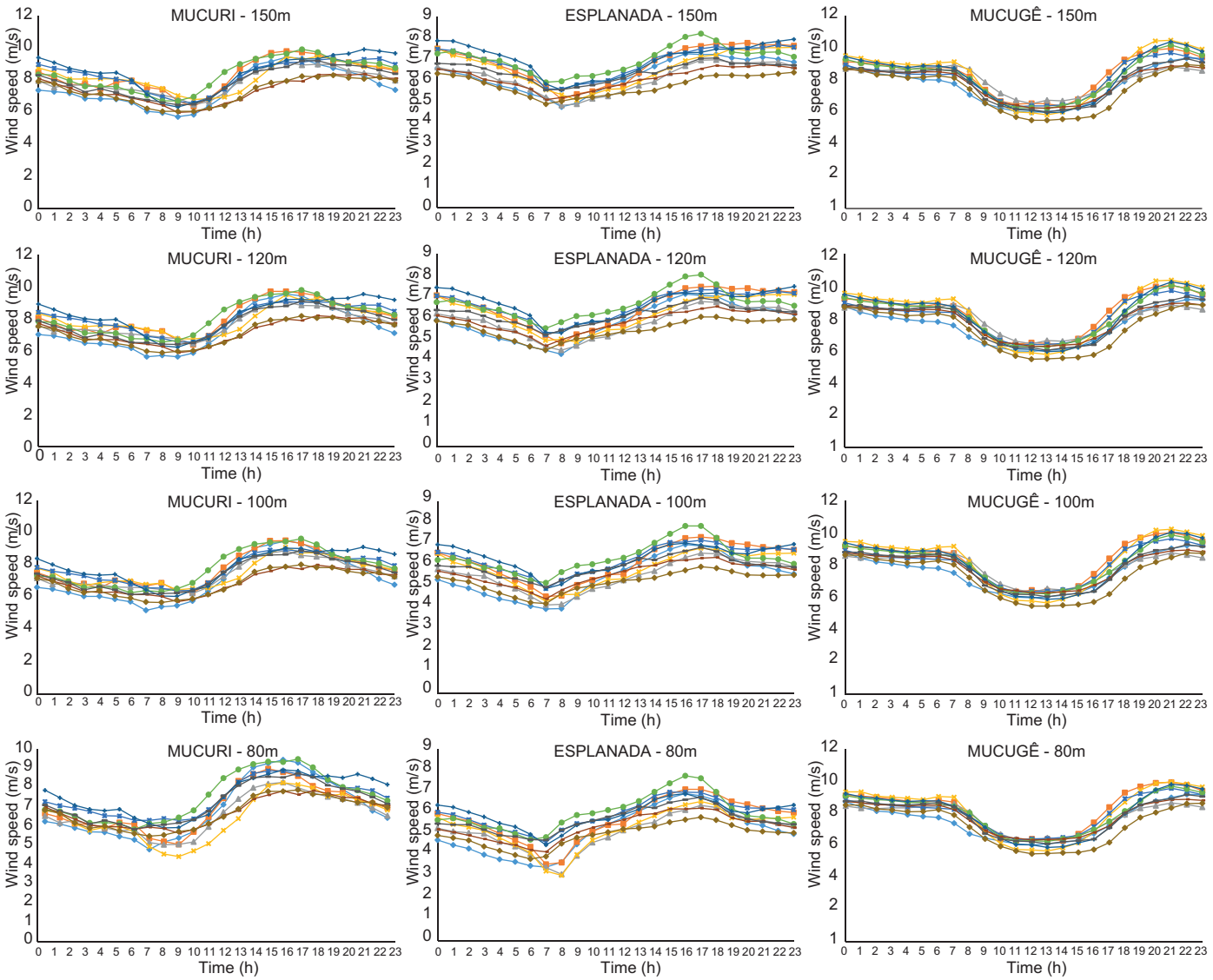

Fig. 5. Average hourly behavior of simulated and observed wind speeds.

It is evident from Figure 5 that the simulated results typically closely follow the trend of the curve of the observed average hourly values. In particular, the data corresponding to Mucugê clearly exhibits a greater congruence between the simulated and observed curves. It is also verified that the values of the average hourly wind speeds vary approximately between 6 and 10 ms^{-1} at Mucuri and Mucugê, and between 3 and 8 ms^{-1} at Esplanada.

The analysis of the graphs in Figure 5 reveal a globally similar trend among the simulated scenarios, indicating that the model is capable of capturing the general behavior of the wind speed, despite small differences in terms of time and magnitude. In order to estimate the scenario that corresponds to the best performances, validation was performed based on the statistics evaluations. In this section, the statistical metrics (MB, RMSE, MAGE, IOA, R, Fac2 and SD) are analyzed in order to estimate the discrepancies between the simulated and observed wind speed and direction corresponding to each case.

Tables III, IV, and V present the values of the statistical metrics obtained by comparing the simulated and observed data on wind speed corresponding to the Mucuri, Esplanada, and Mucugê towers, respec-

Table III. Results of statistical metrics corresponding to the comparison between simulated and observed wind speed data at Mucuri tower.

Height		S1	S2	S3	S4	S5	S6	S7	S8	S9	Obs
150 m	MB	0.846	0.454	0.742	0.854	0.885	1.024	**–0.131**	0.406	–0.227	
	OhitS	3	5	4	2	1	0	8	6	7	
	RMSE	3.187	3.339	3.629	3.139	**2.833**	3.170	3.029	2.895	2.951	
	OhitS	2	1	0	4	8	3	5	7	6	
	MAGE	2.261	2.430	2.614	2.256	**2.070**	2.226	2.289	2.135	2.234	
	OhitS	3	1	0	4	8	6	2	7	5	
	IOA	0.512	0.475	0.436	0.513	**0.553**	0.511	0.506	0.539	0.517	
	OhitS	4	1	0	5	8	3	2	7	6	
	R	0.518	0.450	0.378	0.521	**0.584**	0.512	0.471	0.530	0.472	
	OhitS	5	1	0	6	8	4	2	7	3	
	Fac2	87.8%	85.9%	85.3%	88.7%	**89.1%**	88.1%	87.1%	88.7%	87.2%	
	OhitS	4	1	0	7	8	5	2	7	3	
	SD	3.413	3.449	3.457	3.440	3.229	3.341	3.161	**3.146**	2.954	2.849
	OhitS	3	1	0	2	5	4	6	7	8	
120 m	MB	0.781	0.425	0.653	0.747	0.786	0.915	**–0.169**	0.341	–0.266	
	OhitS	2	5	4	3	1	0	**8**	6	7	
	RMSE	3.146	3.282	3.559	3.005	**2.735**	3.071	2.946	2.801	2.866	
	OhitS	2	1	0	4	**8**	3	5	7	6	
	MAGE	2.258	2.388	2.576	2.192	**2.003**	2.220	2.238	2.081	2.189	
	OhitS	2	1	0	5	**8**	4	3	7	6	
	IOA	0.507	0.478	0.437	0.521	**0.563**	0.515	0.511	0.545	0.522	
	OhitS	2	1	0	5	8	4	3	7	6	
	R	0.508	0.448	0.372	0.520	**0.593**	0.506	0.476	0.538	0.482	
	OhitS	5	1	0	6	**8**	4	2	7	3	
	Fac2	86.9%	85.5%	84.4%	88.4%	**88.6%**	87.8%	86.0%	88.4%	86.9%	
	OhitS	4	1	0	7	**8**	5	2	7	4	
	SD	3.308	3.349	3.345	3.230	3.126	3.175	3.025	3.022	**2.834**	2.859
	OhitS	2	0	1	3	5	4	6	7	8	

MB: mean bias; RMSE: root mean squared error; MAGE: mean absolute gross error; IOA: agreement index; R: Pearson's correlation coefficient; Fac2: Factor of 2; SD: standard deviation; Ohits: objective hit score; S: station; Obs: observed. The best results are highlighted in bold.

Table III. Results of statistical metrics corresponding to the comparison between simulated and observed wind speed data at Mucuri tower.

Height		S1	S2	S3	S4	S5	S6	S7	S8	S9	Obs
100 m	MB	0.649	0.368	0.485	0.668	0.726	0.823	**–0.158**	0.330	–0.259	
	OhitS	3	5	4	2	1	0	8	6	7	
	RMSE	3.080	3.226	3.460	2.890	**2.663**	2.976	2.907	2.762	2.829	
	OhitS	2	1	0	5	**8**	3	4	7	6	
	MAGE	2.222	2.354	2.528	2.133	**1.953**	2.164	2.215	2.064	2.172	
	OhitS	2	1	0	6	8	5	3	7	4	
	IOA	0.513	0.484	0.446	0.532	**0.572**	0.526	0.514	0.547	0.524	
	OhitS	2	1	0	6	8	5	3	7	4	
	R	0.501	0.448	0.367	0.530	**0.604**	0.511	0.477	0.540	0.485	
	OhitS	4	1	0	6	8	5	2	7	3	
	Fac2	86.0%	85.2%	83.7%	88.5%	**88.7%**	87.4%	85.8%	88.6%	86.9%	
	OhitS	3	1	0	6	8	5	2	7	4	
	SD	3.203	3.256	3.201	3.089	3,070	3.058	**2.947**	2.954	2.769	2.866
	OhitS	1	0	2	3	4	5	8	7	6	
80 m	MB	**0.071**	–0.245	–0.431	0.577	0.667	0.714	–0.142	0.322	–0.251	
	OhitS	8	6	3	2	1	0	7	4	5	
	RMSE	**2.540**	2.685	2.835	2.767	2.593	2.865	2.858	2.717	2.783	
	OhitS	8	6	1	4	7	0	2	5	3	
	MAGE	1.914	2.058	2.172	2.058	**1.913**	2.091	2.187	2.043	2.146	
	OhitS	7	5	1	5	8	3	2	6	0	
	IOA	0.580	0.549	0.524	0.549	**0.581**	0.542	0.521	0.552	0.530	
	OhitS	7	5	1	5	8	3	0	6	2	
	R	0.585	0.540	0.483	0.547	**0.616**	0.524	0.483	0.546	0.493	
	OhitS	7	4	1	5	8	3	1	6	2	
	Fac2	87.3%	86.7%	84.3%	88.3%	**88.7%**	86.8%	85.1%	87.5%	86.7%	
	OhitS	5	3	0	7	8	4	1	6	3	
	SD	2.827	2.808	2.690	2.968	3.029	2.951	2.873	2.894	2.710	2.876
	OhitS	6	5	0	3	2	4	8	7	1	
Sum of scores		108	65	22	128	179	94	107	184	128	

MB: mean bias; RMSE: root mean squared error; MAGE: mean absolute gross error; IOA: agreement index; R: Pearson's correlation coefficient; Fac2: Factor of 2; SD: standard deviation; Ohits: objective hit score; S: station; Obs: observed. The best results are highlighted in bold.

Table IV. Results of statistical metrics corresponding to the comparison between simulated and observed wind speed data at Esplanada tower.

Height		S1	S2	S3	S4	S5	S6	S7	S8	S9	Obs
150 m	MB	0.613	–0.160	0.292	0.614	0.730	0.737	–0.171	**0.152**	–0.506	
	OhitS	3	7	5	2	1	0	6	8	4	
	RMSE	2.007	2.193	1.934	2.098	**1.898**	2.029	2.151	1.990	2.074	
	OhitS	5	0	7	2	8	4	1	6	3	
	MAGE	1.448	1.679	1.486	1.574	**1.453**	1.521	1.673	1.491	1.625	
	OhitS	7	0	6	3	8	4	1	5	2	
	IOA	0.446	0.384	0.455	0.423	**0.467**	0.442	0.386	0.453	0.404	
	OhitS	5	0	7	3	8	4	1	6	2	
	R	0.492	0.409	0.465	0.509	**0.530**	0.511	0.398	0.423	0.400	
	OhitS	5	2	4	6	8	7	0	3	1	
	Fac2	93.9%	88.9%	93.2%	93.1%	**94.6%**	94.2%	89.6%	92.2%	90.1%	
	OhitS	6	0	5	4	8	7	1	3	2	
	SD	2.133	2.242	1.976	2.245	**1.912**	2.068	2.154	1.980	1.961	1.708
	OhitS	3	1	6	0	8	4	2	5	7	

MB: mean bias; RMSE: root mean squared error; MAGE: mean absolute gross error; IOA: agreement index; R: Pearson's correlation coefficient; Fac2: Factor of 2; SD: standard deviation; Ohits: objective hit score; S: station; Obs: observed. The best results are highlighted in bold.

Table IV. Results of statistical metrics corresponding to the comparison between simulated and observed wind speed data at Esplanada tower.

Height		S1	S2	S3	S4	S5	S6	S7	S8	S9	Obs
120 m	MB	0.721	–0.026	0.378	0.752	0.824	0.831	**–0,003**	0.319	–0.372	
	OhitS	3	7	4	2	1	0	8	6	5	
	RMSE	2.062	2.174	1.993	2.138	**1.934**	2.081	2.115	1.992	2.023	
	OhitS	4	0	6	1	8	3	2	7	5	
	MAGE	1.540	1.668	1.535	1.609	**1.486**	1.576	1.648	1.493	1.586	
	OhitS	5	0	6	2	8	4	1	7	3	
	IOA	0.434	0.389	0.438	0.411	**0.456**	0.423	0.396	0.453	0.419	
	OhitS	5	0	6	2	8	4	1	7	3	
	R	0.476	0.400	0.438	0.489	**0.520**	0.480	0.390	0.408	0.391	
	OhitS	5	2	4	7	8	6	0	3	1	
	Fac2	92.4%	87.5%	91.8%	91.5%	**93.2%**	92.2%	88.3%	90.7%	88.9%	
	OhitS	7	0	5	4	8	6	1	3	2	
	SD	2.115	2.208	1.975	2.186	**1.877**	2.009	2.097	1.917	1.906	1.700
	OhitS	2	0	5	1	8	4	3	6	7	
100 m	MB	0.750	**0.049**	0.354	0.833	0.882	0.869	0.134	0.456	–0.251	
	OhitS	3	8	5	2	0	1	7	4	6	
	RMSE	2.078	2.139	2.010	2.148	**1.954**	2.101	2.076	1.989	1.978	
	OhitS	3	1	5	0	8	2	4	6	7	
	MAGE	1.557	1.651	1.547	1.623	1.504	1.598	1.617	**1.492**	1.551	
	OhitS	4	0	6	1	7	3	2	8	5	
	IOA	0.436	0.403	0.441	0.413	0.456	0.422	0.416	**0.461**	0.439	
	OhitS	4	0	6	1	7	3	2	8	5	
	R	0.473	0.404	0.422	0.486	**0.527**	0.466	0.401	0.416	0.399	
	OhitS	6	2	4	7	8	5	1	3	0	
	Fac2	90.6%	86.1%	90.1%	89.4%	**91.4%**	90.5%	87.4%	89.6%	88.0%	
	OhitS	7	0	5	3	8	6	1	4	2	
	SD	2.109	2.161	1.956	2.139	**1.877**	1.969	2.053	1.878	1.871	1.713
	OhitS	2	0	5	1	7	4	3	6	8	
80 m	MB	0.606	**–0.022**	0.120	0.829	0.880	0.824	0.202	0.533	–0.184	
	OhitS	3	8	7	1	0	2	5	4	6	
	RMSE	1.990	2.055	**1.898**	2.092	1.935	2.062	2.033	1.975	1.942	
	OhitS	4	1	8	0	7	2	3	5	6	
	MAGE	1.495	1.593	**1.476**	1.589	1.489	1.572	1.583	1.488	1.524	
	OhitS	5	0	8	1	6	3	2	7	4	
	IOA	0.482	0.447	**0.488**	0.449	0.484	0.455	0.451	0.484	0.471	
	OhitS	5	0	8	1	7	3	2	7	4	
	R	0.494	0.433	0.454	0.503	**0.554**	0.473	0.431	0.446	0.423	
	OhitS	6	2	4	7	8	5	1	3	0	
	Fac2	88.8%	84.9%	88.3%	87.7%	**90.0%**	88.5%	86.1%	88.5%	87.0%	
	OhitS	7	0	4	3	8	6	1	6	2	
	SD	2.067	2.081	1.868	2.067	**1.891**	1.913	2.017	1.859	1.845	1.770
	OhitS	1	0	6	2	5	4	3	7	8	
Sum of scores		125	41	157	69	184	106	65	153	110	

MB: mean bias; RMSE: root mean squared error; MAGE: mean absolute gross error; IOA: agreement index; R: Pearson's correlation coefficient; Fac2: Factor of 2; SD: standard deviation; Ohits: objective hit score; S: station; Obs: observed. The best results are highlighted in bold.

Table V. Results of statistical metrics corresponding to the comparison between simulated and observed wind speed data at Mucugê tower.

Height		S1	S2	S3	S4	S5	S6	S7	S8	S9	Obs
150 m	MB	0.700	0.364	0.698	0.420	0.519	0.528	**0.138**	0.159	–0.289	
	OhitS	0	5	1	4	3	2	8	7	6	
	RMSE	2.932	3.237	2.857	2.682	2.567	**2.496**	2.874	2.781	2.622	
	OhitS	1	0	3	5	7	8	2	4	6	
	MAGE	2.280	2.559	2.229	2.144	2.026	**1.943**	2.272	2.197	2.088	
	OhitS	2	0	3	5	7	8	1	4	6	
	IOA	0.450	0.409	0.485	0.505	0.532	**0.552**	0.476	0.493	0.519	
	OhitS	1	0	3	5	7	8	2	4	6	
	R	0.614	0.519	0.619	0.649	0.650	**0.662**	0.598	0.609	0.621	
	OhitS	2	0	4	6	7	8	1	3	5	
	Fac2	88.4%	84.1%	88.8%	89.2%	91.5%	**92.3%**	86.2%	87.4%	87.7%	
	OhitS	4	0	5	6	7	8	1	2	3	
	SD	3.700	3.650	3.491	3.467	3.275	**3.220**	3.536	3.448	3.253	2.691
	OhitS	0	1	3	4	6	8	2	5	7	
120 m	MB	0.857	0.440	0.812	0.581	0.647	0.640	0.311	0.321	**–0.169**	
	OhitS	0	5	1	4	2	3	7	6	8	
	RMSE	2.915	3.184	2.842	2.654	2.547	**2.486**	2.829	2.745	2.559	
	OhitS	1	0	2	5	7	8	3	4	6	
	MAGE	2.255	2.513	2.216	2.113	2.004	**1.934**	2.229	2.160	2.029	
	OhitS	1	0	3	5	7	8	2	4	6	
	IOA	0.440	0.402	0.473	0.498	0.523	**0.540**	0.470	0.486	0.517	
	OhitS	1	0	2	5	7	8	3	4	6	
	R	0.600	0.502	0.602	0.639	0.640	**0.648**	0.558	0.596	0.607	
	OhitS	3	0	4	6	7	8	1	2	5	
	Fac2	89.0%	84.1%	88.6%	89.6%	91.5%	**92.4%**	86.9%	88.2%	88.6%	
	OhitS	5	0	4	6	7	8	1	2	4	
	SD	3.567	3.523	3.366	3.336	3.158	**3.097**	3.421	3.328	3.123	2.605
	OhitS	0	1	3	4	6	8	2	5	7	
100 m	MB	0.779	0.306	0.692	0.517	0.553	0.517	0.264	**0.261**	–0.262	
	OhitS	0	5	1	4	2	4	6	8	7	
	RMSE	2.864	3.142	2.793	2.620	2.507	**2.435**	2.811	2.728	2.557	
	OhitS	1	0	3	5	7	8	2	4	6	
	MAGE	2.210	2.473	2.172	2.079	1.963	**1.891**	2.209	2.143	2.033	
	OhitS	1	0	3	5	7	8	2	4	6	
	IOA	0.450	0.412	0.483	0.506	0.533	**0.550**	0.475	0.491	0.517	
	OhitS	1	0	3	5	7	8	2	4	6	
	R	0.584	0.484	0.585	0.624	0.627	**0.636**	0.572	0.578	0.590	
	OhitS	3	0	4	6	7	8	1	2	5	
	Fac2	89.4%	84.4%	88.8%	90.1%	91.8%	**92.7%**	87.3%	88.6%	89.0%	
	OhitS	5	0	3	6	7	8	1	2	4	
	SD	3.449	3.408	3.261	3.222	3.053	**2.979**	3.323	3.224	3.017	2.607
	OhitS	0	1	3	5	6	8	2	4	7	

MB: mean bias; RMSE: root mean squared error; MAGE: mean absolute gross error; IOA: agreement index; R: Pearson's correlation coefficient; Fac2: Factor of 2; SD: standard deviation; Ohits: objective hit score; S: station; Obs: observed. The best results are highlighted in bold.

Table V. Results of statistical metrics corresponding to the comparison between simulated and observed wind speed data at Mucugê tower.

Height		S1	S2	S3	S4	S5	S6	S7	S8	S9	Obs
80 m	MB	0.736	**0.209**	0.582	0.499	0.503	0.417	0.268	0.246	–0.307	
	OhitS	0	8	1	3	2	4	6	7	5	
	RMSE	2.786	3.056	2.703	2.589	2.475	**2.376**	2.794	2.708	2.547	
	OhitS	2	0	4	5	7	8	1	3	6	
	MAGE	2.145	2.408	2.099	2.047	1.928	**1.843**	2.188	2.123	2.025	
	OhitS	2	0	4	5	7	8	1	3	6	
	IOA	0.454	0.416	0.491	0.503	0.532	**0.553**	0.469	0.485	0.509	
	OhitS	1	0	4	5	7	8	2	3	6	
	R	0.567	0.466	0.568	0.605	0.609	**0.620**	0.548	0.554	0.566	
	OhitS	4	0	5	6	7	8	1	2	3	
	Fac2	89.7%	84.9%	88.8%	90.3%	91.5%	**92.6%**	87.3%	88.7%	89.1%	
	OhitS	5	0	3	6	7	8	1	2	4	
	SD	3.285	3.245	3.097	3.093	2.936	**2.836**	3.204	3.099	2.892	2.557
	OhitS	0	1	4	5	6	8	2	3	7	
Sum of scores		46	27	86	141	173	205	66	107	159	

MB: mean bias; RMSE: root mean squared error; MAGE: mean absolute gross error; IOA: agreement index; R: Pearson's correlation coefficient; Fac2: Factor of 2; SD: standard deviation; Ohits: objective hit score; S: station; Obs: observed. The best results are highlighted in bold.

tively. Superior results have been highlighted in bold and, to evaluate simulations results, the objective hit Score (OhitS) (Penchah et al., 2017) was used. As mentioned previously, low values are better for MB, RMSE and MAGE, and high values are better for IOA, R, and Fac2 in terms of prediction accuracy. In the OhitS method, for SD, each parameterization that has the closest result to the corresponding observation gets an 8 score. The second-best parameterization gets a 7 score, and so on, the worst result gets a 0 score. For MB, RMSE and MAGE, parameterizations gain scores from 8 to 0 for minimum to maximum, respectively. For IOA, R and Fac2, the opposite occurs.

Based on the comparison between measurements obtained in the three towers, the common parameterization with the highest score was obtained by scenario 5 (S5) (parameterization YSU and RUC): 184 points in Esplanada, 179 points in Mucuri, and 173 points in Mucugê. In the meantime, scenario 6 (S6) showed the best performance corresponding to Mucugê (parameterizations YSU and NOAH-MP). However, the results obtained by this scenario were as good as those obtained by scenario 5 in all cases. The underestimation of wind speed only occurred in the simulations performed with ACM2-Noah-MP for all levels and towers, and with ACM2-LSM at all levels for Mucuri and at 120 and 150 m in height for Esplanada. Tyagi et al. (2018) also found that ACM2 underestimated wind speed, while MYJ and YSU overestimated this parameter.

These metrics were also evaluated corresponding to the variable of wind direction. Corresponding to Mucugê, as in the case of wind speed, the combination of the YSU scheme and the NOAH-MP soil surface scheme (S6) produced the best estimates. However, Esplanada and Mucuri exhibited different results compared to the case of wind speed. Corresponding to Mucuri, the combination of MYJ atmospheric boundary layer parameterization and the Noah (S1) model performed better overall compared to other combinations. Finally, corresponding to Esplanada, scenario 3 (S3), which comprises the combination of MYJ and NOAH-MP, obtained a better result than the other combinations.

From the perspective of wind energy, wind speed is a more important variable than wind direction. This can be primarily attributed to the existence of a system that guides the wind turbine rotor in the

direction of the wind and the automatic regulation of the slopes of the blades to optimize the incidence of wind. Thus, the results obtained from the analysis of wind speed is more pertinent to simulations with high resolution domain for the entire state of Bahia, given that it produces a high amount of wind energy. Based on the aforementioned discussion, further analyses of the parametrization performance of the combination of the CLP-YSU and LSM-RUC schemes in terms of wind speed follow.

Figures 6, 7, and 8 illustrate the wind roses obtained using YSU and RUC (scenario 5) as parameterizations of PBL and LSM, respectively, corresponding to the three towers: Mucuri (Fig. 6), Esplanada (Fig. 7) and Mucugê (Fig. 8), based on all data generated during the simulated months.

It is evident from the wind rose diagrams that the studied regions exhibit high potential for wind power generation with intermediate and high wind speed values between 6 to 9 ms^{-1} (yellow) and above 9 ms^{-1} (brown), respectively. It is also apparent that the wind speed was slightly overestimated at all three towers. This agrees with the observations of Giannaros et al. (2017), who highlighted the tendency of the WRF model to overestimate the frequency of strong winds; Stucki et al. (2016), who reported that the WRF model tends to overestimate average winds; and Carvalho et al. (2014), who also reported similar results.

As this study uses data from towers located at three different sites equipped with anemometers at varying heights (80, 100, 120 and 150 m), its observations can be considered to be representative in terms of wind energy. Thus, it is interesting to compare these data with the vertical profiles of wind obtained via the simulations. Figures 9, 10, and 11 depict the vertical profiles of the monthly average wind speeds simulated using scenario 5 (YSU and RUC), for two representative months each corresponding to the dry and rainy periods, compared to data observed at the Esplanada, Mucugê, and Mucuri towers, respectively. Table VI specifies the two representative months for dry and rainy periods for each tower.

Data corresponding to Esplanada is observed to be well characterized with two very distinct periods and abundant rain during the months of May and June (see Fig. 1). The average monthly speed is overestimated by the WRF model at all heights during May (Fig. 9b); however, the maximum difference between simulated and observed data is observed to be 0.64 ms^{-1}. During the less rainy period in January, the difference is observed to be 1.34 ms^{-1} and the tendency of overestimation is also observed (Fig. 9a).

Data corresponding to Mucugê (the farthest from the sea) is also well characterized with two very distinct periods and very little rain during the months of July and August (see Fig. 2). The average monthly speed is overestimated by the WRF model at all heights during July, with a maximum difference of 1.34 ms^{-1} (Fig. 10b). During the rainiest period in January, the maximum difference is observed to be 0.22 ms^{-1}, but a slight tendency of overestimation persists (Fig. 10a).

Data corresponding to Mucuri (the closest to the sea) has been uniform in terms of annual rainfall over the most recent decades (see Fig. 3); however, January and February are representative of the dry season in the region. During this period, the average monthly speed is overestimated by the WRF model at all heights during January, with a maximum difference of 1.66 ms^{-1} (Fig. 11a). During May, which is representative of the rainy season in the region, the maximum difference is observed to be 0.27 ms^{-1} (Fig. 11b).

Therefore, consideration of only the months analyzed disaggregated from the others reveals a greater tendency of overestimation of the monthly average wind speed corresponding to the dry season.

Having identified the best performing combination to be the YSU-RUC scheme, the superiority of its performance can be verified with respect to each location, in terms of the distance from the sea (Table VII), based on the evaluated statistics corresponding to all the simulated months. Unlike tables III, IV and V that compare 9 scenarios, in table VII the three towers are compared, therefore, the scores ranged between 2 and 0. SD was analyzed by the difference between simulated and observed SD (diffSD).

An analysis of Table VII and Figure 5 reveals that proximity to the sea degrades the accuracy of simulation. Mucugê is observed to exhibit a greater congruence between simulated and observed wind speeds (Fig. 5). This tower, located in the center of the state of Bahia (280 km from the sea), exhibits a statistically similar performance to that at Esplanada (40 km from the coast), and corresponds to

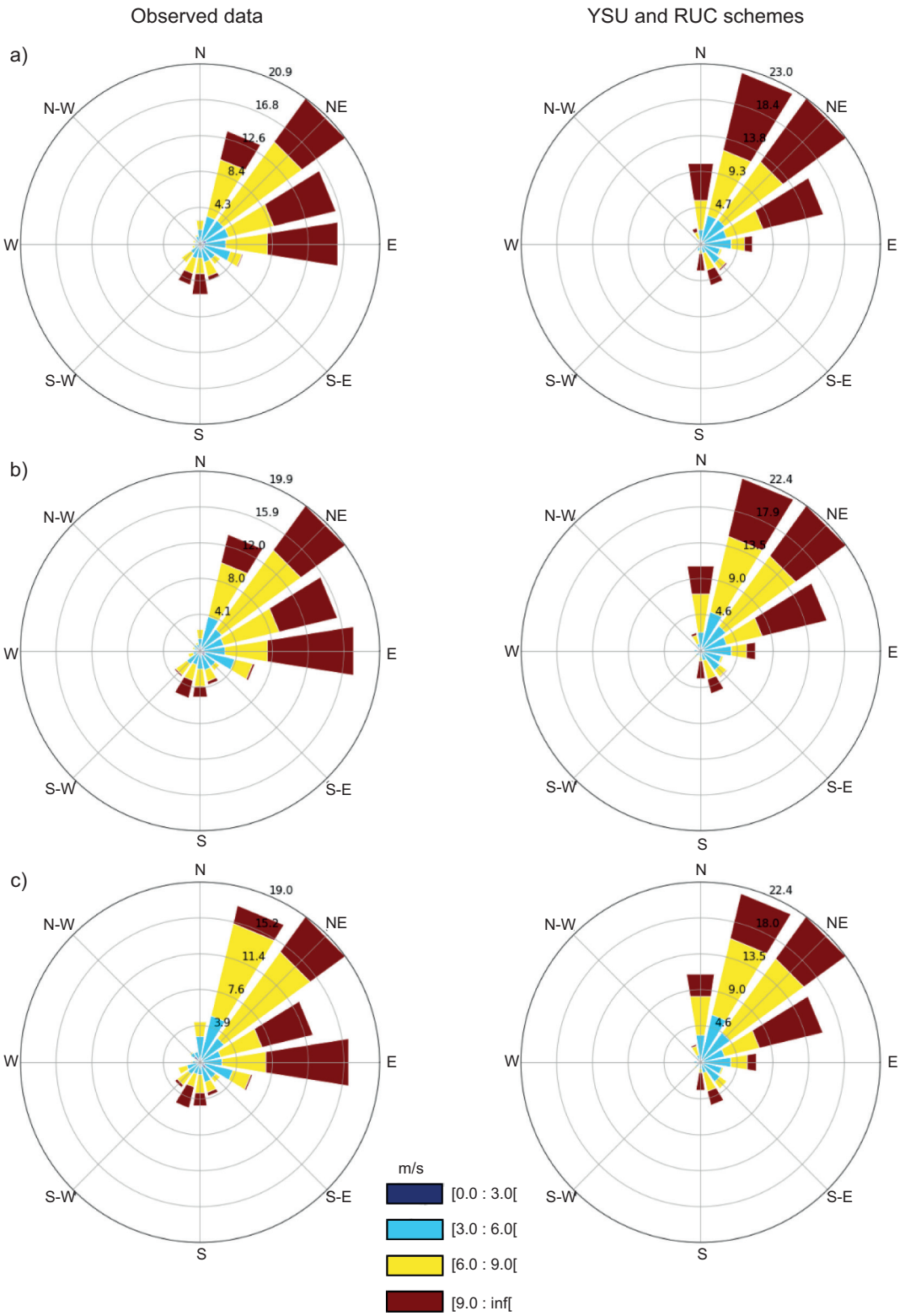

Fig. 6. Comparison of wind roses obtained using the YSU and RUC schemes from the Mucuri tower based on observations at heights of (a) 150 m, (b) 120 m, and (c) 100 m.

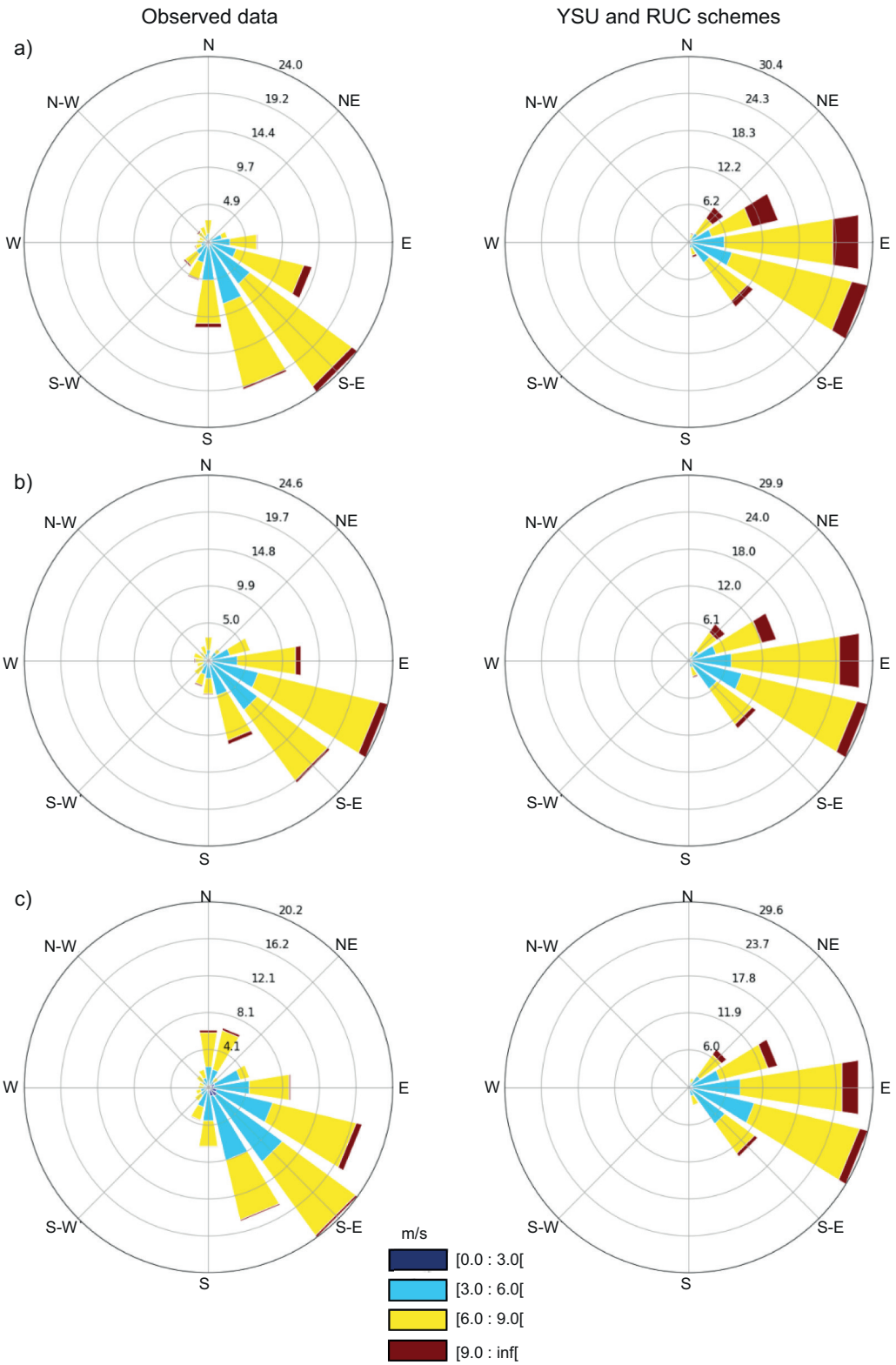

Fig. 7. Comparison of wind roses obtained using the YSU and RUC schemes from the Esplanada tower based on observations at heights of (a) 150 m, (b) 120 m, and (c) 100 m.

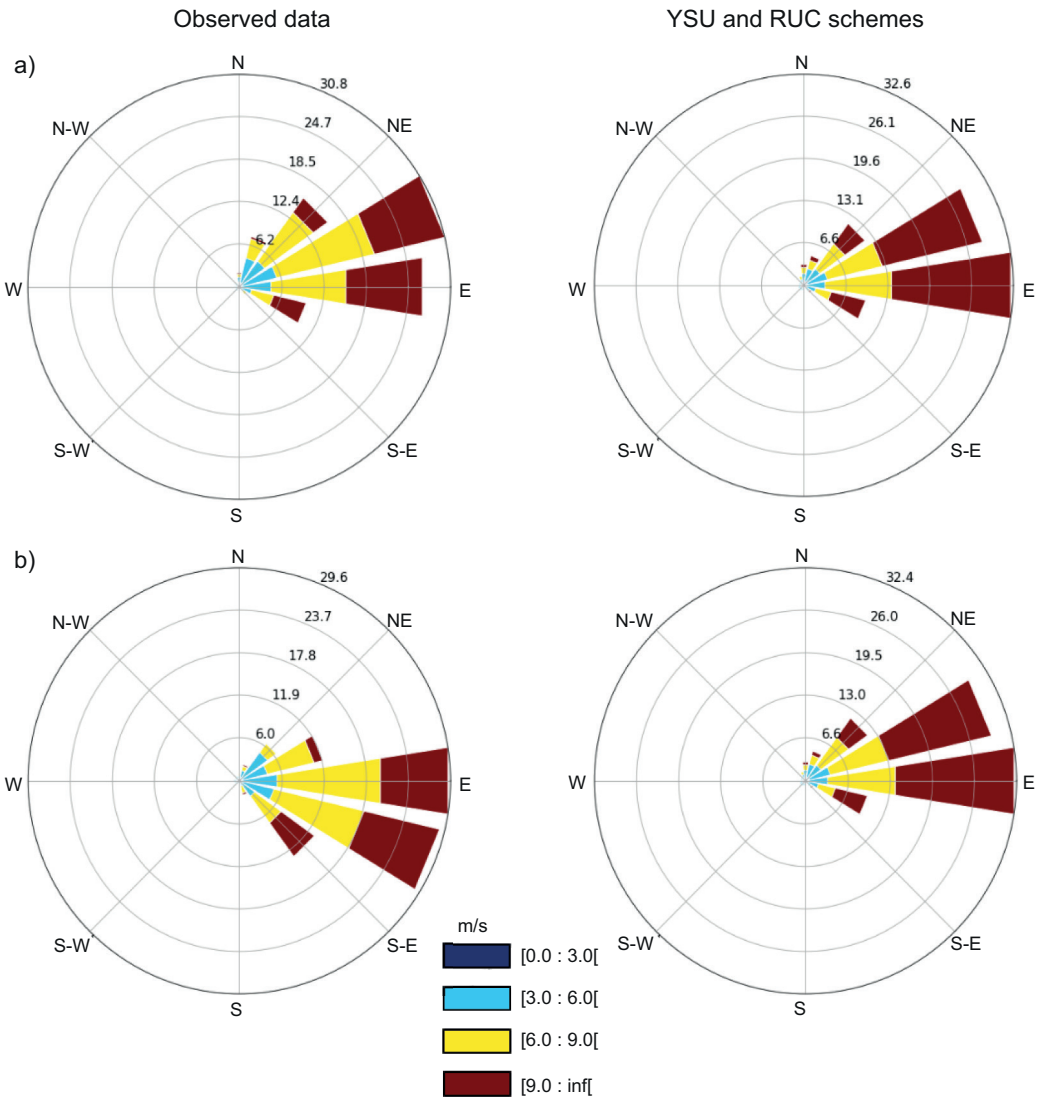

Fig. 8. Comparison of wind roses obtained using the YSU and RUC schemes from the Mucugê tower based on observations at heights of (a) 120 m, and (b) 100 m.

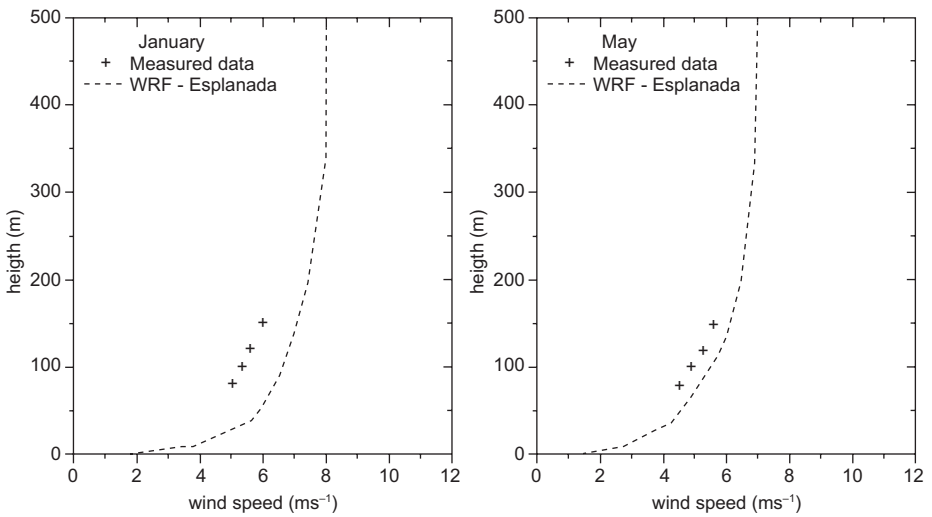

Fig. 9. Vertical profiles of the monthly average wind speed measured by the Esplanada tower and simulated using WRF with the parameters of scenario 5 in (a) January, and (b) May.

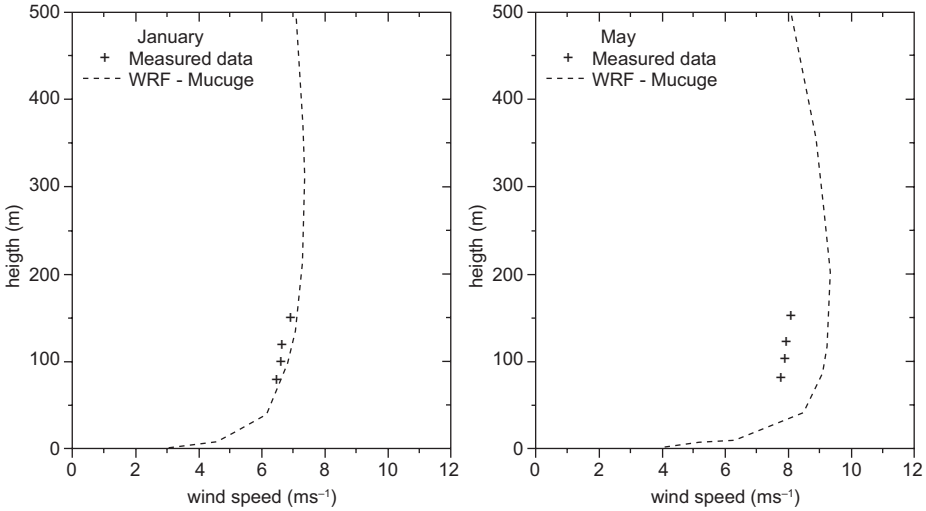

Fig. 10. Vertical profiles of the monthly average wind speed measured by the Mucugê tower and simulated using WRF with the parameterization of scenario 5 in (a) January, and (b) July.

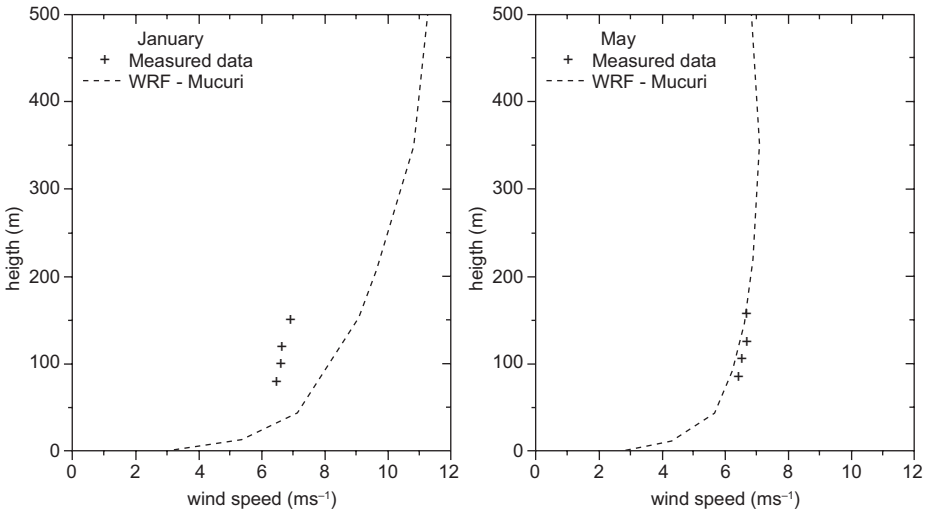

Fig. 11. Vertical profiles of the monthly average wind speed measured by the Mucuri tower and simulated using WRF with the parameterization of scenario 5 in (a) January, and (b) May.

Table VI. Representative months for dry and rainy periods for each tower.

		Esplanada	Mucuri	Mucugê
Month	Rainy	May	May	January
	Dry	January	January	July

Table VII. Wind speed statistics obtained using the YSU and RUC schemes.

Height		Esplanada	Mucuri	Mucugê
150 m	MB	0.730	0.885	**0.519**
	OhitS	1	0	2
	RMSE	**1.898**	2.833	2.567
	OhitS	2	0	1
	MAGE	**1.453**	2.070	2.026
	OhitS	2	0	1
	IOA	0.467	**0.553**	0.532
	OhitS	0	2	1
	R	0.530	0.584	**0.650**
	OhitS	0	1	2
	Fac2	**94.6%**	89.1%	91.5%
	OhitS	2	0	1
	diffSD	**0.204**	0.379	0.584
	OhitS	2	1	0
120 m	MB	0.824	0.786	**0.647**
	OhitS	0	1	2
	RMSE	**1.934**	2.735	2.547
	OhitS	2	0	1
	MAGE	**1.486**	2.003	2.004
	OhitS	2	1	0
	IOA	0.456	**0.563**	0.523
	OhitS	0	2	1
	R	0.520	0.593	**0.640**
	OhitS	0	1	2
	Fac2	**93.2%**	88.6%	91.5%
	OhitS	2	0	1
	diffSD	0.177	0.267	0.552
	OhitS	2	1	0

Height		Esplanada	Mucuri	Mucugê
100 m	MB	0.882	0.726	**0.553**
	OhitS	0	1	2
	RMSE	**1.954**	2.663	2.507
	OhitS	2	0	1
	MAGE	**1.504**	1.953	1.963
	OhitS	2	1	0
	IOA	0.456	**0.572**	0.533
	OhitS	0	2	1
	R	0.527	0.604	**0.627**
	OhitS	0	1	2
	Fac2	91.4%	88.7%	**91.8%**
	OhitS	1	0	2
	diffSD	0.164	0.204	0.447
	OhitS	2	1	0
80 m	MB	0.880	0.667	**0.503**
	OhitS	0	1	2
	RMSE	**1.935**	2.593	2.475
	OhitS	2	0	1
	MAGE	**1.489**	1.913	1.928
	OhitS	2	1	0
	IOA	0.484	**0.581**	0.532
	OhitS	0	2	1
	R	0.554	**0.616**	0.609
	OhitS	0	2	1
	Fac2	90.0%	88.7%	**91.5%**
	OhitS	1	0	2
	diffSD	0.122	0.152	0.379
	OhitS	2	1	0
Sum of scores		31	23	30

MB: mean bias; RMSE: root mean squared error; MAGE: mean absolute gross error; IOA: agreement index; R: Pearson's correlation coefficient; Fac2: Factor of 2; SD: standard deviation; Ohits: objective hit score; S: station; Obs: observed. The best results are highlighted in bold.

a better performance than that recorded at Mucuri (tower closest to the sea, 340 m). This phenomenon can be explained by the differentiated behavior of the model in regions close to large bodies of water compared to more inland areas. At places near the sea, variations in atmospheric conditions are more localized in time and space, due to ill-understood effects of land and sea breezes that are not accounted for by atmospheric models (Salvador et al., 2016b). Proximity to the sea, the interactions between atmospheric flow in the marine boundary layer, and the development of a boundary layer on the continent remain critical factors in the predictions of such models. Further, the adequate simulation of the flow characteristics related to the breeze inlet marine life in coastal regions persists as one of the major challenges for meteorological models (Shin e Hong, 2011; Cheng et al., 2012; de León and Orfila, 2013). It should be noted that the model interpreted the grid cell containing the Mucuri tower as land, which refutes the hypothesis that its poor performance is induced by the mistaken identification of the site as oceanic (given its close proximity to the sea), and supports the hypothesis that proximity to the sea negatively influences the performance of simulations. Tyagi et al. (2018) found a similar result by

indicating that inland locations are better simulated than sea locations, and indicated the reasons for this difference in the fact that performance is likely related to difficulty in reproducing vertical variability at sea, where very shallow PBL development may not be entirely reproduced. In addition, simulated PBL depths at sea locations exhibit a larger bias with respect to observations at inland locations with well-developed convective boundary layers.

Assessment of the wind speed statistics corresponding to periods of more and less rain also reveals scenario 5 (YSU and RUC) to be well-performing. It is interesting to note that the most accurate estimates for wind speed corresponding to the period of less rain in the simulated months in Mucugê (May, June, July, and September), were identical to the one obtained via aggregate analysis (parameterizations YSU and NOAH-MP [S6]). Analysis of the rainy period (January and February) also revealed better estimates corresponding to scenario 5 (YSU and RUC). Following the disaggregation of the two periods (more and less rain) at Esplanada, the data corresponding to the least rainy period (December, January, February, September, and October) continued to support the finding that scenario 5 exhibited the best performance, while better indicators were obtained corresponding to the rainy period (May and June) using scenario 1 (MYJ and NLSM). Corresponding to Mucuri, scenario 5 (YSU and RUC) continued to emerge as the best option for the drier month (January), while the rainy months (December and May) were better characterized by scenario 8 (ACM2 and RUC). However, the results obtained by these scenarios were as good as those obtained by scenario 5 in all cases. Thus, the aforementioned results corroborate the choice of the combination of the PBL-YSU scheme with the LSM-RUC scheme for application in the forecast of wind energy production for the region.

5. Summary and conclusions

This study aimed to evaluate the performance of the various combinations of three PBL parameterization schemes and three LSM parameterization schemes, using the WRF model for a tropical region, in order to identify the optimal parameters to be applied in the analysis of wind energy production based on numerical modeling.

The WRF model was verified to be capable of capturing the general behavior of wind speed. The combination of the YSU and RUC parameterization schemes exhibited the best performance. As expected, the simple scheme of thermal diffusion of the soil, RUC, was better adapted to the case of Bahia, as this region does not experience snowy climate. The RUC scheme was also observed to fit into the intermediate level of complexity.

It is important to highlight that wind speed was overestimated in the simulations and, in general, estimated wind directions were quite similar to the observed data. Further, it was observed that proximity to the ocean degraded the accuracy of the simulations. Thus, the towers at Mucugê and Esplanada, which were farthest from the sea, exhibited the best results for all statistical metrics with respect to wind speed. However, the average hourly time evolution of observed and simulated wind speeds at Mucugê exhibited less dissonance.

Finally, the disaggregated analysis of the results corresponding to rainy and dry periods at each tower revealed YSU-RUC to be the clear front-runner. This fact supports the application of the combination of CLP-YSU and LSM-RUC to the high-resolution simulation of wind energy production for the entire state of Bahia. This should be the next step in order to obtain a high-resolution wind map with verified parameterization that exhibits good results.

Acknowledgments

The authors are grateful to the Supercomputing Center for Industrial Innovation (CIMATEC) for providing the necessary computational structure to perform the simulations; and to the Research Support Foundation of the State of Bahia (FAPESB) and the Coordination for the Improvement of Higher Education Personnel (CAPES) for partially financing the work.

References

ABEEólica. 2017. Boletim anual de geração eólica. Associação Brasileira de Energia Eólica e Novas Tecnologias. Available at: http://abeeolica.org.br/wp-content/uploads/2018/04/Boletim-Anual-de-Geracao-2017.pdf (accessed on June 21, 2019).

Amjad M, Zafar Q, Khan F, Sheikh MM. 2015. Evaluation of weather research and forecasting model for the

assessment of wind resource over Gharo, Pakistan. International Journal of Climatology 35: 1821-1832. https://doi.org/10.1002/joc.4089

Argüeso D, Businger S. 2018. Wind power characteristics of Oahu, Hawaii. Renewable Energy 128: 324-336. https://doi.org/10.1016/j.renene.2018.05.080

Avolio E, Federico S, Miglietta MM, Lo Feudo T, Calidonna CR, Sempreviva AM. 2017. Sensitivity analysis of WRF model PBL schemes in simulating boundary-layer variables in southern Italy: An experimental campaign. Atmospheric Research 192: 58-71. https://doi.org/10.1016/j.atmosres.2017.04.003

Banks RF, Baldasano JM. 2016. Impact of WRF model PBL schemes on air quality simulations over Catalonia, Spain. Science of the Total Environment 572: 98-113. https://doi.org/10.1016/j.scitotenv.2016.07.167

Banks RF, Tiana-Alsina J, Baldasano JM, Rocadenbosch F, Papayannis A, Solomos S, Tzanis CG. 2016. Sensitivity of boundary layer variables to PBL schemes in the WRF model based on surface meteorological observations, lidar, and radiosondes during the HygrA-CD campaign. Atmospheric Research 176-177: 185-201. https://doi.org/10.1016/j.atmosres.2016.02.024

Beljaars ACM. 1994. The parametrization of surface fluxes in large-scale models under free convection. Quartely Journal of the Royal Meteorological Society 121: 255-270. https://doi.org/10.1002/qj.49712152203

Boadh R, Satyanarayana ANV, Rama Krishna TVBPS, Madala S. 2016. Sensitivity of PBL schemes of the WRF-ARW model in simulating the boundary layer flow parameters for their application to air pollution dispersion modeling over a tropical station. Atmósfera 29: 61-81. https://doi.org/10.20937/ATM.2016.29.01.05

Camargo-Schubert. 2013. Atlas eólico: Bahia. Camargo-Schubert, Bahia. Available at: http://www.infraestrutura.ba.gov.br/arquivos/File/publicacoes/atlaseolicobahia2013.pdf (accessed on September 15, 2018).

Carvalho D, Rocha A, Gómez-Gesteira M, Santos CS. 2012. A sensitivity study of the WRF model in wind simulation for an area of high wind energy. Environmental Modelling and Software 33: 23-34. https://doi.org/10.1016/j.envsoft.2012.01.019

Carvalho D, Rocha A, Gómez-Gesteira M, Santos CS. 2014. Sensitivity of the WRF model wind simulation and wind energy production estimates to planetary boundary layer parameterizations for onshore and offshore areas in the Iberian Peninsula. Applied Energy 135: 234-246. https://doi.org/10.1016/j.apenergy.2014.08.082

Chen F, Dudhia J. 2001. Coupling an Advanced Land Surface-Hydrology model with the Penn State-NCAR MM5 modeling system. Part I: Model implementation and sensitivity. Monthly Weather Review 129: 569-585. https://doi.org/10.1175/1520-0493(2001)129<0569:CAALSH>2.0.CO;2

Cheng FY, Chin SC, Liu TH. 2012. The role of boundary layer schemes in meteorological and air quality simulations of the Taiwan area. Atmospheric Environment 54: 714-727. https://doi.org/10.1016/j.atmosenv.2012.01.029

Cheng WYY, Liu Y, Zhang Y, Mahoney WP, Warner TT. 2013. The impact of model physics on numerical wind forecasts. Renewable Energy 55: 347-356. https://doi.org/10.1016/j.renene.2012.12.041

De León SP, Orfila A. 2013. Numerical study of the marine breeze around Mallorca Island. Applied Ocean Research 40: 26-34. https://doi.org/10.1016/j.apor.2012.12.003

Draper M, Guggeri A, Mendina M, Usera G, Campganolo F. 2018. A large eddy simulation-actuator line model framework to simulate a scaled wind energy facility and its application. Journal of Wind Engineering and Industrial Aerodynamics 182: 146-159. https://doi.org/10.1016/j.jweia.2018.09.010

Giannaros TM, Melas D, Ziomas I. 2017. Performance evaluation of the Weather Research and Forecasting (WRF) model for assessing wind resource in Greece. Renewable Energy 102: 190-198. https://doi.org/10.1016/j.renene.2016.10.033

Gunwani P, Mohan M. 2017. Sensitivity of WRF model estimates to various PBL parameterizations in different climatic zones over India. Atmospheric Research 194: 43-65. https://doi.org/10.1016/j.atmosres.2017.04.026

Hariprasad KBRR, Srinivas CV, Singh AB, Rao SVB, Baskaran R, Venkatraman B. 2014. Numerical simulation and intercomparison of boundary layer structure with different PBL schemes in WRF using experimental observations at a tropical site. Atmospheric Research 145-146: 27-44. https://doi.org/10.1016/j.atmosres.2014.03.023

Holtslag AAM, Boville BA. 1993. Local versus nonlocal boundary-layer diffusion in a global climate model. Journal of Climate 6: 1825-2842. https://doi.org/10.1175/1520-0442(1993)006<1825:LVNBLD>2.0.CO;2

Hong SY, Dudhia J, Chen SH. 2004. A revised approach to ice microphysical processes for the bulk parameterization of clouds and precipitation. Monthly Weather Review 132: 103-120. https://doi.org/10.1175/1520-0493(2004)132<0103:ARATIM>2.0.CO;2

Hong SY, Noh Y, Dudhia JA. 2006. A new vertical diffusion package with an explicit treatment of entrainment process. Monthly Weather Review 134: 2318-2341. https://doi.org/10.1175/MWR3199.1

Imran HM, Kala J, Ng AWM, Muthukumaran S. 2017. An evaluation of the performance of a WRF multi-physics ensemble for heatwave events over the city of Melbourne in the southeast Australia. Climate Dynamics 50: 2553-2586. https://doi.org/10.1007/s00382-017-3758-y

Islam T, Srivastava PK, Rico-Ramírez MA, Dai Q, Gupta M, Singh SK. 2015. Tracking a tropical cyclone through WRF–ARW simulation and sensitivity of model physics. Natural Hazards 76: 1473-1495. https://doi.org/10.1007/s11069-014-1494-8

Jain S, Panda J, Rath SS, Devara PCS. 2017. Evaluating land surface models in WRF simulations over DMIC region. Indian Journal of Science and Technology 10: 1-24. https://doi.org/10.17485/ijst/2017/v10i18/103522

Janjic ZI. 1994. The step-mountain eta coordinate model: further developments of the convection, viscous sublayer, and turbulence closure schemes. Monthly Weather Review 122: 927-945. https://doi.org/10.1175/1520-0493(1994)122<0927:TSMECM>2.0.CO;2

Jiménez PA, Dudhia J. 2013. On the ability of the WRF model to reproduce the surface wind direction over complex terrain. Journal of Applied Meteorology and Climatology 52: 1610-1617. https://doi.org/10.1175/JAMC-D-12-0266.1

Kitagawa YKL, Nascimento EGS, Souza NBP, Zucatelli PJ, Aylas GYR, Moreira DM, Salvador N. 2017. Assessment of the sensitivity of the WRF model using different PBL schemes over the Metropolitan Region of Salvador. XXXVIII Ibero-Latin American Congress on Computational Methods in Engineering (CILAMCE 2017). https://doi.org/10.20906/CPS/CILAMCE2017-0647

Kumar RA, Dudhia J, Bhowmik SKR. 2010. Evaluation of physics options of the Weather Research and Forecasting (WRF) model to simulate high impact heavy rainfall events over Indian monsoon region. Geofizika 27:101-125.

Lee CB, Kim J, belorid m, zhao p. 2016. performance evaluation of Four Different Land Surface Models in WRF. Asian Journal of Atmospheric Environment 10: 42-50. https://doi.org/10.5572/ajae.2016.10.1.042

Liu L, Ma Y, Menenti M, Zhang X, Ma W. 2019. Evaluation of WRF modeling in relation to different land surface schemes and initial and boundary conditions: A snow event simulation over the Tibetan plateau. Journal of Geophysical Research: Atmospheres 124: 209-226. https://doi.org/10.1029/2018JD029208

Marjanovic N, Wharton S, Chow FK. 2014. Investigation of model parameters for high-resolution wind energy forecasting: Case studies over simple and complex terrain. Journal of Wind Engineering and Industrial Aerodynamics 134: 10-24. https://doi.org/10.1016/j.jweia.2014.08.007

Mattar C, Borvoran D. 2016. Offshore wind power simulation by using WRF in the central coast of Chile. Renewable Energy 94: 22-31. https://doi.org/10.1016/j.renene.2016.03.005

Mohan M, Bahati S. 2011. Analysis of WRF model performance over subtropical region of Delhi, India. Advances in Meteorology 2011: 1-13. https://doi.org/10.1155/2011/621235

Monin AS, Obukhov AM. 1954. Basic laws of turbulent mixing in the atmosphere. Trudy Instituta Teoreticheskoi Geofiziki, Akademiya Nauk SSSR 24: 163-187 (in Russian).

NCAR/UCAR. 2015. NCEP GDAS/FNL 0.25 degree global tropospheric analyses and forecast grids. Research Data Archive, National Center for Atmospheric Research/ University Corporation for Atmospheric Research. Available at: https://doi.org/10.5065/D65Q4T4Z (accessed on January 23, 2017).

Niu GY, Yang ZL, Mitchell KE, Chen F, Ek MB, Barlage M, Kumar A, Manning K, Niyogi D, Rosero E, Tewari M, Xia Y. 2011. The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. Journal of Geophysical Research 116: D12109. https://doi.org/10.1029/2010JD015139

Obukhov AM. 1971. Turbulence in an atmosphere with a non-uniform temperature. Boundary-Layer Meteorology 2: 7-29. https://doi.org/10.1007/BF00718085

Paulson CA. 1970. The mathematical representation of wind speed and temperature profiles in the unstable atmospheric surface layer. Journal

of Applied Meteorology 9: 857-861. https://doi.org/10.1175/1520-0450(1970)009<0857:TMROWS>2.0.CO;2

Pei L, Moore N, Zhong S, Luo L, Hyndman DW, Heilman WE, Gao Z. 2014. WRF Model sensitivity to land surface model and cumulus parameterization under short-term climate extremes over the southern Great Plains of the United States. Journal of Climate 27: 7703-7724. https://doi.org/10.1175/JCLI-D-14-00015.1

Penchah MM, Malakooti H, Satkin M. 2017. Evaluation of planetary boundary layer simulations for wind resource study in east of Iran. Renewable Energy 111: 1-10. https://doi.org/10.1016/j.renene.2017.03.040

Pleim JE. 2007. A combined local and nonlocal closure model for the atmospheric boundary layer. Part II: Application and evaluation in a mesoscale meteorological model. Journal of Applied Meteorology and Climatology 46: 1396-1409. https://doi.org/10.1175/JAM2534.1

Ramos DNS, Lyra RFF, Júnior RSS. 2013. Previsão do vento utilizando o modelo atmosférico WRF para o estado de Alagoas. Revista Brasileira de Meteorologia 28: 163-172. https://doi.org/10.1590/S0102-77862013000200005

Salamanca F, Zhang Y, Berlage M, Chen F, Mahalov A, Miao S. 2018. Evaluation of the WRF-Urban Modeling System coupled to Noah and Noah-MP land surface models over a semiarid urban environment. Journal of Geophysical Research: Atmospheres 123: 2387-2408. https://doi.org/10.1002/2018JD028377

Salvador N, Loriato AG, Santiago A, Albuquerque TTA, Reis NC, Santos JM, Landulfo E, Moreira G, Lopes F, Held G, Moreira DM. 2016a. Estudo da camada limite interna térmica em condições de brisa do mar, utilizando diferentes parametrizações: Aplicação do modelo WRF na região da Grande Vitória. Revista Brasileira de Meteorologia 31: 593-609. https://doi.org/10.1590/0102-7786312314b20150093

Salvador N, Reis NC, Santos JM, Albuquerque TTA, Loriato AG, Delbarre H, Augustin P, Sokolov A, Moreira DM. 2016b. Evaluation of weather research and forecasting model parameterizations under sea-breeze conditions in a North Sea coastal environment. Journal of Meteorological Research 30: 998-1018. https://doi.org/10.1007/s13351-016-6019-9

Shin HH, Hong SY. 2011. Intercomparison of planetary boundary-layer parameterizations in the WRF model for a single day from CASES-99. Boundary-Layer Meteorology 139: 261-281. https://doi.org/10.1007/s10546-010-9583-z

Skamarock WC, Klemp JB, Dudhia J, Gill DO, Barker DM, Duda MG, Huang XY, Wang W, Powers JG. (2008). A description of the Advanced Research WRF version 3 (No. NCAR/TN-475+STR). University Corporation for Atmospheric Research. https://doi.org/10.5065/D68S4MVH

Smirnova TG, Brown JM, Benjamin SG. 1997. Performance of different soil model configurations in simulating ground surface temperature and surface fluxes. Monthly Weather Review 125: 1870-1884. https://doi.org/10.1175/1520-0493(1997)125<1870:PODSMC>2.0.CO;2

Smirnova TG, Brown JM, Benjamin SG, Kim D. 2000. Parameterization of cold-season processes in the MAPS land-surface scheme. Journal of Geophysical Research 105: 4077-4086. https://doi.org/10.1029/1999JD901047

Soni M, Payra S, Sinha P, Verma S. 2014. A performance evaluation of WRF model using different physical parameterization scheme during winter season over a semi-arid region, India. International Journal of Earth and Atmospheric Science 1: 104-114.

Stucki P, Dierer S, Welker C, Gomez-Navarro JJ, Raible CC, Martius O, Brönnimann S. 2016. Evaluation of downscaled wind speeds and parameterized gusts for recent and historical windstorms in Switzerland. Tellus A 68: 31820. https://doi.org/10.3402/tellusa.v68.31820

Surussavadee C. 2017a. Evaluation of tropical near-surface wind forecasts using ground observations. The 8th International Renewable Energy Congress (IREC 2017). https://doi.org/10.1109/IREC.2017.7926006

Surussavadee C. 2017b. Evaluation of WRF near-surface wind simulations in tropics employing different planetary boundary layer schemes. The 8th International Renewable Energy Congress (IREC 2017). https://doi.org/10.1109/IREC.2017.7926005

Tyagi B, Magliulo V, Finardi S, Gasbarra D, Carlucci G, Toscano P, Zaldei A, Riccio A, Calori G, D'Allura A, Gioli B. 2018. Performance analysis of planetary boundary layer parameterization schemes in WRF modeling set up over southern Italy. Atmosphere 9: 272. https://doi.org/10.3390/atmos9070272

Wharton S, Simpson M, Osuna JL, Newman JF, Biraud SC. 2015. Role of surface energy exchange for simulating wind turbine inflow: A case study in the southern Great Plains, USA. Atmosphere 6: 21-49. https://doi.org/10.3390/atmos6010021

Willmott CJ, Robeson SM, Matsuura K. 2012. A refined index of model performance. International Journal of Climatology 32: 2088-2094. https://doi.org/10.1002/joc.2419

Xie B, Fung JCH, Chan A, Lau A. 2012. Evaluation of nonlocal and local planetary boundary layer schemes in the WRF model. Journal of Geophysical Research 117: D12103. https://doi.org/10.1029/2011JD017080

Xiu A, Pleim JE. 2001. Development of a land surface model. Part I: Application in a mesoscale meteorological model. Journal of Applied Meteorology 40: 192-209. https://doi.org/10.1175/1520-0450(2001)040<0192:-DOALSM>2.0.CO;2

Zempila MM, Giannaros TM, Bais A, Melas D. 2016. Evaluation of WRF shortwave radiation parameterizations in predicting global horizontal irradiance in Greece. Renewable Energy 86: 831-840. https://doi.org/10.1016/j.renene.2015.08.057

Zhang D, Anthes RA. 1982. A high-resolution model of the planetary boundary layer-sensitivity tests and comparisons with SESAME-79 data. Journal of Applied Meteorology 21: 1594-1609. https://doi.org/10.1175/1520-0450(1982)021<1594:AHR-MOT>2.0.CO;2

Zilitinkevich SS. 1995. Non-local turbulent transport: pollution dispersion aspects of coherent structure of convective flows. Transactions on Ecology and the Environment 6: 53-60. https://doi.org/10.2495/AIR950071

Zucatelli PJ, Nascimento EGS, Aylas GYR, Souza NBP, Kitagawa YKL, Moreira DM. 2019. Short-term wind speed forecasting in Uruguay using computational intelligence. Heliyon 5: e01664. https://doi.org/10.1016/j.heliyon.2019.e01664

Atmósfera 36(2), 279-298 (2023)
https://doi.org/10.20937/ATM.53043

CONTENTS

Analyzing and forecasting lightning flashes and the related wind gusts at a wind energy power plant in a hilly region of western Greece

Konstantinos V. KOLOKYTHAS[1]*, Athanassios A. ARGIRIOU[1] and Vassiliki KOTRONI[2]

[1]*Laboratory of Atmospheric Physics, University of Patras, Patras, 265 00 Greece.*
[2]*Institute for Environmental Research and Sustainable Development, National Observatory of Athens, Penteli, 152 36 Greece.*
*Corresponding author: ckkolmet@yahoo.gr

Received: March 1, 2021; accepted: August 13, 2021

RESUMEN

Las plantas de energía eólica son vulnerables a tormentas eléctricas asociadas con la actividad de los rayos y las ráfagas de viento severas y cambios en la dirección del viento que las acompañan. Debido a una variedad de daños que tales fenómenos pueden causar, el conocimiento de la relación entre los sistemas de tormentas y el campo eólico producido es esencial para establecer una planta de energía eólica también durante la fase de construcción y operación. En la primera parte de este estudio se investiga la relación entre las fuertes ráfagas de viento y la actividad de los rayos en un parque eólico de Grecia. Los datos de viento provienen de aerogeneradores que cubren un periodo de tres años (2012-2014), mientras que los datos de rayos corresponden a la red de detección de iluminación ZEUS. Las ráfagas de viento están bien correlacionadas con los rayos. La correlación se maximiza durante el invierno cuando sistemas convectivos bien organizados afectan el área y es mínima en verano como resultado de las tormentas locales debidas a la inestabilidad térmica. En la segunda parte, el estudio se centra en el desarrollo de un modelo de red neuronal artificial con el fin de pronosticar estos dos parámetros a una hora utilizando cuatro variables, a saber, CAPE, TTI, velocidad del viento a 500 hPa y la cizalladura vertical del viento de 0-6 km. El modelo propuesto podría considerarse como una herramienta prometedora para simular la ocurrencia tanto de ráfagas de viento como de relámpagos, proporcionando una evidencia relativamente buena de la posibilidad de que ocurran tales eventos.

ABSTRACT

Wind power plants are vulnerable to abrupt weather changes caused by thunderstorms associated with lightning activity and accompanying severe wind gusts and rapid wind direction changes. Due to the damages that such phenomena may cause, the knowledge of the relationship between storm systems and the produced wind field is essential during the construction and operation phase of a plant. In the first part of this study, the relationship between severe wind gusts and lightning activity in a power plant in Greece is investigated. Wind data are measured at the wind turbines for a 3-year period (2012-2014); the corresponding lightning data come from the ZEUS lighting detection network. Wind gusts are well correlated to lightning strikes. This correlation is maximized during winter when well organized weather systems affect the area and minimized in summer as a result of local storms due to thermal instability. The second part of the study focuses on the development of an artificial neural network (ANN) model in order to forecast these two parameters in a 1-h ahead horizon based on wind speed, wind direction, and maximum observed wind gust measured at the nacelle of a wind turbine and four other variables, namely CAPE, TTI, wind speed at the 500 hPa isobaric level, and the 0-6 km vertical wind shear. The proposed model could be considered as a promising tool in simulating the occurrence both of wind gusts and lightning flashes, providing a relatively good evidence of the possibility of occurrence of such events.

Keywords: wind gusts, lightning strikes, forecast.

1. Introduction

We characterize as severe weather the meteorological conditions which potentially may provoke extremely hazardous situations in any aspect of human life. Consequently, for the energy production industry, severe weather is considered as the conditions that may cause extended disruptions to the energy distribution system and, in the worst case, significant interruptions to the energy production and transportation (Zepka et al., 2008). Such impacts may be caused by lightning discharge as well as by severe wind, phenomena closely related to thunderstorms.

Thunderstorms are weather phenomena related to cumulonimbus clouds and develop due to the atmospheric instability on a local or regional scale. They are one of the most spectacular and, simultaneously, dangerous meteorological phenomena, that may be encountered at any time and in any place of the world. Although thunderstorms have a relatively short duration—as an isolated event—they encompass a tremendous power, producing extreme events of lightning strikes, severe winds, heavy precipitation, and hail. All these are potentially dangerous to human life and property (Litta et al., 2013), which is the main reason why meteorologists pay special attention to thunderstorms, trying to understand the mechanisms of their development and to provide forecasts as accurate as possible.

The main factors which especially favor the development and evolution of deep convection are large atmospheric instability and humidity, and the presence of an appropriate lifting mechanism (e.g., Johns and Doswell, 1992; Doswell et al., 1996). In particular, the primary parameters associated with intense, severe, and well-organized local thunderstorms are large convective potential available energy (CAPE) combined with vertical wind shear. Case studies regarding the occurrence of such meteorological conditions in Europe and the USA for the period 1958-1999 are provided by Brooks (2009) and Brooks et al. (2007). As it emerges, the possibility of occurrence of severe phenomena for certain types of weather is higher in the European continent than in the USA, but the specific meteorological conditions that favor this occurrence are rarely encountered. Furthermore, synoptic meteorological conditions combined with local factors, such as complex topography, play an essential role in the initial stage of thunderstorms in Europe. Cases in which meteorological conditions are combined with local factors are examined in many studies (e.g., Schmid et al., 2000; Kaltenböck, 2000a, b, 2004, 2005; Kaltenböck et al., 2004; Dotzek et al., 2001, 2007; Hannesen et al., 1998, 2000; Mazarakis et al., 2008; Galanaki et al., 2015), where the importance of local factors (like orography, inshore areas or areas of convergence, among others) are particularly examined regarding the development and evolution of medium-scale weather phenomena.

Other studies focus on the climatological frequency of severe local wind events and coexisting meteorological conditions (i.e., Wakimoto, 1985; Johns and Hirt, 1987). Although they do not exclusively examine the observed wind gusts produced by thunderstorms, they provide significant information concerning their frequency and the complex thunderstorm environment which result in the development of severe weather conditions (Smith et al., 2013). Furthermore, the literature reveals also studies analyzing the occurrence of severe wind gusts as a result of passing fronts or convective weather with the use of synoptic weather observations as well as data from weather radars (Bartha, 1994). Nevertheless, the knowledge about the special characteristics, climatology and frequency of appearance of such convective gusty winds and their separation from the turbulent gusts, especially in cases of a mixed weather type, is limited.

In this perspective, the forecast of thunderstorms that may produce intense lightning activity accompanied by strong-to-severe wind gusts, for a time horizon of a few hours would be of great importance since it would contribute to the reduction of the negative impacts on, among others, the energy production and distribution sector. Although many efforts have already been devoted in this direction, thunderstorm forecasting remains a particularly challenging topic due to the temporal and spatial extent of such phenomena, combined with the non-linearity of the factors (dynamical and physical) affecting their evolution.

During the last decades many studies were devoted to comprehending the mechanisms that prevail during a thunderstorm event and the way they drive it. Some of these works deal with thunderstorm phenomena and nowcasting (e.g., Schultz et al., 2011; Chaudhuri and Middey, 2013; Wu et al., 2018;

Mostajabi et al., 2019), while others tackle the subject of thunderstorm and/or lightning nowcasting (e.g., Rasmussen and Blanchard, 1998; Yair et al., 2010; Kohn et al, 2011; Fierro et al., 2014; Giannaros et al., 2015; Das, 2017; Dafis et al., 2018; Wang et al., 2018). Most of these models are empirical, dynamic or combined. More recent studies using artificial neural networks (ANNs), which are applied to a wide range of applications, have contributed to the improvement of such efforts. These statistical models, which can handle non-linear problems, "learn" the relationship between inputs (independent variables) and outputs (dependent variables) by analyzing past data; they ignore data that do not explain a large part of the variance of the underlying process and concentrate instead on those that do so (Kalogirou, 2001).An important number of ANN application studies can be found in literature (e.g., Kalogirou, 1997; Zhang et al., 1998). These models differ in the network architecture, learning, activation, output functions, etc.

Feng and Kitzmiller (2004) discuss the set-up and application of an experimental severe weather nowcasting algorithm, based on a back-propagation neural network (BPNN) and compare it with a multiple linear regression model. The BPNN model uses as input weather radar and upper-air data from numerical models. The methods provided essential improvements to the operational Advanced Weather Interactive Processing System (AWIPS) algorithm developed from a much smaller sample of observational data for operational use, while the BPNN approach exhibited higher forecast scores and, overall, a better performance. Zepka et al. (2008) proposed a cloud-to-ground (CG) lightning forecast system using a back-propagation, multilayer, feed-forward neural network having as inputs lightning data and a number of meteorological (thermodynamic) parameters obtained by the ETA model, which is the operational numerical forecast models run at NCEP, known as the North American Mesoscale (NAM) model. The applied model showed very good results, whose accuracy depends on that of the mesoscale model data and the lightning detection network data used as input. Moreover, Zepka et al. (2014) introduced a lightning forecasting system using neural networks (NN) based on correlations between CG lightning flash data and meteorological variables obtained from MM5 with promising results. In the literature there is a great number of works based on artificial intelligence and data mining techniques (e.g., Sá et al. 2011; Bates et al. 2018; Schön et al., 2019; Mostajabi et al., 2019; Shrestha et al., 2019). An extensive survey of several research papers can be found in Bala et al. (2017).

In this work we examine the occurrence of extreme wind gusts around a wind power plant, located in a hilly region of western Greece, based on the presence of cumulonimbus clouds and lightning activity. Furthermore, the 1-h forecast of lightning flashes and wind gusts for a horizon of 24 h and with the use of ANN statistical models is analyzed. The examined period spans from January 1, 2012 to December 31, 2014.

The paper is organized as follows: section 2 describes the study area, the wind speed and direction data and the ZEUS network from which we got the lightning activity data; section 2 also explains how a wind gust is determined; section 3 presents the statistical analysis of the observed wind gusts as well as the detected lightning flashes in relation to the meteorological conditions; in section 4 the forecasting procedure is described in detail; section 5 presents the overall forecasting results, and the conclusions are discussed in section 6.

2. Data and methodology

The study focuses on a wind power plant at a hilly area close to the town of Nafpaktos, western Greece, at an altitude between 1000 and 1500 m. The study covers an area between 38º 15'-38º 37' N, and 21º 33'-22º 01' E, a grid box with dimensions 20 × 20 km^2 (Fig. 1).

Fig. 1. Study area (light yellow box: 10 ×10 km^2; yellow box: 20 ×20 km^2).

2.1 Wind data

The wind speed and direction are measured at the nacelle of a wind turbine, at a height of 67 magl. Three years of data are used, from January 1, 2012 to December 31, 2014, which consist of the mean wind speed and direction over the last 10 min and the maximum observed wind gust over the same period. These wind data have been aggregated to mean hourly values; the absolute maximum wind gust during the averaging period is also calculated.

The most commonly used approach to determine wind gusts is that of the World Meteorological Organization (WMO), according to which, e.g., the METeorological Aerodrome Report (METAR) and the Aviation SPECIal Weather Report (SPECI), code a wind gust as the maximum horizontal wind speed lasting for at least 3 s if only it exceeds the mean wind speed over the sampling interval by at least 10kts (≈ 5 m s^{-1}) (WMO, 2011a, 2014).

There are also other approaches used on a regional or national level. NOAA (1998) defines the wind gust as the maximum wind speed that lasts 2 s and is 10 kts (≈ 5 m s^{-1}) or higher than the mean wind speed sampled over a 2-min period. In another definition, a gust is the wind speed value lasting for at least 5 s within a 1-h sampling interval (Lombardo et al., 2009; Harris and Kahl, 2017; Letson et al., 2018). According to the WMO (2011b), in Argentina a gust is the maximum hourly average wind speed value that exceeds 30 kts (≈ 15 m s^{-1}). Apart from the difference in the way a wind gust is defined and calculated, it is worth mentioning that there are also numerous approaches concerning the distinction between convective wind gusts and gusts resulting from gradients or local factors (such as topography). De Gaetano et al. (2014) provide a long list of such methods.

In this study we considered as a recorded wind gust the maximum observed wind speed exceeding the average wind speed over the sampling interval by 10 m s^{-1} (20 kts). The reason for not adhering to the WMO definition is that our data are collected at an altitude of about 1350 m, where the differences between mean wind speed and wind gusts, for any sampling interval, frequently exceed 5.1 m s^{-1} by far. In fact, the average difference between the hourly mean and the mean maximum wind speed is 5.0 $\pm$ 2.0 m s^{-1} with a maximum of 19.3 m s^{-1}, while the respective average difference by the absolute maximum wind speed is 6.4 $\pm$ 2.6 m s^{-1} with a maximum of 33 m s^{-1}. On the contrary, the proposed limit of 10 m s^{-1} is observed in fewer cases and mainly when a weather change occurs or is going to occur.

The study focuses on the relationship between wind gusts observed at the wind power plant site and the lightning flashes detected simultaneously inside a specified area (grid box). If one or more wind gusts are recorded at the site of interest, in order to calculate the hourly wind gust values, the data are filtered out to the absolute hourly maximum wind speed and hourly mean wind speed, respectively. Simultaneously an extensive quality control is applied in order to clarify if those wind gusts are due to thunderstorms or to other reasons, e.g., wind speed fluctuations due to local factors such as morphology or even due to possible malfunction of the sensor or the data logger. The data analysis revealed nevertheless that thunderstorms with lightning activity are not always accompanied by gusty winds.

2.2 Lightning data

The required lighting data were provided from the European Network of Lightning Strike Detection ZEUS. This is a large distance network of five receivers placed around Europe with a very good coverage in the central and eastern Mediterranean (Kotroni and Lagouvardos, 2008; Lagouvardos et al., 2009) with a spatial accuracy of the order of 4-5 km. These receivers record the radio signal (sferic) emitted by mainly cloud-to-ground (CG) electric discharges in the VHF range between 7 and 15 kHz. Each receiver captures up to 70 sferics per second; every time such a signal is captured a detection algorithm processes the signal of all the receivers in order to detect the possible sferic candidate, excluding weak signals and noises. Consequently, the location of detected discharge is determined by applying a triangulation technique over the arrival time difference.

It is worth mentioning that the points indicating the occurrence of lightning discharges represent a big portion of the total electric activity, including the IC and CG flashes as well. According also to Maier and Krider (1986) and Williams et al. (1989), the IC activity prevails in the first stages of the developed thunderstorms, while the CGs occur later.

Drüe et al. (2007) state that the detected discharges are merged and if specific criteria are

satisfied then lightning flashes datasets might be created. According to these criteria observed flashes having a spatial difference up to 20 km and a temporal difference of 1 s are considered as one flash. Other studies suggest different approaches. In Cummins et al. (1998) and Diendorfer (2008) the location and time of the first recorded flash consist of an individual lightning datum, while in Piper and Kunz (2017) a day with lightning is considered as when at least five electric discharges are detected in a grid box of 10 ×10 km^2.

Because of the apparent difficulty in separating the IC and CG electric discharges, in this study we developed the hourly lightning dataset by processing and clustering the data according to Drüe et al. (2007). So, if a lightning flash fulfils the criteria, the hour of detection is flagged as a flash data independently of the total number of flashes detected during the same period.

3. Statistical analysis of wind gusts and lightning flashes

The frequency of detected lightning discharges over the eastern Mediterranean basin reaches a maximum in autumn (Yair et al., 2010; Kotroni and Lagouvardos, 2016), while most of them occur over maritime and coastal areas rather than overland when warm waters provide the appropriate conditions for storm development. The lightning activity is produced by cumulonimbus clouds due mainly to synoptic scale meteorological conditions (well-organized cyclones) or, in a lesser extent, to passing troughs.

Graeme and Klugmann (2014) demonstrated a clear preference of thunderstorms with lightning activity to occur over land during the warm season of the year. The opposite happens during the cold season, when the majority of these events befall over the sea. The average annual number of electric discharges per km^2 in Europe ranges between 0.1 and 4. Galanaki et al. (2015) also studied the cloud-to-ground activity over 10 years (2005-2014) in the eastern Mediterranean basin with similar results. Their results revealed an average annual number of electric discharges between 0.1 and 6, with highest densities over land, while the majority of lightning events over sea befell over the Ionian Sea and the west coastal areas of the Balkans.

The analysis of synoptic charts and thermodynamic diagrams revealed that the general meteorological conditions over our area of interest during 2012-2014 were characterized by a sequence of hot and cold periods accompanied by the corresponding weather phenomena, as normally expected. During winter the development and passage of low-pressure systems results in an important number of lightning flashes, especially over the sea. That type of weather led to a high number of events of severe wind gusts which sometimes lasted for more than two days, as a result of long-lasting frontal activity accompanying those systems. These severe gusts were either related directly to the occurrence of a lightning flash or to the wind profile induced by those pressure systems as the result of the air convection and pressure gradient. As shown in Figure 2, most of the observed wind gusts due or to thunderstorms not, as well as the detected lightning flashes are connected to southern wind directions. The main reason was the frequent passage of well-organized low-pressure systems with extended frontal activity, moving from west to the east and overpassing the area of interest.

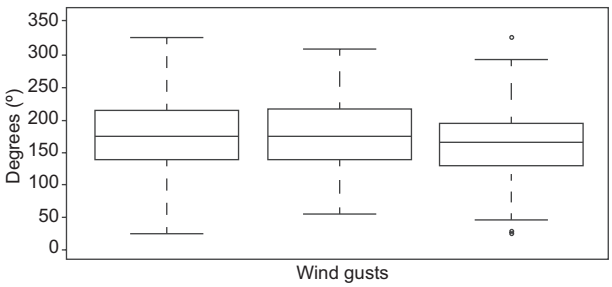

Fig. 2. Frequency distribution of wind direction at the wind turbine height related to the total wind gusts (right), wind gusts in the presence of lightnings (middle) and wind gusts in the absence of lighting (left).

The number of lightning flashes and wind gusts within a 1-h period in the site of interest are presented in Table I. The analysis of the dataset of the area of interest for the period from January 1 2012 to December 31, 2014 revealed that a total of 268 lightning strikes occurred, leading to an annual average of 89.3 strikes. The corresponding number of wind gusts, determined according to the definition given in section 2.1, was 937 with an annual average of 312.3 events. The

Table I. Total number of lightning flashes and wind gusts higher than 10 m s^{-1} in a period of 1 h. The total number of wind gusts related or not to lightning flashes are also shown.

Wind gusts	Lightning flashes	Wind gusts related to lightning flashes	Wind gusts not related to lightning flashes
937	268	636	301

number of wind gusts is divided into those related to lightning strikes and those that are not. From the total of 937 observed wind gusts, 636 (68%) were found to be related to lighting strikes. Nevertheless, only 235 were directly related to the strikes (cases in which a wind gust is observed within the same hour than a lightning strike) while the remaining 401 recorded wind gusts were the result of the general activity of severe weather affecting the surrounding area and producing strong wind gusts sometime before or after the strike. The analysis also revealed that the major portion of the detected lightning strikes (235, 88%) coincided with the observation of strong wind gusts, while the remaining 33 strikes (12%) were not accompanied by gusts or even a slight wind disturbance.

From Table I it is apparent that 301 of the observed wind gusts (32%) were not related to the occurrence of lightning strikes (direct or indirect) or to any synoptic system. This is attributed to the complex topography of the surrounding area, which combined with specific isobaric situations, might produce severe wind gusts (e.g., high pressure systems from the north combined with low pressure systems from the south, which induce strong north-northeast airstreams or strong southern air flows in the low and middle atmosphere, ahead of a slow moving weather system). Furthermore, a small percentage of such wind gusts is due to recording errors (e.g. instrument's fault). In that case, the separation of these wind gusts from those due to lightning strikes is practically inevitable.

Figure 3 presents the daily, monthly and seasonal distribution of the total wind gusts and the detected lightning strikes for the examined period. The number of wind gusts related directly or indirectly to lightning strikes and those that are not related are also shown. As shown in the figure, wind gusts and

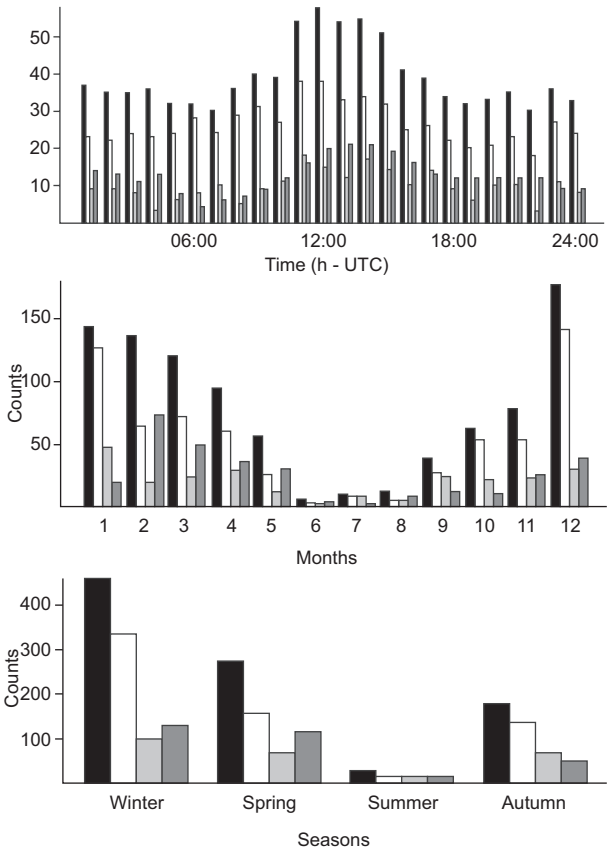

Fig. 3. Daily (top), monthly (middle) and seasonal (bottom) distribution of total wind gusts (black bars), total lightning strikes (grey bars), wind gusts related to lightning strikes (white bars) and wind gusts not related to lightning strikes (dark grey bars).

lightning strikes present a maximum in the afternoon and a second peak of smaller magnitude around midnight. It can also be observed that the activity during winter is intense but decreases significantly during the warm period of the year. This is clearly shown in the seasonal distribution, where the recorded wind gusts reach 49% of the cases and lightning strikes 41% of them. These results partially agree with the findings of related studies analyzing the frequency and distribution of detected lightning flashes in the area of interest (Graeme and Klugmann, 2014; Yair et al., 2010). The main reason for this distribution is an intensive frontal activity during the cold period of the year in the area of interest, which has its maximum between December and February. This activity sometimes leads to the development of severe weather fre-

quently accompanied by thunderstorms with intense lighting activity and a strong wind profile (Houze, 2014), especially at the altitude of the wind power plant. These conditions may last for a long period of time (more than two days) (Galanaki et al., 2018).

Figure 4 emphasizes that the distribution of wind gusts accompanied by lightning strikes is a function of weather types. The different meteorological conditions affecting the surrounding area consist of weather phenomena due to synoptic (e.g., organized cyclones, passing troughs) and atmospheric instability (mainly thermal) conditions. As shown in the figure the wind gusts due to synoptic conditions prevail. These synoptic conditions are more common in Greece during the cold period, i.e., between November and April. As seen in Table II, all wind gusts directly related to lightning strikes were exclusively caused by synoptic weather conditions for most of the months. Of the 199 lightning strikes of this category, 159 (80%) were encountered in this period (between November and April), together with 68% of the total wind gusts directly related to strikes

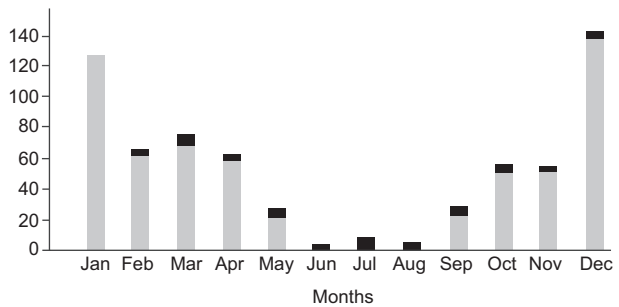

Fig. 4. Causes of wind gusts (y-axis) related to lightning flashes: synoptic scale (grey bars) and instability (black bars) conditions.

Table II also reveals that 32% of the wind gusts due to synoptic weather types were directly related to lightning strikes. As mentioned earlier, this is because the synoptic weather produces a strong wind profile that expands vertically to the upper atmosphere and also horizontally at large distances ahead of the forthcoming frontal zone (e.g., Houze, 2014). In such cases, wind speed starts increasing as frontal zones approach; eventually, strong wind gusts are recorded even if the lightning flashes in the surrounding area occur later on. Additional evidence that confirms this aspect is shown in Table II (4th column), where the rate of wind gusts per lightning strikes in each month is given, resulting in an annual average equal to 2.7. It is obvious that in most months affected by synoptic weather conditions this rate is high and far from 1, with a maximum in December (4.7), followed by February (3.4) and March (3.1).

On the other hand, when atmospheric instability develops within the study area, weather conditions are characterized by isolated and/or clusters of cumulonimbus clouds, which develop over or close to

Table II**.** Wind gusts related to flashes vs. weather types.

Month	Wind gusts related to flashes	Total flashes	Gusts/flashes	Flashes due to instability conditions	Flashes due to synoptic scale conditions
January	127	47	2.7	0	47
February	64	19	3.4	3	16
March	72	23	3.1	2	21
April	59	28	2.1	1	27
May	25	10	2.5	4	6
June	2	2	1.0	2	0
July	7	7	1.0	7	0
August	4	4	1.0	4	0
September	27	22	1.2	6	16
October	54	21	2.6	3	18
November	53	22	2.4	2	20
December	142	30	4.7	2	28
Total/average	636	235	2.7	36	199

the mountains and hills, resulting in local showers and thunderstorms sometimes followed by lightning activity, causing intense fluctuations of wind speed and direction. Wind speed increases when these phenomena appear, while the corresponding gusts are produced more or less simultaneously with the lightning flashes. These gusts may be as severe as those of the previous case but of shorter duration, occuring mainly during the presence of isolated thunderstorms and constrained temporally and spatially by the thunderstorm. When thunderstorms dissipate, such wind gusts disappear.

The corresponding percentage of gusts due to isolated thunderstorms directly related to lightning strikes was only 5%. Nevertheless (in contrast to the cases of synoptic weather conditions) in months when instability (mainly thermal) conditions prevail (May to September), the corresponding monthly rate of wind gusts per lightning strikes is too low and (especially during summer) equal to 1 (Table II). This clearly leads to the conclusion that almost all the wind gusts observed during that period of the year and in the area of interest are produced only by isolated thunderstorms mainly over the surrounding hills, and are directly related to the occurrence of lightning strikes.

4. Wind gusts and lightning flashes forecasting

4.1 Thermodynamic parameters

The development of cumulonimbus clouds and thunderstorms is a result of the simultaneous occurrence of a number of factors such as atmospheric instability, large amount of moisture, vertical wind shear and potential available energy (e.g., Johns and Doswell, 1992; Doswell et al., 1996; Rudolf et al., 2010).

In our study, we analyzed the correlation between wind gusts and flashes using four well-known parameters, namely convective available potential energy (CAPE), total totals index (TTI), wind speed at 500 hPa, and the 0 to 6 km vertical wind shear. These parameters are measured using radiosondes or satellites. They can also be obtained from numerical weather prediction models (Davis, 2001). In this study we used data provided by the Copernicus Climate Data Store, generated using Copernicus Climate Change Service (C3S) information (C3S, 2019). The data consist of hourly values with a 0.25º × 0.25º spatial resolution. In what follows we discuss each of the above parameters in detail. We also examine their correlation with flashes and gusts.

4.1.1 Convective potential available energy

The convective potential available energy (CAPE) is one of the main indexes of the possibility of thunderstorms occurrence. CAPE is obtained with Eq. (1):

$$CAPE = g\int_{LFC}^{EL}\left(\frac{\theta_{e(LFC)} - \theta_{es}}{\theta_{es}}\right)dz \qquad (1)$$

where LFC is the level of free convection, EL the equilibrium level, θ_e the equivalent potential temperature of the air parcel, and θ_{es} the saturated equivalent potential temperature of the atmospheric environment. $CAPE$ stands for positive differences between θ_e and θ_{es}, meaning that the pseudo-adiabatic of the displaced air parcel is warmer than that of the environment resulting in instability situations (Sá et al., 2011; Das, 2017). In other words, it is the amount of energy an air-parcel would have in the case it was vertically lifted to a certain height in the atmosphere.

Its value of 250 J kg^{-1} is a critical limit for the discrimination between thunderstorm and non-thunderstorm classes (Kaltenböck et al., 2009). It is also used to distinguish ordinary thunderstorms from severe events that might cause heavy rain, hail and gusty winds. Figure 5 illustrates the relation between the CAPE and the distribution of the total observed wind speed gusts related and not related to lightning and of the total number of lightning flashes. In this work, it is mainly used to investigate the occurrence of thunderstorms with lightning flashes producing severe wind gusts. In should be noted that the CAPE is mostly connected to instability events not resulting from well-organized cyclones, which are due to local causes such as unstable weather in summer months or the presence of an extended upper low accompanied by cold temperatures. So, in this work the CAPE is not considered as the main representative index for the assessment of lightning activity because (as mentioned in section 3), the lightning flashes are mainly due to frontal weather phenomena and, to a lesser extent, to isolated thunderstorms due to atmospheric instability conditions.

4.1.2 Total totals index

The total totals index (TTI) is an instability index extensively used for a first estimation of

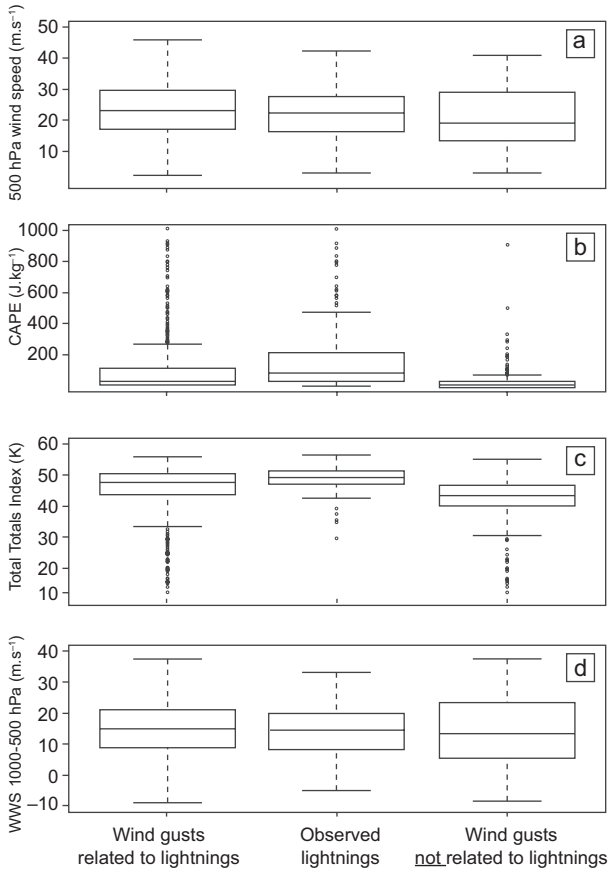

Fig. 5. Box and whisker plots of the total amount of lightning, wind speed gusts related and not related to lightnings, for (a) CAPE, (b) TTI, (c) wind speed at 500 hPa and (d) vertical wind speed between 1000 hPa and 500 hPa. The lower box boundary indicates the 25th percentile, the line within the box the mean, the upper boundary of the box indicates the 75th percentile, while bars above and below the box indicate the 90th and 10th percentiles. The 5% and 95% percentiles are marked as points.

thunderstorm events followed by severe weather phenomena, especially in North America (Peppier, 1988). This parameter gives an indication of the probability of occurrence of a thunderstorm and its severity by using the vertical gradient of temperature and humidity. The TTI consists of the arithmetic combination of vertical totals (VT; 850-500 hPa temperature difference) and cross totals (CT; the difference between the 850 hPa dew point and 500 hPa temperature) according to Eq. (2):

$$TT = T_{850} + Td_{850} - 2T_{500} \quad (2)$$

The probability of deep convection tends to increase with increasing lapse rate and atmospheric moisture content, while TTI values also vary slightly by geographic location and season. Moreover, 44 K is considered as the value-threshold for the probability of thunderstorms occurrence. Values between 44 and 50 K indicate likely thunderstorms, while values higher than 50 K are a strong indication of the development of severe mesoscale convective systems (e.g., Maddox, 1983; Velasco and Fritsch, 1987). Figure 5 presents the relation between the TTI and our parameters of interest. It can be seen that lightning flashes and wind gusts related to them occur preferentially for TTI values around 50 K. It is also shown that when the observed wind gusts are not the result of lightnings the TTI ranges between 40 and 50 K, probably revealing that these gusts are due to other reasons than lightning.

4.1.3 500 hPa wind speed

The 500 hPa isobaric level is particularly important for the evaluation of the atmosphere's status . Therefore, the wind at that level is a useful parameter for the study of the mid-tropospheric circulation. High wind speeds at 500 hPa are related to severe weather, mainly to gusty winds, because they provide an indication of the downward motion of the developed thunderstorms (Dotzek et al., 2009). In particular the rear-flank downdraft or dry rear inflow of well-organized local thunderstorms is mostly observed when the 500 hPa wind speed is high enough (Kaltenböck, 2004). As shown by the available data, wind speeds of 20 m s^{-1} or higher constitute a good index of the occurrence of lightning discharges, directly or indirectly accompanied by severe wind gusts (Fig. 5). In cases when wind gusts are not related to thunderstorm presence, the wind speeds at 500 hPa are lower.

4.1.4 0-6 km vertical wind shear

The difference of wind speed between the low (1000 hPa) and upper (500 hPa) atmosphere (vertical wind shear, 0-6 km VWS) is another useful index of imminent atmospheric instability. Higher differences are observed in cases of thunderstorms accompanied with lightning activity and strong wind gusts. When the wind gusts are not due to lightning flashes the VWS values interval is bigger with an

average approximately equal to 10 m s^{-1}. On the other hand, more than 50% (between the 25th and 75th percentiles) of the events of lightning discharges and severe wind gusts are related to VWS of about 15 m s^{-1} (Fig. 5). Therefore, this value is considered as a limit for the discrimination of wind gusts due to lightning those due to other reasons. These findings are in accordance with related works investigating events of severe weather phenomena closely related to thunderstorms and their general activity (e.g., Rasmussen and Blanchard, 1998; Schmid et al., 2000; Kaltenböck et al., 2009).

4.2 Model selection

In this work we propose a multilayer perceptron feed-forward neural network (MLP) model back-propagation learning algorithm in order to predict probable lightning flashes and wind gusts 1 h ahead for a forecasting horizon of 24 h. This type of neural networks are widely applied in various disciplines mainly because they are capable of arbitrary mapping the input-output relationship (Zhang et al., 1995; Wei et al., 2000). Typically, such a network consists of multiple layers of nodes; the first and last layers are the input and the output layers, respectively. Between them there are one or more so-called "hidden" layers of neurons fully interconnected to each other with the use of proper weights. An example of a feed-forward neural network model is presented in Figure 6.

One of the most crucial aspects during an ANN model setup is the selection of the appropriate input values (Alexiadis et al., 1998; Zhang et al., 1998). The model topology, i.e., the number of hidden layers, weights and biases values, the training method, the least acceptable error and the number of iterations during training are the key factors that should be defined in advance, while their optimal combination is heuristic, using a trial-and-error approach. For our purpose we selected feed-forward neural network with a single hidden layer which allows skip-layer connections. As activation function *fh* of the hidden layer we selected the sigmoid function:

$$fh(x) = \frac{1}{1 + \exp(-x)} \tag{3}$$

and the linear function as the output layer function, *fo*. The model output y_k is given by Eq. (4):

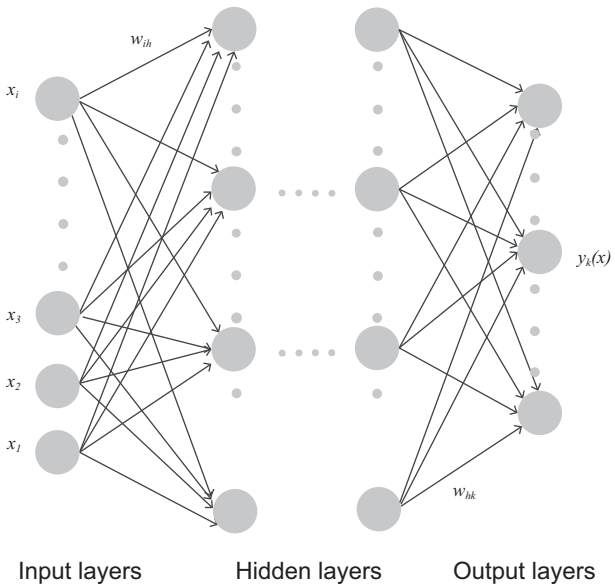

Fig. 6. Example of a typical feed forward neural network.

$$yk = fo\left(ak + \sum_h whk f_h\left(ah + \sum_i wihxi\right)\right) \tag{4}$$

where x_i is the inputs, α_k and α_h are the biases, and w_{hk} and w_{ih} are the set of weights of every link in the network. The sets of weights are adjusted by a general quasi-Newton optimization procedure, the BFGS algorithm (BFGS method), published simultaneously by Broyden (1970), Fletcher (1970), Goldfarb (1970), and Shanno (1970). This is an iterative algorithm for solving unconstrained non-linear optimization problems. Its advantage is that it gets close to a local minimum reaching it to the machine accuracy after a few iterations (Ripley, 1996). The algorithm uses function values and gradients to build up a picture of the surface to be optimized (Fletcher, 2000) and has been already applied in a variety of works resulting to better model fitting during training that led to more precise forecasts (Liu et al., 2013, 2018).

The inputs of our model are first lags of hourly data of CAPE, TTI, wind speed at 500 hPa and the 0-6 km VWS. Additional inputs tested are the corresponding lags of the difference between mean maximum and mean hourly wind speed as well as the first lag of the absolute maximum wind speed in the same hourly interval. The output layer of our model consists of one neuron providing the 1-h ahead forecast results for the next 24 h, whether or not a

wind gust or a flash have been recorded. The output of both neurons is binary, i.e., equal to 0 if no event is forecasted or 1 if an event is forecasted. In order to produce the forecasts, the outputs of the model are parameterized as follows: (*i*) the "no existence" of the phenomena corresponds to 0; (*ii*) the "existence" of the phenomena corresponds to 1. Data of a whole year, from January 1, 2012 until December 31, 2012, were used for determining the appropriate model parameters as well as its training. As already mentioned, the activation function is the sigmoid function (Eq. 1) and the linear function as the output function. The error, in order to achieve the optimum model performance is set equal to 10-4, while the maximum number of 1000 to iterations. Table III shows the tested model configurations.

Finally, the overall model performance is evaluated using the root mean square error,

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N} e_t^{\,2}} \tag{5}$$

and the mean absolute percentage error,

$$MAE = \frac{1}{N}\sum_{i=1}^{N} \mathrm{abs}(e_t) \tag{6}$$

where N is the number of observations, $e_t=o_t- F_t$ the predicted error, o_t the actual observation at time t, and F_t the corresponding forecast value. The Pearson correlation coefficient (R^2) is also used in order to provide an insight of the model ability to simulate the occurrence of wind gusts and flashes rather than its overall forecasting performance.

5. Results

5.1 Model training

The statistical metrics of the training phase are shown in Table IV. As it can be seen (a) the best forecasts are given by model 6 for the wind gusts and model 5 for flashes, and (b) the training errors decrease each time an additional parameter is used as input. This effect is more pronounced for the wind gust models, the RMSE and MAE, which are finally improved by 52 and 74%, respectively. The corresponding improvement for the flash models is 22 and 29%, respectively. It should also be noted that model configuration 7 using as input the two thermodynamic parameters (CAPE, TTI) as well as wind speed at 500 hPa and the 0-6 km VWS, presents the highest training errors. This could be due either to the uncertainty of these parameters (being the result of reanalysis) or that they provide a more general prediction of future weather

Table III**.** Tested model configurations.

Model configurations			
Model	Inputs	Network topology	
		Wind gusts	Flashes
1	Lag-1: absolute maximum wind speed	1-2-1	1-2-1
2	Lag-1: absolute maximum wind speed, TTI	2-4-1	2-2-1
3	Lag-1: absolute maximum wind speed, TTI, CAPE	3-5-1	3-4-1
4	Lag-1: absolute maximum wind speed, TTI, CAPE, 500 hPa wind speed	4-7-1	4-6-1
5	Lag-1: absolute maximum wind speed, TTI, CAPE, 500 hPa wind speed, 0-6 km VWS	5-10-1	5-7-1
6	Lag-1: absolute maximum wind speed, TTI, CAPE, 500 hPa wind speed, 0-6 km VWS, mean maximum wind speed-mean wind speed difference	7-14-1	7-12-1
7	Lag-1: TTI, CAPE, 500 hPa wind speed, 0-6 km VWS	4-8-1	4-6-1

VWS: vertical wind shear.

Table IV. Statistic results of the training phase of the various models.

Parameter		Applied model						
		1	2	3	4	5	6	7
Wind gusts	RMSE	0.13	0.11	0.12	0.11	0.1	0.063	0.14
	MAE	0.034	0.027	0.029	0.026	0.026	0.009	0.044
	R^2	0.35	0.46	0.45	0.53	0.55	0.84	0.24
Flashes	RMSE	0.092	0.09	0.088	0.087	0.072	0.081	0.087
	MAE	0.017	0.017	0.016	0.017	0.012	0.014	0.016
	R^2	0.06	0.11	0.13	0.15	0.43	0.28	0.16

conditions, such as the possibility of thunderstorms and their intensity.

5.2 Forecast evaluation

In order to test the model effectiveness in forecasting wind gusts and lightning activity in the area of interest, eight time periods, not belonging to the training sample, are used. These periods are expanded from 24 to 72 h and presented in Table V. These periods correspond to three typical meteorological conditions accompanied by a range of weather phenomena and affect the surrounding area during the examined three-year interval (2012-2014). Cases 1 to 3 correspond to well-organized low-pressure systems accompanied by frontal activity. Cases 4 to 6 correspond to thermal instability conditions. Cases 7 and 8 correspond to consecutive passing troughs in the upper (500 hPa) atmosphere.

The forecasting results for both wind gusts and lightning flashes for the eight selected cases are given in Table VI. It can be seen that wind gust forecasts are more accurate than those of lightning flashes in every case. According to the presented results the model performs better under thermal instability conditions (cases 4 to 6) as well as when gusts of wind are considered as the forecasting parameter. The worst performance was observed for the frontal activity cases (1 to 3).

Table V. Forecast cases.

Case study	Date	Weather conditions
1	18/1/2013	Frontal activity
2	24/1/2013	Frontal activity
3	28-29/12/2014	Frontal activity
4	13/5/2013	Thermal instability
5	5/6/2013	Thermal instability
6	8-10/7/2013	Thermal instability
7	27-28/3/2014	Trough
8	23-24/10/2014	Trough

Figures 7 to 12 show the forecasting results. As already mentioned, the model forecasts wind gusts with remarkable accuracy. Nevertheless, there is an apparent difference concerning the amplitude of the

Table VI. Forecasting results of wind gusts and flashes.

		Case study								
		1	2	3	4	5	6	7	8	Average
Wind gusts	RMSE	0.23	0.26	0.31	0.11	0.13	0.16	0.10	0.38	0.21
	MAE	0.14	0.20	0.19	0.023	0.028	0.034	0.028	0.18	0.10
	R^2	0.78	0.75	0.68	0.99	0.99	0.94	0.91	0.39	0.80
Lightning flashes	RMSE	0.50	0.53	0.34	0.20	0.21	0.20	0.24	0.36	0.32
	MAE	0.35	0.36	0.14	0.048	0.049	0.048	0.090	0.15	0.15
	R^2	0.0	0.0	0.29	0.0	0.0	0.0	0.0	0.18	0.06

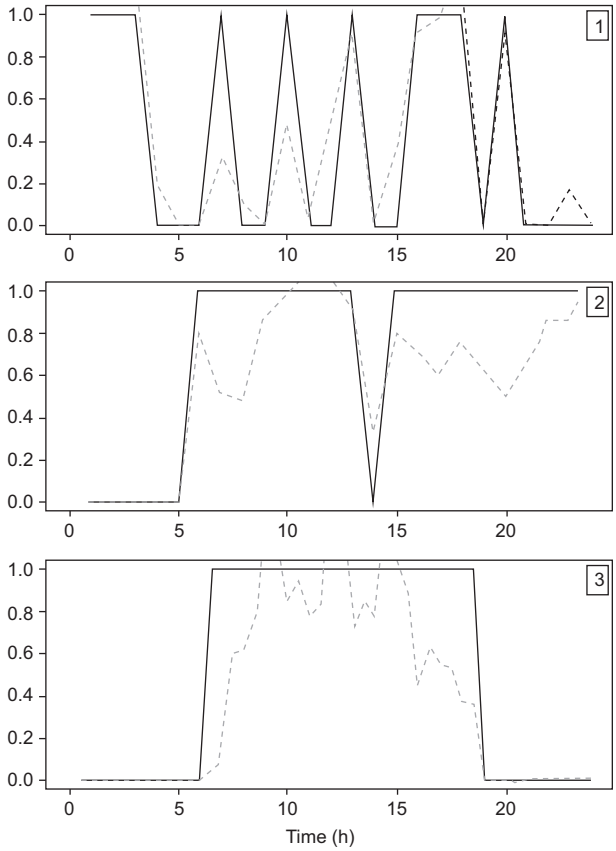

Fig. 7. Forecasted (dashed lines) vs. observed (solid lines) wind gusts (y-axis), for three cases of frontal weather conditions.

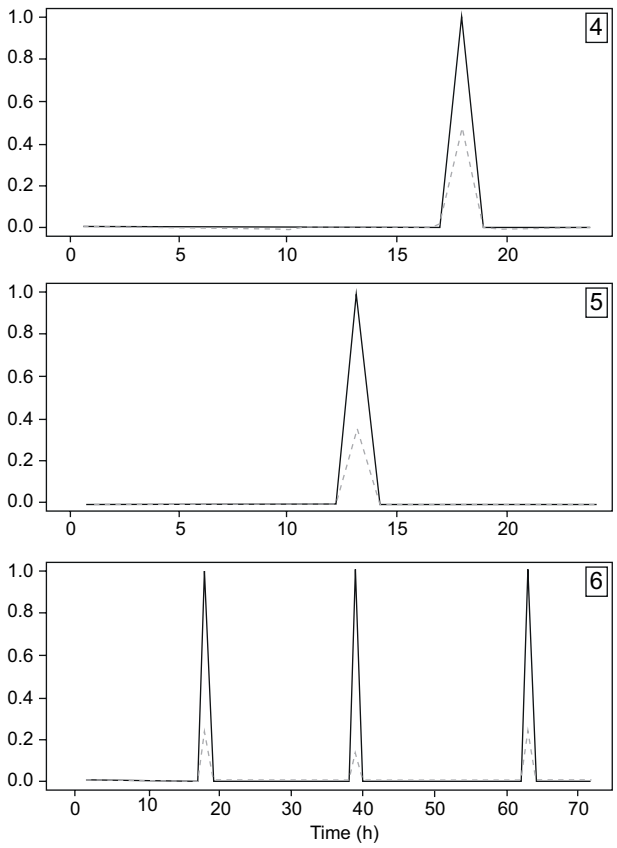

Fig. 8. As in Figure 7 but for three cases of thermal instability weather conditions.

forecasting curve revealing that the model underestimates future wind gusts. This is probably due to the first phase of forecasts, and especially to input pre-processing and the model training process. On the other hand, lightning flashes forecasts deviate significantly from the real events and are also underestimated, which leads to a worse performance concerning the output flash events for the a few next hours, bigger forecasting errors and, eventually, results of lower credibility and usefulness.

The relation between forecasts and the prevailing meteorological conditions at the area of interest needs to be highlighted. When weather phenomena ensued from atmospheric instability, the forecasts of both wind gust and lighting strike events were closer, at least temporally, to the observed data and produce smaller errors. When weather phenomena ensued resulted from well-organized low-pres-

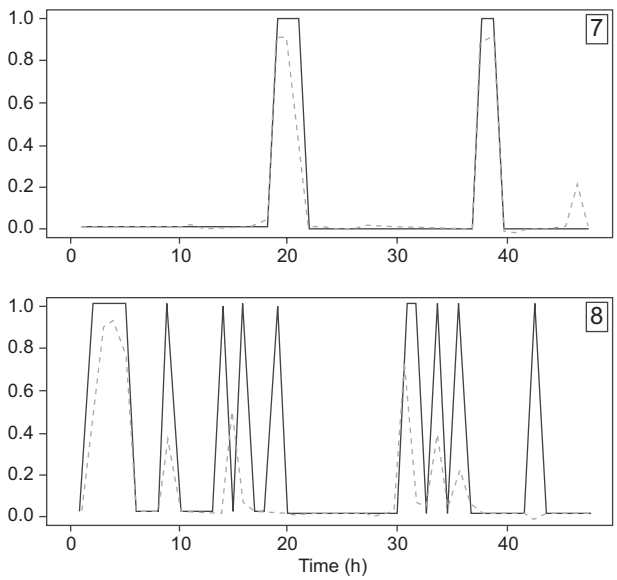

Fig. 9. As in Figure 7 but for two cases of trough weather conditions.

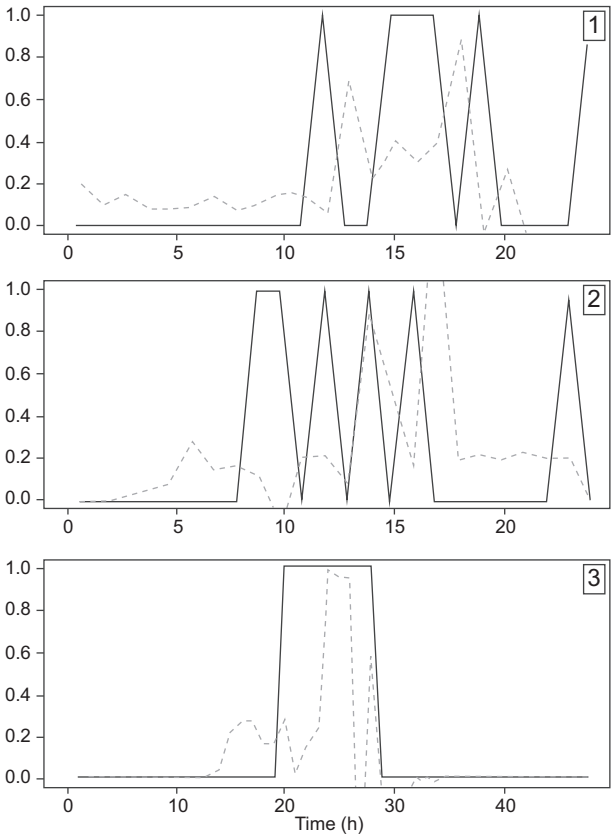

Fig. 10. Forecasted (dashed lines) vs. observed (solid lines) lightning indices (y-axis), for three cases of frontal weather conditions.

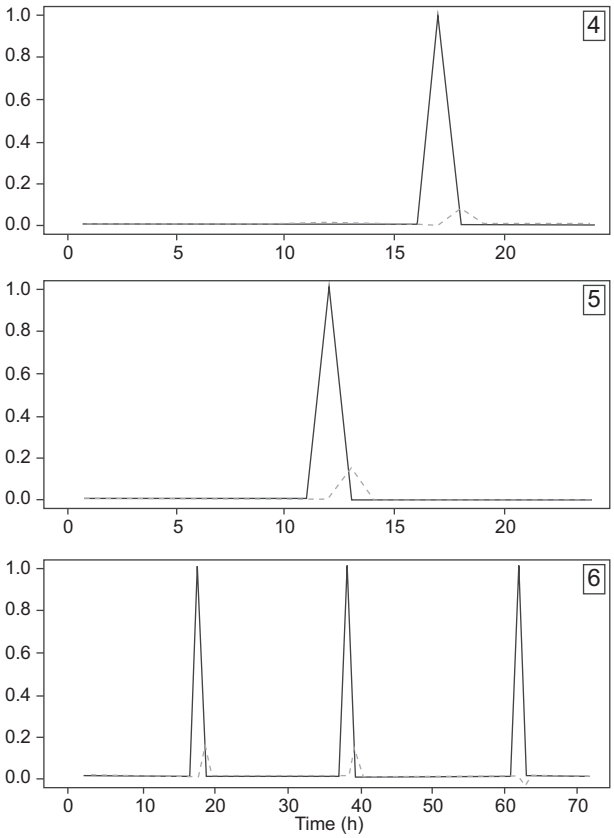

Fig. 11. As in Figure 10 but for three cases of thermal instability weather conditions.

sure systems or were due to a sequence of upper troughs, the forecasted events followed in general the observed ones, but with larger deviations. Again, wind gust forecasts were more accurate compared to those of lightning flashes. A probable cause could be that the number of gusts and flashes occurring during well-organized low-pressure systems differs from that occurring during atmospheric instabilities. When thunderstorms with their lightning activity result from atmospheric instability, the number of observed wind gusts is small and isolated, but when they are due to synoptic weather conditions their number increases significantly. Additionally, in the case of synoptic weather conditions, the phenomena last longer and present important fluctuations during their occurrence, while most of times these phenomena are followed by highly fluctuated gusty wind speeds for a longer period. It should also be

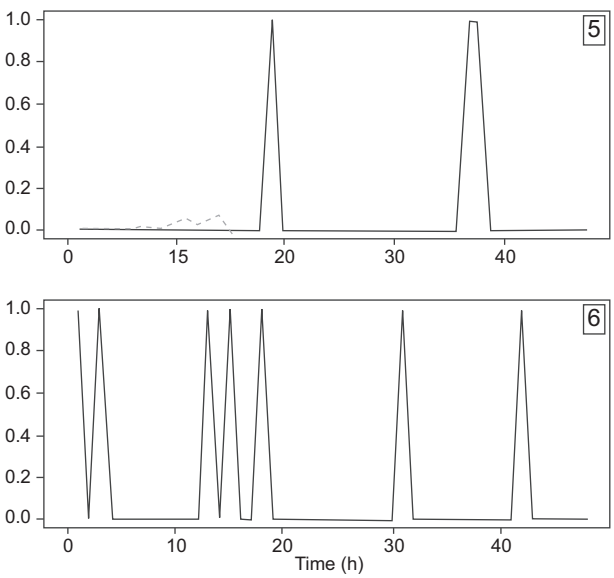

Fig. 12. As in Figure 10 but for two cases of trough weather conditions.

noted that the simulated indices and particularly the lightning flashes present significantly lower amplitude compared to those of observations.

6. Discussion and conclusions

In the first part of this study we investigated the relationship between the wind speed gusts recorded at a wind power plant in a hilly area of western Greece and the corresponding lightning strikes detected in the surrounding area for a 3-year (2012-2014) period. The observed wind gusts are strongly correlated with the detected lightning flashes ($r = 0.74$), both on a seasonal and on a daily basis. Both parameters show a maximum during winter months. On a daily basis, a maximum is observed in the afternoon and a secondary peak during midnight. More than 65% of the overall recorded wind gusts were found to be related directly or indirectly to the presence of cumulonimbus clouds accompanied by lightning activity. On the other hand, the highest percentage of the lightning strikes (almost 90%) detected in the study area was followed by wind gusts and only a small percentage was not accompanied by any change in wind speed or direction.

A major part of the gusty wind patterns affecting the area of interest was found to be due, directly or indirectly, to the development and passage of thunder clouds accompanied by lightning activity. Nevertheless, a percentage of the total amount of the recorded gusts (30-35%) was due to other factors such as strong air streams resulting from the morphology of the area, turbulence in cases of strong upper winds when an atmospheric disturbance approaches without the development of significant weather phenomena, or even the passage over the area of a trough's edge without causing any weather changes. However, it may be concluded that the presence of certain meteorological conditions that are able to produce severe weather like thunderstorms with lightning activity, causes strong to severe gusty winds. Consequently, those conditions might disrupt the operation of the wind power plant since they may delay or cancel maintenance works, damage the wind turbines or other equipment, or even lead to a complete shutdown in order to protect the wind turbines.

In the second part, we focused on the relation between wind speed gusts and lightning flashes with CAPE, TTI, wind speed at the 500 hPa isobaric level and the 0-6 km vertical wind shear. We also proposed a back-propagation feed forward ANN in order to produce a 1-h ahead forecast of these two variables. The proposed model manages to simulate the variability of the occurrence both of wind gusts and lightning flashes, especially when the four selected parameters were combined with wind data coming directly from wind turbines, such as absolute maximum and mean wind speeds. The forecasted wind speed gust values were more accurate and presented smaller errors, compared to those of flashes, which were underestimated. Nevertheless, despite the apparent underestimations the model manages to capture the fluctuation of wind gusts and, in a smaller extent, of flashes occurrence. Based on these conclusions, we state that our model might provide good 1-h ahead forecasts of the wind speed gusts for a forecasting horizon of 1 to 72 h beforehand. Also, in the case of lightning flashes the proposed method may lead to a good evidence of the possibility of occurrence of such events, although such predictions still need great advances in scientific knowledge.

Lightning flashes are characterized by important spatial and temporal variability. Therefore, the performance of the proposed model is expected to be further improved if data of higher spatial and temporal resolution are available. In addition, we suggest combining the output of our model with those of high-resolution numerical models covering a longer time period and a larger area, in order to improve accuracy. Also, understanding the significance of local and medium scale factors during the development and evolution of severe phenomena is a key parameter. In this context, data from other sites might also give additional insight of the way in which weather phenomena produce severe events. Such an analysis needs to be precisely designed in advance and include both coastal and mountainous areas as well as their combination, according to their special wind and temperature profile (Rudolf et al., 2010). This type of analysis would be very useful, especially for areas like Greece, characterized by a complex morphology and abrupt terrain variations and changes between land and sea, where many sites could be potentially used for the construction of wind power plants.

Acknowledgments
The authors would like to thank the company Terna Energy for providing the wind data and the National Observatory of Athens operating the ZEUS VLF lightning detection network for providing the lightning data used in this work.

References

Alexiadis MC, Dokopoulos PS, Sahsamanoglou HS, Manousaridis IM. 1998. Short-term forecasting of wind speed and related. Solar Energy 63: 61-68. https://doi.org/10.1016/S0038-092X(98)00032-2

Bala K, Choubey DK, Paul S. 2017. Soft computing and data mining techniques for thunderstorms and lightning prediction: A survey. In: Proceedings of the 2017 International Conference of Electronics, Communication and Aerospace Technology (ICECA). Coimbatore, India. https://doi.org/10.1109/ICECA.2017.8203729

Bartha I. 1994. Development of a decision procedure for forecasting maximum wind gusts associated with thunderstorms. Meteorological Applications 1: 103-107. https://doi.org/10.1002/met.5060010202

Bates BC, Dowdy AJ, Chandler RE. 2018. Lightning prediction for Australia using multivariate analyses of large-scale atmospheric variables. Journal of Applied Meteorology and Climatology 57: 525-534. https://doi.org/10.1175/JAMC-D-17-0214.1

Brooks HE, Anderson AR, Riemann K, Ebbers I, Flachs H. 2007. Climatological aspects of convective parameters from the NCAR/NCEP reanalysis. Atmospheric Research 83: 294-305. https://doi.org/10.1016/j.atmosres.2005.08.005

Brooks HE. 2009. Proximity sounding for severe convection for Europe and the United States from reanalysis data. Atmospheric Research 93: 546-553. https://doi.org/10.1016/j.atmosres.2008.10.005

Broyden CG. 1970. The convergence of a class of double-rank minimization algorithms. Journal of the Institute of Mathematics and Its Applications 6: 76-90. https://doi.org/10.1093/imamat/6.1.76

C3S. 2019. Copernicus Climate Data Store. Copernicus Climate Change Service. Available at: https://cds.climate.copernicus.eu/#!/search?text=upper%20air (accessed on September 25, 2019).

Chaudhuri S, Middey A. 2013. Nowcasting lightning flash rate and peak wind gusts associated with severe thunderstorms using remotely sensed TRMM-LIS data. International Journal of Remote Sensing 34: 1576-1590. https://doi.org/10.1080/01431161.2012.723834

Cummins KL, Murphy MJ, Bardo EA, Hiscox WL, Pyle RB, Pifer AE. 1998. A combined TOA/MDF technology upgrade of the US National Lightning Detection Network. Journal of Geophysical Research 103: 9035-9044. https://doi.org/10.1029/98JD00153

Dafis S, Fierro A, Giannaros Th, Lagouvardos K, Kotroni V, Mansell E. 2018. Exploring the skill of explicit lightning forecasting. Journal of Geophysical Research-Atmospheres 123: 5130-5148. https://doi.org/10.1029/2017JD027930

Das S. 2017. Severe thunderstorm observation and modeling – A review. Vayu Mandal 43: 1-19.

Davis RS. 2001. Flash flood forecast and detection methods, in severe convective storms. Meteorological Monographs of the American Meteorological Society 50: 481-526. https://doi.org/10.1175/0065-9401-28.50.481

De Gaetano P, Repetto MP, Repetto T, Solari G. 2014. Separation and classification of extreme wind events from anemometric records. Journal of Wind Engineering and Industrial Aerodynamics 126: 132-143. https://doi.org/10.1016/j.jweia.2014.01.006

Diendorfer G. 2008. Some comments on the achievable accuracy of local ground flash density values. In: Proceedings of the 29th International Conference on Lightning Protection. Uppsala, Sweden.

Doswell III CA, Brooks HE, Maddox RA. 1996. Flash flood forecasting: An ingredients-based methodology. Weather and Forecasting 11: 560-580. https://doi.org/10.1175/1520-0434(1996)011<0560:FFFAIB>2.0.CO;2

Dotzek N, Höller H, Théry C, Fehr T. 2001. Lightning evolution related to radar-derived microphysics in the 21 July 1998 EULINOX supercell storm. Atmospheric Research 56: 335-354. https://doi.org/10.1016/S0169-8095(00)00085-5

Dotzek N, Lang P, Hagen M, Fehr T, Hellmiss W. 2007. Doppler radar observation, CG lightning activity, and aerial survey of a multiple downburst in southern Germany on 23 March 2001. Atmospheric Research 83: 519-533. https://doi.org/10.1016/j.atmosres.2005.08.016

Dotzek N, Groenemeijer P, Feuerstein B, Holzer AM. 2009. Overview of ESSL's severe convective storms research using the European Severe Weather Database

ESWD. Atmospheric Research 93: 575-586. https://doi.org/10.1016/j.atmosres.2008.10.020

Drüe C, Hauf T, Finke U, Keyn S, Kreyer O. 2007. Comparison of a SAFIR lightning detection network in northern Germany to the operational BLIDS network. Journal of Geophysical Research 112: D18114. https://doi.org/10.1029/2006JD007680.

Fletcher R. 1970. A new approach to variable metric algorithms. Computer Journal 13: 317-322. https://doi.org/10.1093/comjnl/13.3.317

Fletcher R. 2000. Practical methods of optimization. 2nd edition. John Wiley & Sons, New York. https://doi.org/10.1002/9781118723203

Feng Y, Kitzmiller DH. 2004. An experimental severe weather nowcast algorithm based on a back-propagation neural network and a comparison with an algorithm based on multiple linear regression. In: Proceedings of the 17th Conference on Probability and Statistics in the Atmospheric Sciences—84th AMS Annual Meeting. Seattle, USA.

Fierro A, MacGorman D, Mansell T, Ziegler C. 2014. Forecasting lightning activity by using an explicit charging and discharging scheme in WRF-ARW. In: Proceedings of the 23rd International Lightning Detection Conference-5th International Lightning Meteorology Conference. Tucson, Arizona, USA.

Galanaki E, Kotroni V, Lagouvardos K, Argiriou A. 2015. A ten-year analysis of cloud-to-ground lightning activity over the Eastern Mediterranean region. Atmospheric Research 166: 213-222. https://doi.org/10.1016/j.atmosres.2015.07.008

Galanaki E, Lagouvardos K, Kotroni V, Flaounas E, Argiriou A. 2018. Thunderstorm climatology in the Mediterranean using cloud-to-ground lightning observations. Atmospheric Research 207: 136-144. https://doi.org/10.1016/j.atmosres.2018.03.004

Giannaros T, Kotroni V, Lagouvardos K. 2015. Predicting lightning activity in Greece with the Weather Research and Forecasting (WRF) model. Atmospheric Research 156: 1-13. https://doi.org/10.1016/j.atmosres.2014.12.009

Goldfarb D. 1970. A family of variable metric updates derived by variational means. Mathematics of Computation 24: 23-26. https://doi.org/10.1090/S0025-5718-1970-0258249-6

Graeme A, Klugmann D. 2014. A European lightning density analysis using 6 years of ATDnet data. 23rd International Lightning Detection Conference, Tucson, Arizona, USA.

Hannesen R, Dotzek N, Gysi H, Beheng KD. 1998. Case study of a tornado in the Upper Rhine valley. Meteorologische Zeitschrift 7: 163-170. https://doi.org/10.1127/metz/7/1998/163

Hannesen R, Dotzek N, Handwerker J. 2000. Radar analysis of a tornado over hilly terrain on 23 July 1996. Physics and Chemistry of the Earth B 25: 1079-1084. https://doi.org/10.1016/S1464-1909(00)00156-8

Harris AR, Kahl JD. 2017. Gust factors: Meteorologically stratified climatology, data artifacts and utility in forecasting peak gusts. Journal of Applied Meteorology and Climatology 56: 3151-3166. https://doi.org/10.1175/JAMC-D-17-0133.1

Houze RA. 2014. Cloud dynamics. 2nd ed. Academic Press, New York, 496 pp. (International Geophysics, 104).

Johns RH, Hirt WD. 1987. Derechos: Widespread convectively induced windstorms. Weather and Forecasting 2: 32-49. https://doi.org/10.1175/1520-0434(1987)002%3C0032:DWCIW%3E2.0.CO;2

Johns RH, Doswell III CA. 1992. Severe local storms forecasting. Weather and Forecasting 7: 559-569. https://doi.org/10.1175/1520-0434(1992)007%3C0588:SLSF%3E2.0.CO;2

Kalogirou SA, Neocleous C, Schizas C. 1997. Artificial neural networks for the estimation of the performance of a parabolic trough collector steam generation system. In: Proceedings of the EANN'97 International Conference. Stockholm, Sweden.

Kalogirou S. 2001. Artificial neural networks in renewable energy systems applications: A review. Renewable and Sustainable Energy Reviews 5: 373-401. https://doi.org/10.1016/S1364-0321(01)00006-5

Kaltenböck R. 2000a. Der einfluß der vertikalenwindscherung auf die bildung von Gewittern. CD-Rom, Austro Control, Vienna, Austria.

Kaltenböck R. 2000b. Nowcasting of mesoscale phenomena in Austria: Thunderstorm initiation. Convergence lines north of the Alps, dry and wet cases. International Training Workshop on Nowcasting in Sydney. WMO World Weather Research Program, Forecasting Demonstration Project, Sydney 2000. CD-Rom. Bureau of Meteorology, Sydney, Australia.

Kaltenböck R. 2004. The outbreak of severe storms along convergence lines northeast of the Alps. Case study of the 3 August 2001 mesoscale convective system with a pronounced bow echo. Atmospheric Research 70: 55-75. https://doi.org/10.1016/j.atmosres.2003.11.003

Kaltenböck R, Kerschbaum M, Pistotnik G, Pöttschacher W, Ratheiser M, Svabik O. 2004. Orographic influences on a tornadic supercell formation over Vienna, Austria. In: Proceedings of the 3rd European Conference on Severe Storms. León, Spain.

Kaltenböck R. 2005. Nowcasting of thunderstorms using mesoscale modified low-level wind in Austria. WMO Word Weather Research Programme Symposium on Nowcasting and Very Short-Range Forecasting, CD-Rom, Meteo-France, Toulouse, France.

Kaltenböck R, Diendorfer G, Dotzek N. 2009. Evaluation of thunderstorm indices from ECMWF analyses, lightning data and severe storm reports. Atmospheric Research 93: 381-396. https://doi.org/10.1016/j.atmosres.2008.11.005

Kohn M, Galanti E, Price C, Lagouvardos K, Kotroni V. 2011. Now-Casting thunderstorms in the Mediterranean region using lightning data. Atmospheric Research 100: 489-502. https://doi.org/10.1016/j.atmosres.2010.08.010

Kotroni V, Lagouvardos K. 2008. Lightning occurrence in relation with elevation, terrain slope and vegetation cover over the Mediterranean. Journal of Geophysical Research 113: D21118. https://doi.org/10.1029/2008JD010605.

Kotroni V, Lagouvardos K. 2016. Lightning in the Mediterranean and its relation with sea-surface temperature. Environmental Research Letters 11: 034006. https://doi.org/10.1088/1748-9326/11/3/034006

Lagouvardos K, Kotroni V, Betz H-D, Schmidt K. 2009. A comparison of lightning data provided by ZEUS and LINET networks over Western Europe. Natural Hazards and Earth System Sciences 9: 1713-1717. https://doi.org/10.5194/nhess-9-1713-2009, 2009

Letson F, Pryor S, Barthelmie R, Hu W. 2018. Observed gust wind speeds in the coterminous United States and their relationship to local and regional drivers. Journal of Wind Engineering and Industrial Aerodynamics 173: 199-209. https://doi.org/10.1016/j.jweia.2017.12.008

Litta AJ, Idicula SM, Mohanty UC. 2013. Artificial neural network model in prediction of meteorological parameters during pre-monsoon thunderstorms. International Journal of Atmospheric Sciences 2013: 525383. https://doi.org/10.1155/2013/525383

Liu H, Tian H, Chen C, Li Y. 2013. An experimental investigation of two wavelet-MLP hybrid frameworks for wind speed prediction using GA and PSO optimization. International Journal of Electrical Power and Energy Systems 52: 161-173. https://doi.org/10.1016/j.ijepes.2013.03.034

Liu H, Duan Z, Han F, Li Y. 2018. Big Multi-step wind speed forecasting model based on secondary decomposition, ensemble method and error correction algorithm. Energy Conversion and Management 156: 525-541. https://doi.org/10.1016/j.enconman.2017.11.049

Lombardo FT, Main JA, Simiu E. 2009. Automated extraction and classification of thunderstorm and non-thunderstorm wind data for extreme-value analysis. Journal of Wind Energy and Industrial Aerodynamics 97: 120-131. https://doi.org/10.1016/j.jweia.2009.03.001

Maier LM, Krider EP. 1986. The charges that are deposited by cloud-to-ground lightning in Florida. Journal of Geophysical Research 91: 275-289. https://doi.org/10.1029/JD091iD12p13275.

Maddox RA. 1983. Large-scale meteorological conditions associated with mid-latitude, mesoscale convective complexes. Monthly Weather Review 111: 1475-1493. https://doi.org/10.1175/1520-0493(1983)111%3C1475:LSMCAW%3E2.0.CO;2

NOAA, 1998. ASOS user's guide. National Oceanic and Atmospheric Administration, 61 pp. Available at: http://www.nws.noaa.gov/asos/aum-toc.pdfMazarakis N, Kotroni V, Lagouvardos K, Argiriou AA. 2008. Storms and lightning activity in Greece during the warm periods of 2003-2006. Journal of Applied Meteorology and Climatology 47: 3089-3098. https://doi.org/10.1175/2008JAMC1798.1

Mostajabi A, Finney DL, Rubinstein M, Rachidi F. 2019. Nowcasting lightning occurrence from commonly available meteorological parameters using machine learning techniques. NPJ Climate and Atmospheric Science 2: 41. https://doi.org/10.1038/s41612-019-0098-0

Peppier RA. 1988. Review of static stability indices and related thermodynamic parameters. SWS Miscellaneous Publication 104. Illinois State Water Survey Division, Climate and Meteorology Section, 94 pp.

Piper D, Kunz M. 2017. Spatio-temporal variability of lightning activity in Europe and the relation to the North Atlantic Oscillation teleconnection pattern. Natural Hazards and Earth System Science Discussions 17: 1319-1336. https://doi.org/10.5194/nhess-2017-35

Rasmussen EN, Blanchard DO. 1998. A baseline climatology of sounding-derived supercell and tornado forecast parameters. Weather and Forecasting 13: 1148-1164.

https://doi.org/10.1175/1520-0434(1998)013%3C1148:ABCOSD%3E2.0.CO;2

Ripley BD. 1996. Pattern recognition and neural networks. Cambridge University Press. https://doi.org/10.1017/CBO9780511812651

Rudolf B, Becker A, Schneider U, Meyer-Christoffer A, Ziese M. 2010. The new "GPCC Full Data Reanalysis Version 5" providing high quality gridded monthly precipitation data for the global land-surface is public available since December 2010. GPCC Status Report, December 2010, pp. 1- 7.

Sá JAS, Almeida AC, Rocha BRP, Mota MAS, Souza JRS, Dentel LM. 2011. Lightning forecast using data mining techniques on hourly evolution of the convective available potential energy. In: Proceedings of the 10th Brazilian Congress on Computational Intelligence (CBIC' 2011). Fortaleza, Ceará, Brazil.

Shanno DF. 1970. Conditioning of quasi-Newton methods for function minimization. Mathematics of Computation 24: 647-656: https://doi.org/10.1090/S0025-5718-1970-0274029-X

Schön C, Dittrich J, Müller R. 2019. The error is the feature: How to forecast lightning using a model prediction error. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Anchorage, Alaska.

Schmid W, Schiesser HH, Furger M, Jenni M. 2000. The origin of severe winds in a tornadic bow-echo storm over northern Switzerland. Monthly Weather Review 128: 192-207. https://doi.org/10.1175/1520-0493(2000)128%3C0192:TOOSWI%3E2.0.CO;2

Smith BT, Castellanos TE, Winters AC, Mead CM, Dean AR, Thompson RL. 2013. Measured severe convective wind climatology and associated convective modes of thunderstorms in the contiguous United States, 2003-09. Weather and Forecasting 28: 229-236. https://doi.org/10.1175/WAF-D-12-00096.1

Shrestha Y, Zhang Y, Doviak R, Chan PW. 2019. Application of Artificial Intelligence in Lightning Detection and Nowcasting. In: Proceedings of the AMS 99th Annual Meeting. Phoenix, Arizona, USA.

Schultz CJ, Petersen WA, Carey LD. 2011. Lightning and severe weather: A comparison between total and cloud-to-ground lightning trends. Weather and Forecasting 26: 744-755. https://doi.org/10.1175/WAF-D-10-05026.1

Velasco I, Fritsch JM. 1987. Mesoscale convective complexes in the Americas. Journal of Geophysical Research 92: 9591-9613. https://doi.org/10.1029/JD092iD08p09591

Wakimoto RM. 1985. Forecasting dry microburst activity over the High Plains. Monthly Weather Review 113: 1131-1143. https://doi.org/10.1175/1520-0493(1985)113%3C1131:FDMAOT%3E2.0.CO;2

Wang Y, Yang Y, Jin S. 2018. Evaluation of lightning forecasting based on one lightning parameterization scheme and two diagnostic methods. Atmosphere 9: 99. https://doi.org/10.3390/atmos9030099

Wei F, Xu ZC, Qu Z, Milbrandt J, Zhuo M. 2000. Role of EGR1 in hippocampal synaptic enhancement induced by tetanic stimulation and amputation. Journal of Cell Biology 149: 1325-1334. https://doi.org/10.1083/jcb.149.7.1325

Williams ER, Weber ME, Orville RE. 1989. The relationship between lightning type and convective state of thunderclouds. Journal of Geophysical Research 94: 213-220. https://doi.org/10.1029/JD094iD11p13213

WMO. 2011a. –Manual on codes, vol. I.1. Part A-alphanumeric codes. Technical report. World Meteorological Organization, Geneva, Switzerland.

WMO. 2011b. WMO Publication No. 306–Manual on codes, vol. II. Regional codes and national coding practices. Technical report. World Meteorological Organization, Geneva, Switzerland.

WMO. 2014. Measurement of surface wind. In: Guide to meteorological instruments and methods of observation. World Meteorological Organization, Geneva, Switzerland.

Wu F, Cui X, Zhang DL. 2018. A lightning-based nowcast-warning approach for short-duration rainfall events: Development and testing over Beijing during the warm seasons of 2006-2007. Atmospheric Research 205: 2-17. https://doi.org/10.1016/j.atmosres.2018.02.003

Yair Y, Lynn B, Price C, Kotroni V, Lagouvardos K, Morin E, Mugnai A, Llasat M. 2010. Predicting the potential for lightning activity in Mediterranean storms based on the Weather Research and Forecasting (WRF) model dynamic and microphysical fields. Journal of Geophysical Research 115: D04205. https://doi.org/10.1029/2008JD010868

Zepka GS, Pinto Jr O, Farias WRG, Carretero MA. 2008. A forecast cloud-to-ground lightning system based on a neural network – Preliminary results. In: Proceedings of the 20th International Lighting Detection Conference. Tucson, Arizona, USA.

Zepka GS, Pinto O, Saraiva ACV. 2014. Neural network for lightning forecasting in southeastern Brazil. Ciencia e Natura 36: 538-547. https://doi.org/10.5902/2179460X13721

Zhang PG, Patuwo EB, Hu MY. 1995. Wavelet neural networks for function learning. IEEE Transactions on Signal Processing 43: 1485-1497. https://doi.org/10.1109/78.388860

Zhang PG, Patuwo E, Hu MY. 1998. Forecasting with artificial neural networks: The state of the art. International Journal of Forecasting 14: 35-62. https://doi.org/10.1016/S0169-2070(97)00044-7.

Atmósfera 36(2), 299-316 (2023)
https://doi.org/10.20937/ATM.53037

CONTENTS

Fuzzy cognitive maps to explore the repercussions of less precipitation on the water supply service of the Mexico City Metropolitan Area

Norma Elizabeth OLVERA-FUENTES[1]* and Carlos GAY-GARCÍA[2]

[1] *Posgrado en Ciencias de la Tierra, Universidad Nacional Autónoma de México, Ciudad Universitaria, 04510 Ciudad de México, México.*
[2] *Instituto de Ciencias de la Atmósfera y Cambio Climático, Universidad Nacional Autónoma de México, Ciudad Universitaria, 04510 Ciudad de México, México.*
*Corresponding author: norma.olvera@atmosfera.unam.mx

Received: February 15, 2021; accepted: July 27, 2021

RESUMEN

Actualmente la operación del sistema de servicio de abastecimiento de agua potable de la Zona Metropolitana de la Ciudad de México (ZMCM) enfrenta serios problemas, los cuales generan impactos altamente negativos en los sectores ambiental, social y de abastecimiento de este sistema. Dada la presencia del cambio climático, consideramos el posible escenario en el que se genere una disminución de la precipitación media anual en la zona. Para evaluar su impacto en cada uno de los sectores se construyó el mapa cognitivo difuso (MCD) asociado a cada uno de ellos. Este marco permitió el estudio de un sistema cuya información disponible presentaba amplios rangos de incertidumbre, además de ser altamente complejo. Con base en los resultados obtenidos, presentamos una medida de mitigación para cada sector, con el fin de proponer acciones eficientes a los tomadores de decisiones. La simulación matemática se realizó mediante programación en Python.

ABSTRACT

Currently the drinking water supply service system of the Mexico City Metropolitan Area (MCMA) faces serious problems in its operation, which generate highly negative impacts on the impacts on the environmental, social and provider sectors of this system. Given the presence of climate change, we consider the possible scenario in which a decrease in average annual precipitation occurs in the area. To evaluate its impact on each of the sectors, the fuzzy cognitive map (FCM) associated with each of them was constructed. This framework allowed the study of a system whose available information presented large ranges of uncertainty, in addition to being highly complex. Based on the results obtained, we present a mitigation measure for each sector, in order to provide efficient actions to decisions makers. The mathematical simulation was carried out programming in Python.

Keywords: Fuzzy cognitive maps, climate change, decrease in average annual precipitation, Mexico City Metropolitan Area, Python programming.

1. Introduction

The water supply service of the Mexico City Metropolitan Area (MCMA) has to provide the vital liquid to a population of more than 20 million inhabitants (INEGI, 2018), for which an unsustainable exploitation of water sources is carried out (Birkle, 1998; World Bank, 2013). According to Soto and Herrera (2009, 2019), about 70% of the water consumed in the MCMA comes from the extraction of the local aquifer (SEMARNAT, 2018). If this process continues, the available water resources would be exploited to its exhaustion, which would generate

unquantifiable negative social, environmental and economic consequences. On the other hand, since the 20th century, this fragile water supply system has faced a continuous and increasingly intense forced rise in the average global temperature of the planet derived from anthropogenic climate change.

Since human activities and the corresponding emissions for this century can develop in different ways, the Intergovernmental Panel on Climate Change (IPCC) indicates that, in the most optimistic scenario, there would be an increase of 1.5 ºC by the year 2100, and in the worst scenario, it could be of 4.8 ºC (IPCC, 2014). Although at this time the range of uncertainty associated with the final temperature is wide, all scenarios indicate the occurrence of extreme climatic events characterized by excessively high temperatures, torrential rains and severe droughts. This climatic disturbance will generate profound changes in the patterns of humidity and precipitation around the world, including the MCMA. However, there is great uncertainty in the time in which temperature increases and their corresponding effects on the Mexican Republic will occur (Sánchez, 2017). Climate estimates of what will happen on our planet over time and up to the year 2100 are made considering four scenarios or representative concentration trajectories (RCP). These are characterized by their time series of emissions and concentration of all greenhouse gases; aerosol sprays; chemically active gases; land use, and vegetation cover. The trajectories are: RCP2.6, RCP4.5, RCP6.0, and RCP8.5 (IPCC, 2013), in which the number refers to the radiative forcing that is achieved measured in Wm^{-2}. The GFDL-CM3, HADGEM2-ES, and MPI-ESM-LR models considered from the Fifth Assessment Report (AR5) of the IPCC (2013), which present maps for the update of climate change scenarios projections (Fernández-Eguiarte et al., 2015), show that for the MCMA the representative concentration trajectories RCP4.5 and RCP8.5 project lower precipitation values compared to the base climatology of the period 1902-2010 (Fernández-Eguiarte et al., 2015). For both RCPs a decrease in precipitation is expected, for which it is estimated that with the increase in the planetary global temperature, the annual average precipitation over the MCMA will decrease.

The vulnerability (Adger, 2006) of the water supply service system for the MCMA will be aggravated by the effects of climate change (Escolero et. al., 2009; Soto and Herrera, 2019). It is urgent to resort to methods for exploring, in relatively short periods, the future scenarios that may arise. We have chosen to use the fuzzy cognitive map (FCM) framework to explore the effects that climate change will generate on the water supply service system of the MCMA because FCM allows working with data that may present great uncertainty and/or lack of precision.

Below we present the literature review of the role that FCMs have played in the environmental and climatological areas. Following that, we explain our method in detail, and include the results obtained when applying these mathematical models. We present a brief discussion of the results and focus our attention on detecting the elements that appear to be most affected in each of the three sectors (environmental, social and provider). An intervention action is proposed for each case. Finally, conclusions are presented.

2. Theoretical background

2.1 Fuzzy cognitive maps in the environmental and climatological areas

FCMs have been used in various fields. This has been illustrated in the modeling of simulated domains (Dickerson and Kosko, 1994); in the analysis of systems, to detect their failures and analyze their effects (Peláez and Bowles, 1996); in the creation of ecological models based on the knowledge of people (Özesmi and Özesmi, 2004); in modeling complex systems (Stylios and Groumpos, 2004); for decision-making and environmental management (Papageorgiou and Kontogianni, 2012); for developing land cover scenarios in the Brazilian Amazon (Soler et.al., 2012); for evaluating the future of water in the Seyhan basin (Cakmak et al., 2013); in the assessment of impacts and community preparedness to cyclones (Singh and Chudasama, 2017); for modeling the problem of high energy consumption in the construction sector considering the context of climate change (Mpelogianni and Groumpos, 2018); in modeling the behavior of the immediate consequence of high intensity rainfall in Kampala, Uganda (Nair et al., 2019); in sustainable agriculture (Bahri et al., 2020); and to analyze the causes and hazardous effects of malathion (organophosphorus insecticide

and pesticide) in the environmental components (air, water, and soil) (Poomagal et al., 2021). However, its application in issues related to climate change and its effects has been very recent. Pioneering contributions arose with the work of Gay-García and Paz-Ortiz in the study of stability in terrestrial climate systems (Paz-Ortiz, 2011; Gay-García and Paz-Ortiz, 2014), in the modeling, simulation and evaluation of the terrestrial climate system based on a framework of planetary limits (Paz-Ortiz and Gay-García, 2014), as well as in the analysis of the impact that climate change will have on Mexico City's water supply sources (Paz-Ortiz and Gay-García, 2013). It is worth highlighting that this last study utilized scenarios in which there was an increase or decrease of the average annual precipitation in Mexico City. In both cases the water supply sources were seriously compromised. The specific question for this study was how can we expand the field and apply FCMs to explore the repercussions that a decrease in average annual precipitation as a consequence of climate change will have on the water supply service of the MCMA. Based on the works carried out by Soto and Herrera (2009, 2019), we propose the environmental, social and supplier sectors as the basic and fundamental subsystems of the highly complex system that makes up the water supply service of the MCMA. Since all subsystems and their segments are closely interrelated with each other, any change generated in any subsystem will affect the other subsystems, and therefore, the system itself. These interactions occur in a non-linear and dynamic way, where circular or feedback processes occur, as well as a continuous exchange of energy and information between the components of the system and between the system and its environment. All these characteristics of a study system allow it to be identified as a complex system (Lara-Rosano et al., 2018).

2.2 Theory of fuzzy cognitive maps

The focus of the FCM is to consider that any system can be represented by means of cause and effect relationships between its different variables (Ross, 2010). The FCM graphically captures the causal relationships established between the different actors, which can be facts and things, processes and values, or policies and objects (Kosko, 1986). In this network, events affect others to some degree (Kosko, 1996) and the various actions affect the whole in an indissoluble way. Due to their characteristics, FCMs allow both quantitative and qualitative data to be combined, allowing to obtain a qualitative, panoramic and comprehensive understanding of the system, with the additional advantage of requiring light computing power, allowing the acquisition of results in short periods of time. Although the construction of an FCM starts from documented sources, it also exhibits a subjective nature due to the incorporation of information from the consultation of experts and the application of questionnaires. This ostensible disadvantage is its greatest quality, since the incorporation of more information allows for improving the model's veracity (Dickerson and Kosko, 1994; Kosko, 1996). The structure of an FCM corresponds to that of a weighted diagram (Harary et al., 1965), whose nodes represent variable concepts and the arcs cause-effect relationships between the concepts (Fig. 1). The concepts are represented by fuzzy sets, so the values in this set are between zero and one (Zadeh, 1965).

For an FCM of N concepts, these are identified with the labels C_1, C_2,…, C_N. The weighted arc (e.g., e_{ij}) indicates that the direction in the causal relationship begins on C_i and ends at the concept effect C_j. The values or weights assigned to e_{ij} can be both numeric and linguistic values. Linguistic values allow us to take advantage of real world input in the FCM. A positive causal relationship from concept C_i to concept C_j is indicated by $C_i \rightarrow^{+} C_j$; that is, if the concept C_i grows, so will the concept C_j. Conversely, a negative causality denoted by $C_i \rightarrow^{-} C_j$ means that if C_i grows, the concept C_j will decrease. In general these changes do not occur in an absolute linear fashion. For this reason, FCMs only allow answering questions like what if, that is, forward processes but never in the inverse. Due to this non-linearity, the model cannot explain exactly how the variables are altered (Kosko, 1996).

The fuzzy causal relationships established between the nodes of the FCM are indicated by means of the corresponding adjacency matrix E with fuzzy weights (Kosko, 1986, 1996). Since the flexible structure of an FCM allows adding, eliminating or modifying the concepts considered as the weights of the relationships in the network, we can explore an immense variety of different scenarios. The evaluation

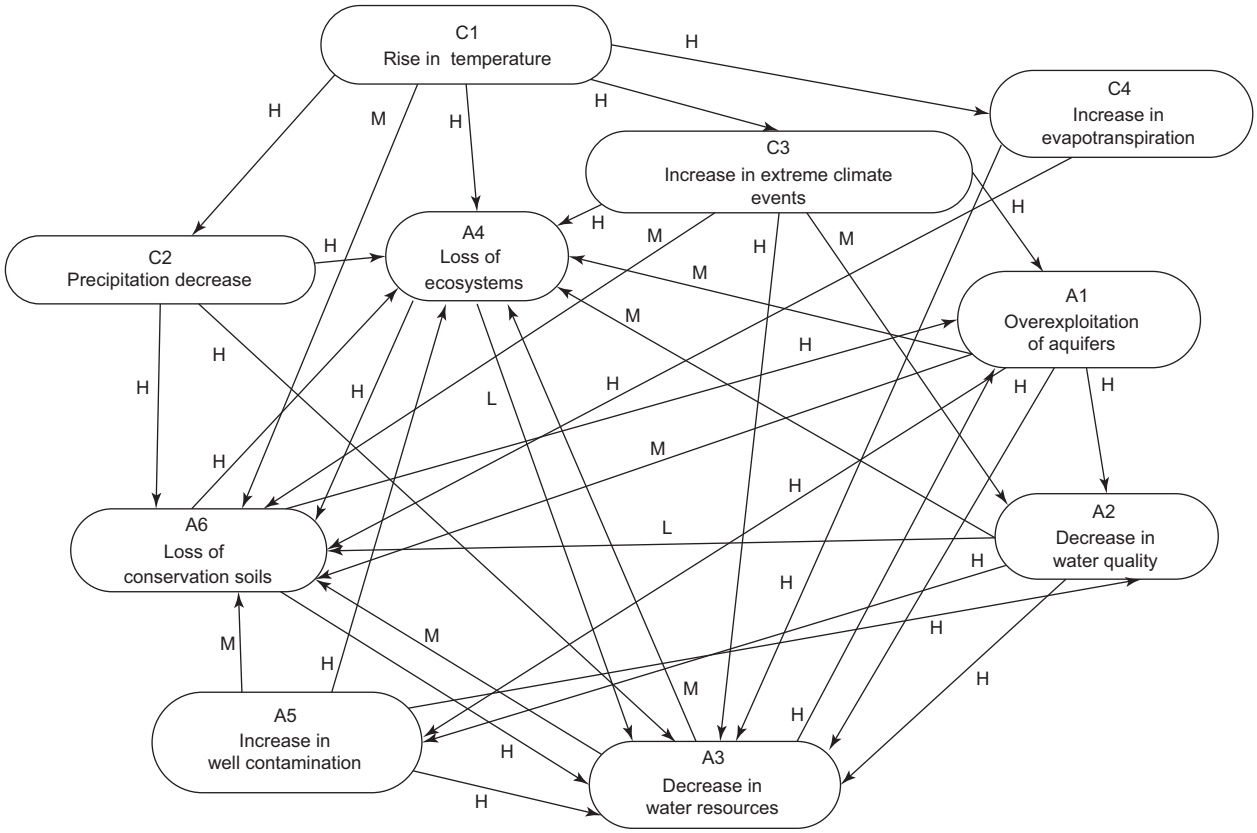

Fig. 1. Concepts and relationships between the environmental and climate subsystems. The linguistic weights of the causal relationships are denoted by H (high), M (medium) and L (low).

of an FCM is based on how well it reflects reality. To evaluate how good an FCM is, the information it provides is compared with the actual situation being modeled. If the results are contrary to what is known to be happening within the system, as it could be that without any intervention the water quality will increase, it is an indication that the FCM has been modeled in the wrong way.

2.2.1 Balance and normalization

Once the rules of causality have been established along with their respective weights, the FCM is activated by turning on one or more of its concepts. Starting from the vector that describes the initial state of the system $C(0) = [C_1(0), C_2(0),\ldots, C_N(0)]$, it is multiplied by the corresponding matrix E. The vector resulting from this operation is multiplied by the adjacency matrix E and so on. Following this iteration process from the vectors obtained after each multiplication, one of the following three possibilities will be obtained: a state of equilibrium located at a fixed point; a limit cycle or a chaotic attractor (Kosko, 1996). The procedure described here corresponds to the Kosko inference rule (Kosko, 1986) and is the one applied in this work in Eq. (1). Since all the constituent elements of our subsystems interact, we do not have the independence condition, which is why for our subsystems it is valid to apply the Kosko inference rule (Nápoles et al., 2018). To ensure that after each iteration process the values of each of the concepts remain within the interval [0,1], a normalization process is applied.

$$C_i(t_{n+1}) = S\left[\sum_{k=1}^{N} e_{ki}(t_n)C_k(t_n)\right] \quad (1)$$

Since FCMs can be seen as a type of neural network (Kosko, 1992), the logistic signal function is applied:

$$S(x) = \frac{1}{1 + e^{-x}} \quad (2)$$

It is important to note that the iterations do not have a direct representation of time in reality, but they are rather relative to the other components of the systems.

2.2.2 Concept indexes

Within the theory of digraphs, there are three measures that quantify the importance of concepts in a network (Harary et al., 1965). On the one hand, the index od(C_i) measures the flow of causality that leaves the node C_i to the network. From considering the matrix of absolute values of E, all the components of the i-th row are added.

$$od(C_i) = \sum_{k=1}^{N} \overline{e_{ik}} \quad (3)$$

The index id (C_i) measures the flow of causality that the node C_i receives from the other concepts in the network. This is obtained by adding all the components of the i-th column.

$$id(C_i) = \sum_{k=1}^{N} \overline{e_{ki}} \quad (4)$$

The centrality of a node C_i measures the importance of this concept in relation to the network; this measures is obtained by adding the values of its indices $od(C_i)$ and $id(C_i)$.

$$cen(C_i) = od(C_i) + id(C_i) \quad (5)$$

A node for which its index $od(C_i)$ is null, and its $id(C_i) > 0$ is called the "receiver concept". If instead the $id(C_i)$ is null and the $od(C_i) > 0$, the concept will be a "transmitter". If the $od(C_i) > 0$, and the $id(C_i)$, the concept will be "mixed".

3. Results

In the proposed model, the three illustrative diagrams represent the basic elements that make up each of the three subsystems —environmental, social and supplier— that encompass the water supply service of the MCMA as well as those of the climate system with which it interacts (Figs. 1, 2, 3).

Environmental concepts are included in Figure 1: exploitation of aquifers (A1); decrease in water quality (A2); decrease in water resources (A3); loss of ecosystems (A4); increase in contamination of wells (A5), and loss of conservation soils (A6).

Social concepts are included in Figure 2: industrial, population and tourism growth (S1); increase in water

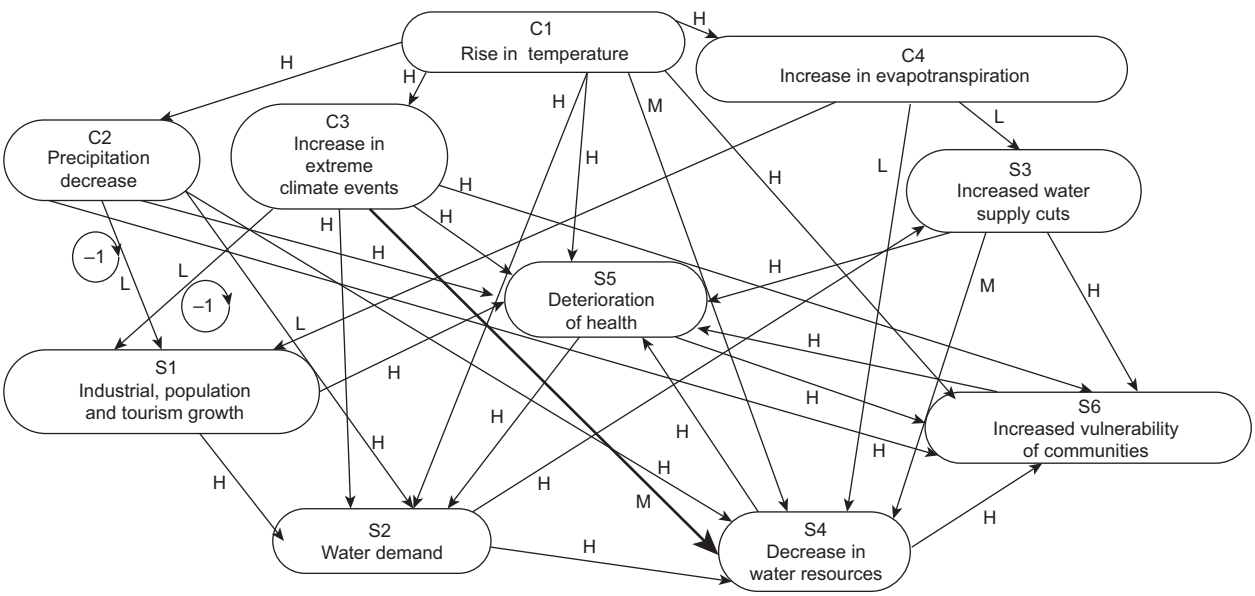

Fig. 2. Concepts and relationships between the social and climate subsystems. The linguistic weights of the causal relationships are denoted by H (high), M (medium) and L (low).

demand (S2); increased cuts in water supply (S3); decrease in water quality (S4); deterioration of health (S5), and increased vulnerability of communities (S6).

Provider and climatic concepts are included in Figure 3: increased exploitation of supply sources (P1); increased water sanitation costs (P2); increased deficiency in infrastructure for the distribution of drinking water (P3); decrease in the efficiency of the drainage system (P4); generation of social, political, economic and environmental problems (P5); increase in floods (P6); increase in temperature (C1); decrease in precipitation (C2); increase in extreme climate events, framed in a scenario characterized by an increase in the frequency and intensity of droughts (C3), and increase in evapotranspiration as an effect of the increase in temperature (C4).

3.1 Causal relationships

3.1.1 Climate subsystem

In the scenario in which an increase in temperature (C1) generates a decrease in the average annual precipitation received in a positive way by the active region, the following concepts are affected: there is a decrease in precipitation (C2); an increase in climate extreme events (C3), characterized by an increase in the frequency and intensity of droughts, and an increase in evapotranspiration (C4) (IPCC, 2007).

3.1.2 Environmental and climate subsystems

As documented by the IPCC (2007), increases in temperature (C1) have had a direct effect on the increase in loss of ecosystems (A4) and conservation soils (A6). The decrease in annual precipitation (C2) results in a decrease in the amount of water available for the recharge of dams, lakes and aquifers, thus directly generating a decrease in water resources (A3). The lack of rainfall (mainly in times of low water) generates the loss of ecosystems (A4) and conservation soils (A6). We consider that the increase in the occurrence of extreme climatic events (C3), such as stronger periods of drought, will have an impact on the quality of available water (A2), the reduction of

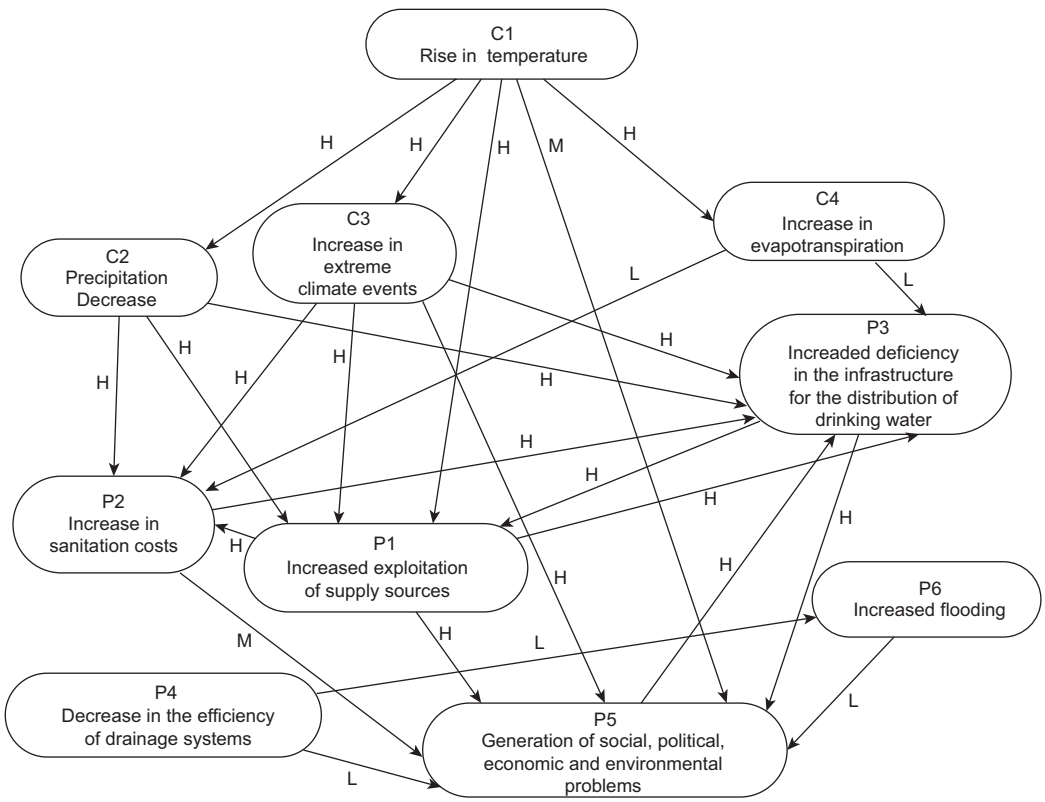

Fig. 3. Concepts and relationships between the provider and climate subsystems. The linguistic weights of the causal relationships are denoted by H (high), M (medium) and L (low).

water resources (A3), and the need for overexploitation of the aquifers (A1) to meet the population's demand. When (A1) occurs, the groundwater levels in the aquifers decrease, which causes a horizontal flow inversion in some parts of the aquifer, such as occurs in Iztapalapa and Texcoco, resulting in a salinization of the water (Soto and Herrera, 2009), decreasing its quality (A2) or polluting it (A5), thus reducing the available water resources (A3), conservation soils (A6) and ecosystems (A4). Periods of frequent or intense droughts generate an increase in the loss of ecosystems (A4) and of conservation soils (A6). Currently the voracious growth of urban areas and immoderate tree felling (A6), generates an overexploitation of aquifers (A1), a decrease in water resources (A3) and a loss of ecosystems (A4). The increase in the loss of moisture from soils due to evaporation and transpiration of vegetation (C4) reduces the amount of infiltrated water to recharge the aquifers, generating a loss of conservation soils (A6) and a decrease in the available water resources (A3). In turn, (A3) generates an overexploitation of aquifers (A1), loss of ecosystems (A4) and loss of conservation soil (A6). If water in the aquifers is deficient (A2), or if contamination of wells is increased (A5), this contributes to the reduction of available water resources (A3), the loss of ecosystems (A4) and of conservation soils. In turn, an increase in contamination of the wells decreases the quality of water (A2). Good quality water indicates that there is a biological balance, so the loss of ecosystems (A4) contributes to the reduction of organisms necessary to keep the water in good condition, thus reducing water resources (A3) with loss of conservation soils (A6) (Soto and Herrera, 2009). The FCM associated with these cause-effect relationships is shown in Figure 1.

3.1.3 Social and climate subsystems

Faced with increases in temperature (C1) or in the occurrence of extreme weather events (C3), high temperatures cause the population to increase its water consumption (S2), as occurs when personal hygiene is done more than once a day. This generates an increase in the amount of required water, which is obtained from an increase in the exploitation of aquifers, generating a decrease in the quality of water (S4). Temperatures higher than usual contribute to the detriment of health (S5) and increase the vulnerability of communities (S6). We consider that the decrease in precipitation (C2) or the increase in extreme climatic events (C3) would negatively affect concept S1, which refers to industrial, population and tourism growth, because a decrease in precipitation or the presence of extreme climatic events hinder human activity as well as being factors that deteriorate health (S5). Both increase the vulnerability of communities (S6), favor greater water consumption (S2), and impact on reducing water quality (S4). An increase in the evapotranspiration process (C4) will have among its innumerable consequences a negative effect on farmland. We suppose that an increase in the reduction of water resources may lead to this phenomenon, contributing to migration towards the MCMA in search of the vital liquid (S1). Faced with a decrease in water resources, an increase in shortages of water supply is expected (S3), as well as a decrease in its quality (S4). From the study by Soto and Herrera (2019) it is known that industrial, population and tourism growth (S1) demands a greater amount of water to cover its needs (S2). In turn, this growth contributes to the deterioration of the population's health in various ways (S5), whether in the emission of more pollutants into the atmosphere, water pollution, or contact with bacteria, among others. If we assume that the amount of available water is limited, and given the current conditions, to respond to the incessant increase in water demand (S2) it will be necessary to cut back on its supply (S3), in order to distribute it. To satisfy this demand for water, the population resorts to various means, such as renting pipes or buying jugs of water, which in many cases is of very poor quality (S4). We consider that an increase in cuts in water supply (S3) generates a decrease in its quality (S4). Both the water cuts (S3) and the decrease in its quality (S4) contribute to the deterioration of health (S5), increasing the vulnerability of the community (S6), a situation widely palpable in these times of the Covid-19 pandemic. In turn, the deterioration of health (S5) generates an increase in the demand for water (S2), and therefore an increase in the vulnerability of the community (S6). If vulnerability increases (S6), this has an effect on deteriorating health (S5). The FCM associated with these cause-effect relationships is shown in Figure 2.

3.1.4 Provider and climate subsystems

As well as the water supply service of the MCMA, increases in temperature (C1) will force the supplier to overexploit the supply sources (P1) to meet the growing needs of the population. On the other hand, it is assumed that changes in temperature will generate social, political, economic and environmental problems (P5). We consider that simply due to time issues, the presence of problems between the different sectors of society will affect the mechanisms that allow the implementation of better infrastructure for the distribution of drinking water, which is why this system will be more deficient (P3), as the many water leaks from the hydraulic system have shown. We assume that if the system is not very efficient, it will be necessary to resort to an increase in the exploitation of supply sources (P1), which in turn will contribute to generating social, political, economic and environmental problems (P5). A decrease in precipitation (C2) will generate a decrease in the amount of recharge water. Since the demand for the vital liquid never yields and is exacerbated by the increase in temperature, we assume that the exploitation of supply sources will increase (P1). As there is little water in the reservoirs, the amount of dissolved particles may be important in proportion, which would lead to an increase in the costs of water sanitation (P2). Given this condition, an increase in the deficiency of the infrastructure for the distribution of drinking water (P3) is assumed, such as the lack of running water or pipes. For the case in which extreme climate events occur in the region (C3), the same causal relationships will exist, with the additional factor that, under extreme climate conditions, the generation of social, political, economic, and environmental problems will increase (P5). In a similar way, a decrease in the amount of water due to an increase in evapotranspiration (C4) will generate increases in the costs of water sanitation (P2), as well as an increase in the deficiency of the infrastructure for the distribution of drinking water (P3). We consider that both the increase in the exploitation of supply sources (P1) and the increase in water sanitation costs (P2) generate an increase in the deficiency of the infrastructure for the distribution of drinking water (P3), as well as social, economic, and environmental problems (P5). For the water supply service to be efficient, it must have an adequate sewage collection system. A decrease in the efficiency of drainage systems (P4), for example due to damage to their infrastructure, or the fact that part of the population does not have this service leads to the generation of social problems (P5). Furthermore, as is already the case in many situations, during the rainy season, this service is exceeded, which causes severe flooding (P6). The increase in floods (P6) generates social, economic and environmental problems (P5). The FCM associated with these cause-effect relationships is shown in Figure 3.

3.2 Feedback process

The weights of the causal relationships can be worth {–1,0,1} (Kosko, 1992). In our maps, the only cases with negative values occur between the concept C2, C3 and S1. To study the feedback processes of the system, one or more concepts are turned on in the initial state vector and, without keeping them as forcing, the resulting vector after each iteration already normalized, applying Eq. (2), is multiplied by the adjacency matrix. This process is repeated until it is possible to analyze the equilibrium state that the system reaches (Kosko, 1986, 1996; Ross, 2010). For each of the three subsystems, two cases were considered, (a) the one in which extreme climate events do not occur, and (b) where they do occur.

Since in case (a) the three subsystems exhibit the same patterns as in case (b) below, we only show the results in a scenario in which there are extreme climatic events. The only difference between these two scenarios is when extreme events are triggered, being the result that there are higher values of the concepts from the first iteration. This indicates that the presence of extreme events accelerates the asymptotic growth of the values of the concepts.

In all cases, the first four components of the vector correspond to the climatic concepts and appear in the order in which they are numbered in each of the fuzzy cognitive maps. For example [C_1 C_2 C_3 C_4 A_1 A_2 A_3 A_4 A_5 A_6].

3. 2. 1 Environmental and climatic subsystems

We turn on the concept of extreme events (C3), along with (C1) and (C2). The initial state vector is v0 = [1 1 1 0 0 0 0 0 0 0], without these concepts remaining as forcing. After nine iterations, the vector N9 = [0.5 0.62245933 0.62245933 0.62245933 0.93215231

0.91824864 0.99861344 0.9984338 0.86417417 0.99915809] is obtained. The first point to highlight is that all concepts turn on even if concepts C1, C2, and C3 are no longer active, which indicates that the causal flow will continue to circulate through the cognitive map continuously. The most affected concept in this panorama is the loss of conservation soils (A6), followed by a decrease in water resources (A3), loss of ecosystems (A4), overexploitation of aquifers (A1), decrease in water quality (A2), and the increase in contamination of wells (A5) (Fig. 4).

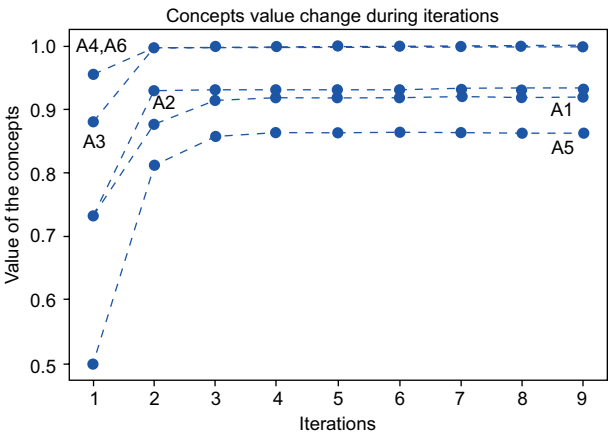

Fig. 4. Feedback process in the environmental subsystem.

3.2.2 Social and climate subsystems

The initial state vector is v0 = [1 1 1 0 0 0 0 0 0 0]. For the ninth iteration, the state vector N9 = [0.5 0.62245933 0.62245933 0.62245933 0.34922232 0.95634201 0.82903469 0.9845228 0.99258827 0.9895543] is obtained. The deterioration in the health of the population (S5), the increased vulnerability of the communities (S6), the decrease in water quality (S4) and the increase in the demand for water (S2) are the concepts most affected by this scenario. On the other hand, in a unique way, industrial, population and tourism growth (S1) starts with a value lower than 0.3. One possible interpretation is that human activities will be affected in their response capacity, but later their resilience will increase, which however will stagnate, remaining constant. It is very important to highlight that, under this scenario, the health of the population and its vulnerability will be the most affected, and that all the concepts remain on even when the concepts (C1), (C2) and (C3) are turned off (Fig. 5).

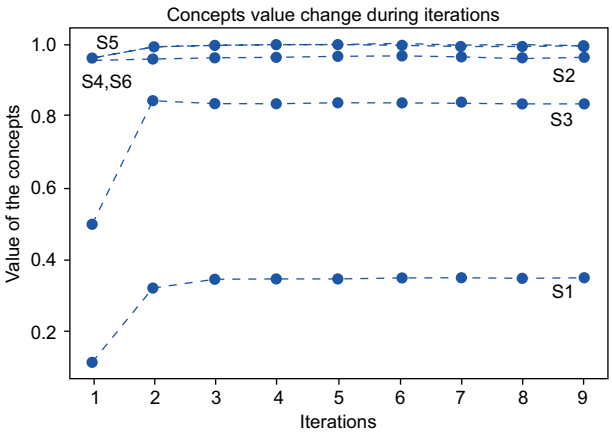

Fig. 5. Feedback process in the social subsystem.

3.2.3 Provider and climate subsystems

Given the initial state vector v0 = [1 1 1 0 0 0 0 0 0 0], at the end of nine iterations it leads to vector N9 = [0.5 0.62245933 0.62245933 0.62245933 0.93913417 0.94302672 0.99136783 0.5 0.9940503 0.62245933]. Figure 6 shows that the values of the concepts are displayed in the following decreasing order: generation of social, political, economic and environmental problems (P5), followed by an increase in the deficiency of the infrastructure for

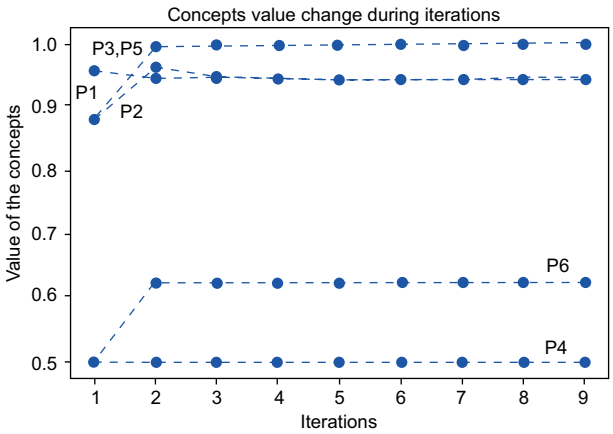

Fig. 6. Feedback process in the provider subsystem.

the distribution of drinking water (P3), an increase in water sanitation costs (P2), an increase in the exploitation of supply sources (P1), an increase in floods (P6), and a decrease in the efficiency of the drainage system (P4). The fact that concept P4 appears with the lowest relative and constant value among the concepts is explained by considering that the initial conditions consider a decrease in the annual average rainfall, so that problems derived from floods will not appear.

3.3 Hidden patterns

The weights of the causal relationships can be worth {–1,0,1} (Kosko, 1992). In our maps the only case with a negative value occurs between concepts C3 and S1. In all cases, for the analysis of the hidden patterns in each of the subsystems to be considered, the concepts that are turned on are forced to maintain that state (Kosko, 1988, 1996). Below we present the effects that climate change has on each system by maintaining its action continuously.

For each of the three subsystems, two cases were considered, (a) the one in which extreme climate events do not occur, and (b) where they do occur. Since in case (a) the three subsystems exhibit the same patterns as in case (b) below, we only show the results in a scenario in which there are extreme climatic events. However, the values obtained from the first iteration in case (b) are higher, which indicates a great sensitivity of the concepts to the presence of extreme events that remain forcing.

3.3.1 Environmental and climate subsystems

The initial state vector v0 = [1 1 1 0 0 0 0 0 0 0] leads in the ninth iteration to the state vector F9 = [1 1 1 0.73105858 0.95253982 0.94385435 0.99944518 0.9995759 0.86948287 0.99979575]. Figure 7 reveals that the concepts that tend asymptotically to 1 are: the loss of conservation soils (A6), the loss of ecosystems (A4), and the decrease in water resources (A3). To a lesser degree it is followed by the over-exploitation of aquifers (A1), the decrease in water quality (A2), and the increase in contamination of wells (A5). In this scenario, where an increase in temperature, a decrease in annual average rainfall, and the presence of extreme events persist as forcings, there is a loss of conservation soils and ecosystems and a decrease in water resources.

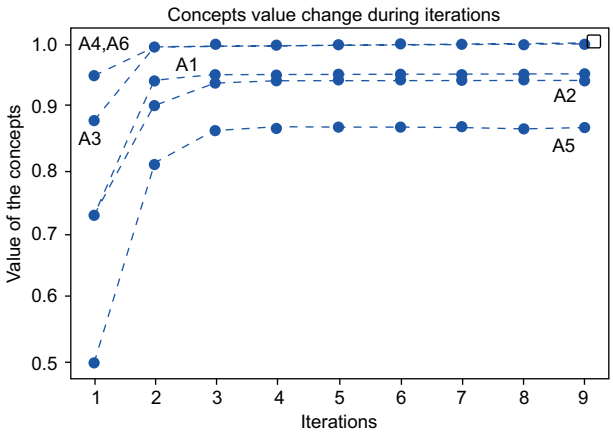

Fig. 7. Detection of hidden patterns in the environmental subsystem.

3.3.2 Social and climate subsystems

Starting from the initial state vector v0 = [1 1 1 0 0 0 0 0 0 0], after the ninth iteration, the state vector F9 = [1 1 1 0.73105858 0.21943852 0.98547297 0.84768154 0.99618204 0.99767198 0.99710403] is obtained. Figure 8 reveals that the concepts that tend asymptotically to 1 are: deterioration of health (S5); increased vulnerability of communities (S6); decrease in water quality (S4). To a lesser degree it is followed by increase in water demand (S2); increased cuts in water supply (S3); and industrial, population and tourism growth (S1). In a highly notorious way, in this scenario there is a situation in which the concept

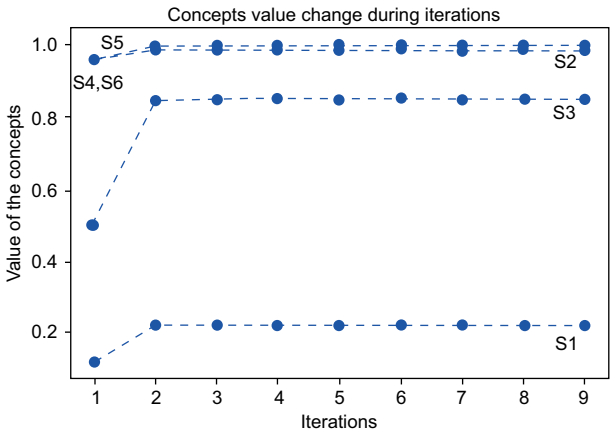

Fig. 8. Detection of hidden patterns in the social subsystem.

(S1) acquires the lowest relative and constant value with respect to the other concepts. This would indicate that the growth of the population's activities would be seriously impacted in a negative way.

3.3.3 Provider and climate subsystems

From the initial state vector v0 = [1 1 1 0 0 0 0 0 0 0], the state vector F9 = [1 1 1 0.73105858 0.98195402 0.97617747 0.99662137 0.5 0.99771069 0.62245933] is obtained in the ninth iteration. Figure 9 displays the relative values of the concepts in decreasing form: generation of social, political, economic and environmental problems (P5); increase in the deficiency of the infrastructure for the distribution of drinking water (P3); increase in the exploitation of the sources of supply (P1); increase in water sanitation costs (P2); increase in floods (P6), and decrease in the efficiency of the drainage system (P4). This last concept remains with a constant value of 0.5.

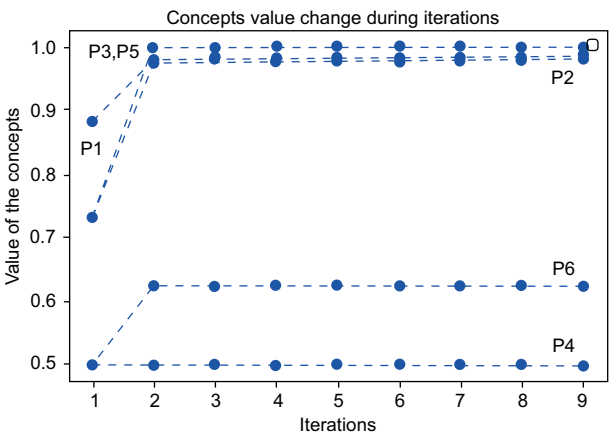

Fig. 9. Detection of hidden patterns in the provider subsystems.

We consider that the coincidence between the results obtained for causes that occur for a certain period of time and for those that remain is due to the fact that each of the three subsystems is already in a highly vulnerable state even without considering the effects of climate change. The problems already present will be exacerbated by the new climatic conditions.

Until now, all interactions have been considered to have the same intensity; however, it is clear that in real life, there will be causal relationships that are more intense than others. To describe the intensity of causal relationships, based on the literature review carried out so far, we use a linguistic description to describe "how intense is the action of one concept on another", since the intensities start from a linguistic assessment. Cognitive maps acquire a fuzzy character.

3.4 Fuzzy values

The intensity of the relationship between the different concepts is estimated through the linguistic variables high (H), medium (M) or low (L) (the causal relationships with linguistic values are observed in Figs. 1, 2, and 3). To analyze the behavior of each of the subsystems by means of the adjacency matrix, the interval [0,1] that describes the degree of belonging of a subset to a set is divided into three parts. For each subset a representative value is selected; in this case, a value of 0.3 is considered low (L), 0.6 as medium (M) and 0.9 as high (H). These values are fuzzy since they are within the interval [0,1]. On the other hand, they are very close to the values suggested by Papageorgiou et al. (2006): 0.25, 0.50, and 0.75, respectively. Due to the nature of fuzzy cognitive maps, experts can modify or assign the value of the weights based on what they consider best represents the interaction. To each subsystem a matrix with fuzzy values is constructed. For each of the three subsystems, concept C1 is turned on, keeping it as forcing, and the state vector of the system obtained in the ninth iteration is analyzed.

3.4.1 Environmental and climate subsystems

We turn on the C1 concept and keep it as forcing. From the initial state vector v0 = [1 0 0 0 0 0 0 0 0 0], the vector F9 = [1 0.7109495 0.7109495 0.7109495 0.91914398 0.88133865 0.99586773 0.99594413 0.83485498 0.99404144] is obtained for the ninth iteration. The evolution of the concepts' values as a function of the number of iterations is shown in Figure 10. As it can be seen, after the first iteration, concepts A4, A3 and A6 quickly tend to 1. This indicates that, in the continuous presence of an increase in temperature, the most affected concepts in descending order are: loss of ecosystems (A4), decrease in water resources (A3), and conservation soils (A6). These three concepts correspond to those most

affected in all the cases previously studied (Figs. 4 and 7)

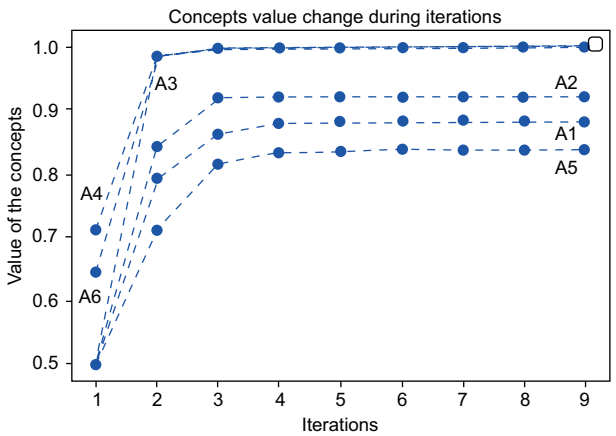

Fig. 10. Tendency to the equilibrium state in the environmental subsystem.

3.4.2 Social and climate subsystems

From the initial state vector v0 = [1 0 0 0 0 0 0 0 0 0 0], the state vector F9 = [1 0.7109495 0.7109495 0.7109495 0.60505425 0.9513964 0.74450973 0.92708948 0.98869984 0.98081253] is obtained in the ninth iteration. In a similar way to all the above-mentioned cases (Figs. 5 and 8), the four concepts that rapidly tend to 1 are: deterioration of health (S5), increased vulnerability of communities (S6), increased demand for water (S2), and decreased water quality (S4) (Fig. 11).

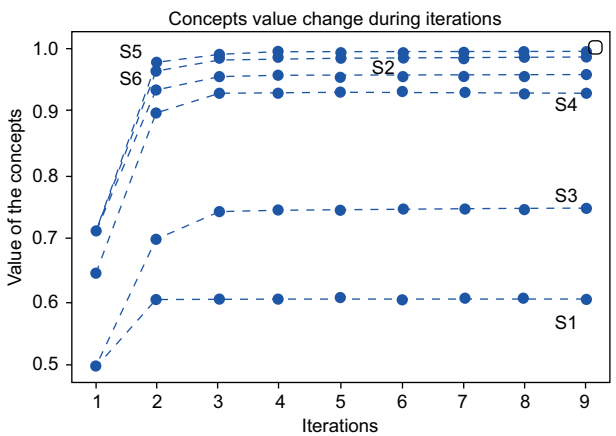

Fig. 11. Tendency to the equilibrium state in the social subsystem.

3.4.3 Provider and climate subsystems

The initial state vector of this system is v0 = [1 0 0 0 0 0 0 0 0 0 0]. After 9 iterations the vector F9 = [1 0.7109495 0.7109495 0.7109495 0.95539954, 0.91316077 0.98297375 0.5 0.97903095 0.53742985] is obtained. The three concepts with the highest growth in descending order are: increased deficiency in infrastructure for the distribution of drinking water (P3); generation of social, political, economic and environmental problems (P5), and increased exploitation of supply sources (P1) (Fig. 12). This same behavior is observed in all the previous cases (Figs. 6 and 9).

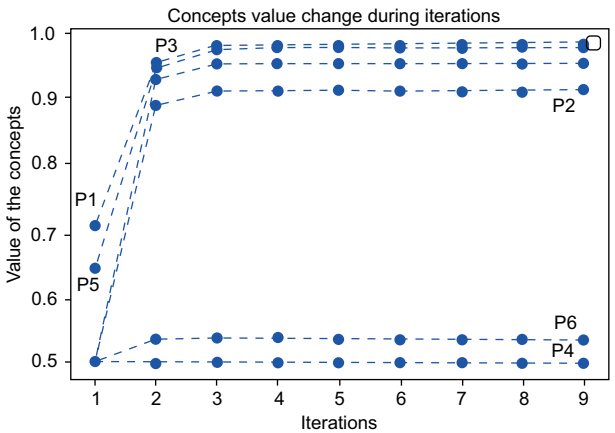

Fig. 12. Tendency to the equilibrium state in the provider subsystem.

4. Discussion

The results obtained so far indicate that the final states of equilibrium for each of the subsystems considered are undesirable. The question that follows is: What can we do about it? We consider our results to answer this question.

4.1 Centrality and feedback

The centrality of each of the concepts is calculated in order to detect which nodes are most important within each network. From (3), (4), and (5) centrality values are obtained for each subsystem. The three concepts with highest centrality are considered. Each of these three concepts also affects the others mutually through a feedback process within each network.

4.1.1 Environmental and climate subsystems

The three concepts with greatest centrality are (in descending order): decrease in water resources (cen(A3) = 8.7), loss of conservation soils (cen(A6) = 8.7) and ecosystems (cen(A4) = 7.5). In turn, Figure 1 shows that these three concepts create a feedback process given by A3, A4, A6, and A3. At this point it is very important to mention that there are more feedback processes such as A4, A6, A4 or A1, A4, A3, A1, etc. We will try to intervene in the feedback process that has more weight on the network.

4.1.2 Social and climate subsystems

According to their centrality, the main nodes of this network in descending order are: deterioration of health (cen(S5) = 7.2), increased vulnerability of communities (cen(S6) = 5.4), and increased demand for water (cen(S2) = 5.4). However, as can be seen in Figure 2, the concept S6 is outside a feedback process, therefore it is excluded from the selection, and the next node in importance is considered, which is the concept S4 with cen(S4) = 4.8. From these concepts, the feedback process [S5, S2, S4] is established.

4.1.3 Provider and climate subsystems

For this subsystem, the three most important concepts in descending order are: increase in infrastructure deficiency for the distribution of drinking water (cen(P3) = 6.6), increase in the exploitation of supply sources (cen(P1) = 6.3) and generation of social, political economic and environmental problems (cen(P5) = 5.4). The feedback process between these concepts follows the process [P3, P1, P5, P3] (Fig. 3).

4.2 Intervention

Three interventions are proposed to decision makers in order to reduce the negative effect generated in the aforementioned feedback processes. Each proposed measure is based on what we consider most urgent. To analyze what happens with each of the systems, the values of the causal relationships are considered to be worth {-1,0,1}.

4.2.1 Environmental and climate subsystems

For the environmental subsystem, among many other measures, an increase in the effective monitoring and conservation of conservation soils is urgent (Paavola and Primmer, 2019). Although there is legislation regarding the territorial limit of conservation soils, in practice these are invaded, either by the increase of the original population or by migration to the MCMA. The extinction of conservation soils is destroying the remaining ecosystems in this region together with the possibility of recharging the aquifers, as the water infiltration process is hampered. The effect of concept I1 on the feedback process can be observed in Figure 13. We consider that the causal relationships for this subsystem have integer values, as shown in Figure 13. The figure first shows the initial state in which concepts A3, A4, A6 are turned on, indicating that the problem is already present (as is currently the case) together with concept I1. So, v0 = [1 1 1 1]. The last component corresponds to the value of concept I1, which is the only one that will remain as a forcing throughout the iteration process. After the ninth iteration, the resulting vector is F9 = [0.76393078 0.58693396 0.58693396 1], where it can be observed that the value of the concepts shows a decreasing behavior (Fig. 14).

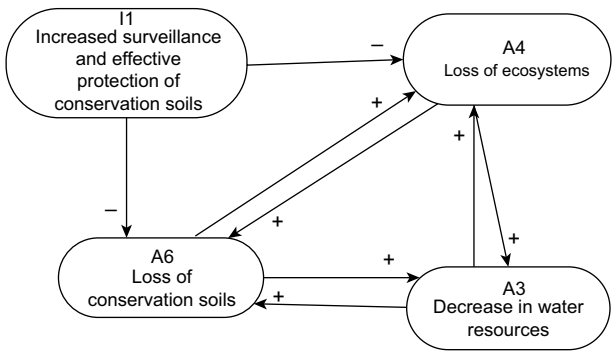

Fig. 13. Feedback process between concepts A3, A4 and A6 together with the action of intervention I1.

The concepts that most benefited from the suggested intervention are loss of conservation soils (A6) and loss of ecosystems (A4), whose obtained value is approximately 58 % of the original one. For its part, even though this measure does not drastically reduce the decrease in water resources (A3), it does contribute to its mitigation. To have a greater effect on this concept, more measures should be carried out to reduce its value to the maximum.

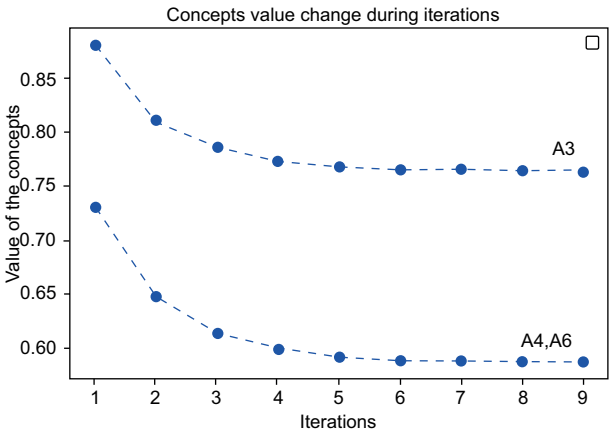

Fig. 14. The intervention concept I1: Increase in the surveillance and protection of conservation soils is enforced and maintained as a force.

4.2.2 Social and climate subsystems

As Soto and Herrera (2019) point out, the poorest population is more likely to obtain an unequal distribution of the liquid. The water supply in their homes is enabled for a few hours each week, in addition to the fact that the liquid is of poor quality. To cope with the shortage, it is common to store water as well as purchasing pipe and bottled water. These time and money factors disproportionately affect low income households.

From this study we have obtained that the deterioration in health due to lack of supply or lack of drinking water is one of the most urgent problems to be addressed. Figure 15 reconsiders the feedback process between concepts S2, S4 and S5, together with intervention I2. If we start from the initial state vector in which all concepts are on, and we maintain only I2 as forcing, the initial state vector is v0 = [1 1 1 1]. After the ninth iteration, the state vector is F9 = [0.4124232 0.60166935 0.64603877 1]. Figure 16 shows that the value of the concepts decreases, with S2 being the most benefited, followed by S3 and S5. If the intervention proposal we present is carried out, the amount of water required by the system would be significantly reduced (S2), which in turn would improve the quality of the water that the population consumes (S4), contributing positively to a reduction in deterioration of health (S5).

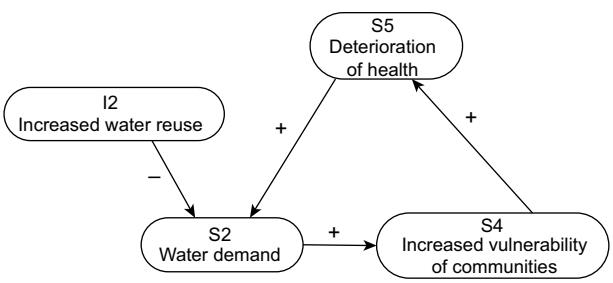

Fig. 15. Feedback process between concepts S2, S4 and S5 together with the action of the I2 intervention.

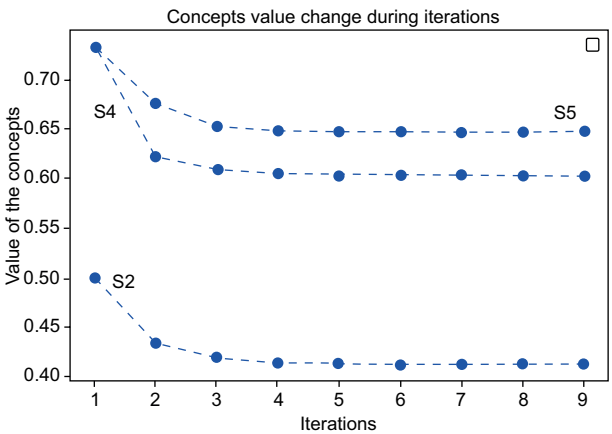

Fig. 16. The concept of intervention I2: increase in the reuse of wáter is turned on and kept as a force.

4.2.3 Provider and climate subsystems

According to Mexico City's Water System (Sacmex), more than 40 % of the water that the city receives is lost in numerous leaks affecting the hydraulic system. Although the percentage will be different for the MCMA, this area faces great losses of the vital liquid due to damage to its infrastructure. The proposed I3 intervention aims to an increase in monitoring, supervision and effective control of drinking water leaks, for which it will be necessary to change the hydraulic network in areas with the greatest impact. Figure 17 presents the feedback process between nodes P1, P3 and P5, together with intervention I3. If we start from the initial state vector in which all the concepts are on and we keep I3 as a forcing concept, v0 = [1 1 1 1]. After the ninth iteration the state vector of the system is F9 = [0.34538251 0.3605633 0.42702804 1]. These

values indicate that the values of concepts P1, P3 and P5 have decreased to 34, 36 and 42% of their original value, respectively (Fig. 18).

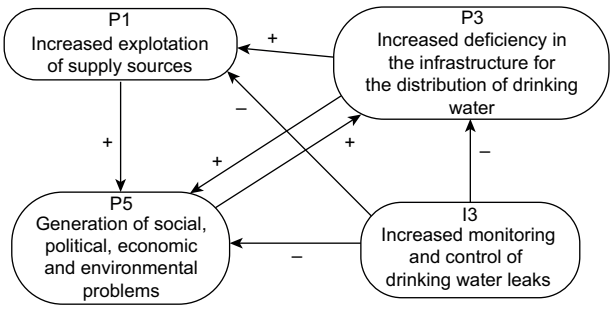

Fig. 17. Feedback process between concepts P1, P3 and P5 together with the action of intervention I3.

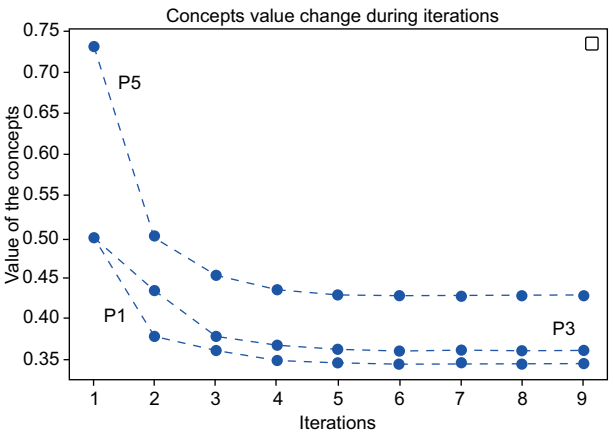

Fig. 18. The intervention concept I3 is enforced and maintained: Increase in the effective monitoring and control of drinking water leaks.

The concept that most benefited from this intervention is P1. The exploitation of supply sources would be reduced. This result has profound consequences, since projects are being carried out to bring water from more distant regions to meet the needs of the MCMA population. However, these solutions are far from being friendly to the environment, as well as being expensive. With this proposal, the aim is to make the most of the resources available, without disrupting other sources or endangering the current ones. For its part, a better infrastructure that does not present leaks will improve the distribution of drinking water, contributing in part to a reduction in social, political, economic and environmental problems.

5. Conclusions

Although the MCMA water supply service is a highly complex system, of which we ignore important information and where there are large ranges of uncertainty, through the use of fuzzy cognitive maps we have been able to qualitatively explore the possible effects of climate change on the environmental, social, and provider subsystems. This study has allowed us to determine which of the repercussions they will face from an increase in temperature and a decrease in precipitation are the most sensitive.

For the environmental-climatic subsystem, it was found that the most pressing problems are the decrease in water resources and the increase in the loss of conservation ecosystems and soils. The intervention that we recommend to decision-makers is to increase surveillance and effective protection of conservation soils and, where appropriate, recover invaded areas, so that both soils and the ecosystem may recover. If the proposed intervention is carried out, although the decrease in water resources and the loss of ecosystems and conservation soils will not be eliminated, the rate at which these processes occur will be reduced. If this intervention is strengthened with appropriate actions, the mitigation effect can be enhanced.

In the social-climatic system, the scenario constantly increasing temperature leads us to a state of equilibrium characterized by an increase in the of water by the population, a decrease in water quality and a deterioration in health. The intervention that we recommend to decision-makers is to work in mechanisms that permit the reuse of water in an efficient way.

For the provider-climate system, a sustained increase in temperature and a decrease in precipitation lead to an increase in the exploitation of supply sources and an increase in the deficiency of the infrastructure for the distribution of drinking water, with a consequent generation of social, political, economic and environmental problems. The control of water leaks will allow making the most of the

available resource without wasting it, thus reducing the exploitation of supply sources of water. This will contribute to a state in which there are no social conflicts due to the lack of the vital liquid.

Acknowledgments

This work was supported by CONACYT (CVU/31010). We thank the reviewers for their meticulous work and valuable contributions. We gratefully acknowledge Melanie Slone for the Spanish edition of this text; Corinne Otto for her dedication in the multiple readings she made of this article, invaluable suggestions as well as the English edition; Arlette Paola Olvera Fuentes for her valuable contributions to the improvement of this work, and Óscar Sánchez-Meneses for his contributions regarding the scenarios we analyze. Finally, acknowledge the Posgrado en Ciencias de la Tierra, UNAM, for the facilities provided for carrying out this research.

References

Adger WN. 2006. Vulnerability. Global Environmental Change 16: 268-281. https://doi.org/10.1016/j.gloenvcha.2006.02.006

Bahri O, Mourhir A, Papageorgiou EI. 2020. Integrating fuzzy cognitive maps and multi-agent systems for sustainable agriculture.Euro-Mediterranean Journal for Environmental Integration 5: 1-10. https://doi.org/10.1007/s41207-020-0143-8

Birkle P, Rodríguez VT, Partida EG. 1998. The water balance for the Basin of the Valley of Mexico and implications for future water consumption. Hydrogeology Journal 6: 500-517. https://doi.org/10.1007/s100400050171

Cakmak EH, Dudu H, Eruygur O, Ger M, Onurlu S, Tonguç Ö. 2013. Participatory fuzzy cognitive mapping analysis to evaluate the future of water in the Seyhan Basin. Journal of Water and Climate Change 4: 131-145. https://doi.org/10.2166/wcc.2013.029

Dickerson JA, Kosko B. 1994. Virtual worlds as fuzzy cognitive maps. Presence: Teleoperators and Virtual Environments 3: 173-189. https://doi.org/10.1162/pres.1994.3.2.173

Escolero Ó, Martínez S, Kralisch S, Perevochtchikova M. 2009. Vulnerabilidad de las fuentes de abastecimiento de agua potable de la Ciudad de México en el contexto de cambio climático. Informe final. Centro Virtual de Cambio Climático en México-Centro de Ciencias de la Atmósfera, Universidad Nacional Autónoma de México. Available at: http://www.cvcccm-atmosfera.unam.mx/sis_admin/archivos/agua_escolero__inffinal_org.pdf (accessed on September 12, 2020).

Fernández-Eguiarte A, Zavala-Hidalgo J, Romero-Centeno R, Conde-Álvarez AC, Trejo-Vázquez RI. 2015. Actualización de los escenarios de cambio climático para estudios de impactos, vulnerabiliad y adaptación. Centro de Ciencias de la Atmósfera, Universidad Nacional Autónoma de México-Instituto Nacional de Ecología y Cambio Climático, Secretaría de Medio Ambiente y Recursos Naturales. Available at: http://atlasclimatico.unam.mx/AECC/servmapas (accessed on September 13, 2020).

Gay-García C, Paz-Ortiz I, 2014. Stability analysis of climate system using fuzzy cognitive maps. In: Simulation and modeling methodologies, technologies and applications (Obaidat M, Filipe J, Kacprzyk J, Pina N, Eds.). Springer, Cham, 211-222. Advances in Intelligent Systems and Computing, 256). https://doi.org/10.1007/978-3-319-03581-9_15

Harary F, Norman RZ, Cartwright D. 1965. Structural models: an introduction to the theory of directed graphs. 2nd ed. John Wiley and Sons, New York, London, Sydney.

INEGI. February 2018. Delimitación de las zonas metropolitanas de México 2015. Available at: https://www.inegi.org.mx/contenidos/productos/prod_serv/contenidos/espanol/bvinegi/productos/nueva_estruc/702825006792.pdf (accessed on May 20, 2020).

IPCC. 2007. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change (Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averty KB, Tignor M, Miller HL, Eds.). Cambridge University Press, Cambridge, United Kingdom and New York. Available at: https://www.ipcc.ch/report/ar4/wg1/ (accessed on May 16, 2021).

IPCC. 2013. Climate change 2013: The physical science basis. Contributions of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (Stocker TF, Qin D, Plattner GK, Tignor M, Allen SK, Boschung J, Navels A, Xia Y, Bex V, Midgley PM, Eds.). Cambridge University Press, United Kingdom and New York. https://doi.org/10.1017/CBO9781107415324

IPCC. 2014. Climate change 2014: Synthesis report contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. (Pachauri RK, Meyer LA, Eds.). IPCC, Geneva. Available at: https://www.ipcc.ch/site/assets/uploads/2018/02/SYR_AR5_FINAL_full_es-.pdf (accessed on May 16, 2021).

Kosko B. 1986. Fuzzy cognitive maps. International Journal of Man-Machine Studies 24: 65-75. https://doi.org/10.1016/S0020-7373(86)80040-2

Kosko B. 1988. Hidden patterns in combined and adaptive knowledge networks. International Journal of Approximate Reasoning 2: 377-393. https://doi.org/10.1016/0888-613X(88)90111-9

Kosko B. 1992. Neural networks and fuzzy systems: A dynamical systems approach to machine intelligence. Prentice-Hall, New Jersey.

Kosko B. 1996. Fuzzy engineering. Prentice-Hall, New Jersey.

Lara-Rosano F, Gallardo A, Almanza S. 2018. Teorías, métodos y modelos para la complejidad social: Un enfoque de sistemas complejos adaptativos. Colofón Ediciones Académicas, Mexico City.

Mpelogianni V, Groumpos PP. 2018. Re-approaching fuzzy cognitive maps to increase the knowledge of a system.AI & Society 33: 175-188. https://doi.org/10.1007/s00146-018-0813-0

Nair A, Reckien D, van Maarseveen MF. 2019. A generalised fuzzy cognitive mapping approach for modelling complex systems.Applied Soft Computing 84: 105754. https://doi.org/10.1016/j.asoc.2019.105754

Nápoles G, Espinosa ML, Grau I, Vanhoof K. 2018. FCM expert: Software tool for scenario analysis and pattern classification base on fuzzy cognitive maps. International Journal on Artificial Intelligence Tools 27: 1860010. https://doi.org/10.1142/S0218213018600102

Özesmi U, Özesmi SL. 2004. Ecological models based on people's knowledge: A multi-step fuzzy cognitive mapping approach. Ecological Modelling 176: 43-64. https://doi.org/10.1016/j.ecolmodel.2003.10.027

Paavola J, Primmer E. 2019. Governing the provision of insurance value from ecosystems.Ecological Economics164: 106346. https://doi.org/10.1016/j.ecolecon.2019.06.001

Papageorgiou EI, Stylios C, Groumpos PP. 2006. Unsupervised learning techniques for fine-tuning fuzzy cognitive map causal links.International Journal of Human-Computer Studies64, 727-743. https://doi.org/10.1016/j.ijhcs.2006.02.009

Papageorgiou E, Kontogianni A. 2012. Using fuzzy cognitive mapping in environmental decision-making and management: A methodological primer and an application. In: International perspectives on global environmental change (Young SS, Silvern SE, Eds.). InTech, Rijeka, 427-450.

Paz-Ortiz I. 2011. Uso de mapas cognitivos para el estudio de la estabilidad en sistemas climáticos terrestres. M.Sc. thesis. Universidad Nacional Autónoma de México.

Paz-Ortiz I, Gay-García C. 2013. A fuzzy cognitive map for México City's water availability system. In: SIMULTECH 2013–3rd International Conference on Simulation and Modeling Methodologies, Technologies and Applications, 495-503. Available at: https://pdfs.semanticscholar.org/d9a8/e5f071c5a82a742a316c-72417fa0b7afccdb.pdf (accessed on January 30, 2020).

Paz-Ortiz I, Gay-García C. 2014. Using fuzzy cognitive mapping and nonlinear Hebbian learning for modeling, simulation and assessment of the climate system, based on a planetary boundaries framework. In: Proceedings of the 4th International Conference on Simulation and Modeling Methodologies, Technologies and Applications, Vienna, Austria, 852-862. https://doi.org/10.5220/0005140608520862

Peláez CE, Bowles JB. 1996. Using fuzzy cognitive maps as a system model for failure modes and effects analysis. Information Sciences 88: 177-199. https://doi.org/10.1016/0020-0255(95)00161-1

Poomagal S, Sujatha R, Kumar PS, Vo DVN. 2021. A fuzzy cognitive map approach to predict the hazardous effects of malathion to environment (air, water and soil). Chemosphere 263: 127926. https://doi.org/10.1016/j.chemosphere.2020.127926

Ross TJ. 2010. Fuzzy logic with engineering applications. John Wiley & Sons. 3rd ed. The Atrium, Southern Gate, Chichester, West Sussex.

Sánchez Ó. 2017. Estimación de rangos de incertidumbre en las fechas para alcanzar los valores de incremento en la temperatura promedio global 1.0, 1.5 y 2.0 ºC y las implicaciones para la República Mexicana como producto de la Sexta Comunicación Nacional de México ante la Convención Marco de las Naciones Unidas sobre el Cambio climático. Instituto Nacional de Ecología y Cambio Climático-Programa de las Naciones Unidas para el Desarrollo.

SEMARNAT. 2018. Diario Oficial de la Federación. Acuerdo por el que se actualiza la disponibilidad media anual de agua subterránea de los 653 acuíferos de los Estados Unidos Mexicanos, mismos que forman parte de las regiones hidrológico-administrativas que se indican. Secretaría de Medio Ambiente y Recursos Naturales, Mexico. Available at: https://www.dof.gob.mx/nota_detalle_popup.php?codigo=5510042 (accessed on May 16, 2021).

Singh PK, Chudasama H. 2017. Assessing impacts and community preparedness to cyclones: A fuzzy cognitive mapping approach.Climatic Change 143: 337-354. https://doi.org/10.1007/s10584-017-2007-z

Soler LS, Kok K, Camara G, Veldkamp A. 2012. Using fuzzy cognitive maps to describe current system dynamics and develop land cover scenarios: a case study in the Brazilian Amazon. Journal of Land Use Science 7: 149-175. https://doi.org/10.1080/1747423X.2010.542495

Soto G, Herrera M. 2009. Estudio sobre el impacto del cambio climático en el servicio de abasto de agua de la Zona Metropolitana de la Ciudad de México. Available at: http://www.ampres.com.mx/assets/impacto-en-el-cambio-climatico-en-el-df.pdf (accessed on January 30, 2020)

Soto G, Herrera M. 2019. Cambio climático y agua en ciudades: impactos en la ciudad de México. Aspectos científicos y políticas públicas. UAM-Conagua, Mexico City. Available at: https://www.casadelibrosabiertos.uam.mx/contenido/contenido/Libroelectronico/Cambio-climatico-y-agua-en-ciudades.pdf (accessed on May 16, 2021).

Stylios CD, Groumpos PP. 2004. Modeling complex systems using fuzzy cognitive maps. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans 34: 155-162. https://doi.org/10.1109/TSMCA.2003.818878

World Bank. 2013. Urban water in the Valley of Mexico: A green path for tomorrow? World Bank, Washington. Available at: https://documents.worldbank.org/en/publication/documents-reports/documentdetail/295391468049168354/mexico-agua-urbana-en-el-valle-de-mexico-un-camino-verde-para-ma-241-ana (accessed on May 16, 2021).

Zadeh LA. 1965. Fuzzy sets. Information and Control 8: 338-353. https://doi.org/10.1142/9789814261302_0021

Atmósfera 36(2), 317-327 (2023)
https://doi.org/10.20937/ATM.53032

CONTENTS

Remote sensing of atmospheric nitrogen dioxide in an urban area in central northern Mexico

Julio HERNÁNDEZ-DE-TORRES[1]*, Clara ÁVILA-MONTOYA[1], Luis Felipe LASTRAS-MARTÍNEZ[1], Alfonso LASTRAS-MARTÍNEZ[1], David Enrique FLORES-JIMÉNEZ[2], Abraham CÁRDENAS[3] and Marcos ALGARA-SILLER[3]

[1]*Instituto de Investigación en Comunicación Óptica, Universidad Autónoma de San Luis Potosí, Av. Karakorum 1470, Lomas cuarta sección, 78210 San Luis Potosí, San Luis Potosí, México.*
[2]*Instituto de Ingeniería, Universidad Autónoma de Baja California, Blvd. Benito Juárez y Calle Normal, Colonia Insurgentes Este, 21280 Mexicali, Baja California, México.*
[3]*Centro de Investigación y Estudios de Posgrado Facultad de Ingeniería, Universidad Autónoma de San Luis Potosí, Manuel Nava 8, Zona Universitaria, 78290 San Luis Potosí, San Luis Potosí, México.*
*Corresponding author: jcmex15@hotmail.com

Received: January 20, 2021; accepted: July 27, 2021

RESUMEN

Se midió la concentración atmosférica de dióxido de nitrógeno (NO_2) en el área urbana de San Luis Potosí (México) mediante la técnica de espectroscopia de absorción óptica diferencial (DOAS, por su sigla en inglés), utilizando una configuración cenital. Las mediciones se realizaron de julio a agosto de 2015. En esta técnica, se recolecta la luz solar dispersa en la atmósfera por medio de un telescopio y se hace incidir sobre un espectrómetro que cuenta con un arreglo lineal de detectores a nivel del suelo. Durante el periodo de medición, los niveles máximos de NO_2 (2.3×10^{16} moleculas cm^{-2}) se encontraron alrededor de las 11:00 LT, mientras que los niveles de NO_2 cayeron en los días con velocidades de viento inferiores a 1.5 m s^{-1}. Los niveles de NO_2 se compararon con el comportamiento de los datos meteorológicos como velocidad del viento, humedad relativa y temperatura superficial obtenidos del modelo de pronóstico de investigación meteorológica (WRF).

ABSTRACT

Nitrogen dioxide (NO_2) was measured in the urban area of San Luis Potosí (Mexico) using a remote sensing technique based on differential optical absorption spectroscopy (DOAS) in zenith-sky configuration. Measurements were taken from July to August 2015. In this technique, scattered sunlight in the atmosphere is collected by a telescope and carried onto a linear array-based spectrometer at ground level. During the measurement period, the maximum NO_2 levels (2.3×10^{16} molecules cm^{-2}) were found at around 11:00 LT, whereas NO_2 levels fell on days with wind speeds lower than 1.5 m s^{-1}. NO_2 levels were compared and explained with the behavior of meteorological data such as wind speed, relative humidity, and surface temperature obtained from the Weather Research and Forecasting (WRF) model.

Keywords: optical absorption, spectroscopy, atmospheric gases.

1. Introduction

Pollutants are monitored in many countries because of their risk to human health and environmental impact (SEMARNAT, 2013). Most are directly derived from combustion processes at fixed or mobile sources. The main criteria pollutants around the world are carbon monoxide (CO), sulfur dioxide (SO_2), nitrogen dioxide (NO_2), tropospheric ozone

(O_3), and particulate matter (PM) (Tyler et al., 2013). The health effects depend on exposure time and concentration (Mapoma et al., 2014). In 2012, the World Health Organization (WHO) estimated that one out of every nine deaths resulted from air pollution-related conditions (WHO, 2016).

Combustion processes are involved in the emission of most pollutants and increase the amount of gases in the atmosphere at different scales. In particular, SO_2, NO_x, and CO are generated by industrial activities and the burning of fossil fuels. The latter two also contribute to the formation of tropospheric ozone (O_3) (Lazaridis, 2011) through photochemical reactions. This process depends on energy from solar radiation and is further exacerbated by increasing concentrations of NO_2 and volatile organic compounds in the atmosphere (Sillman, 1999). It is important to analyze the interaction of these pollutants with meteorological parameters on a local scale.

Specifically, understanding the behavior or tendency of NO_2 levels in urban areas is key for monitoring and mitigation strategies because of its relation with other pollutants and greenhouse gases (Gaffney and Marley, 2003). This knowledge, for example, has enabled the implementation of strategies to reduce emissions in several European cities (Henschel et al., 2015). On the other hand, in emerging cities and countries where industrial activities have increased in recent years, such as in Beijing, China, severe pollution warnings are emitted (Hou et al., 2016). In urban areas of Japan such as Tokyo and Osaka, a correlation was found between increasing NO_2 concentrations and temperature, exacerbating heat islands (Gotoh, 1993).

The remote sensing of atmospheric NO_2 and O_3, among other atmospheric pollutants, can be performed using differential optical absorption spectroscopy (DOAS) (Platt and Stutz, 2008). In this technique, scattered light from the sun is carried onto a linear array-based spectrometer. Many measurement campaigns around the world have analyzed NO_2 in the atmosphere using this technique, especially in large cities in Europe (Platt and Perner, 1980), China (Bernard et al. 2015), the United States (Spinei et al., 2015), and in several Latin American urban areas and volcanic regions (Grutter et al., 2008; Frins et al., 2011).

In Mexico City and its surroundings, the presence of NO_2 has similarly been analyzed by spectroscopy techniques (Grutter et al., 2008, Melamed et al., 2009; Rivera, 2013). The DOAS technique has mostly been used in Mexico City to identify the vertical distribution of NO_2 in the troposphere and its relationship with meteorological parameters such as wind speed and temperature (Melamed et al., 2009; Arellano et al., 2016). Atmospheric monitoring in Mexico City has been carried out since 1986, when the Red Automática de Monitoreo Atmosférico (Automatic Atmospheric Monitoring Network, RAMA) began to measure several pollutants (SEDEMA, 2017). The Sistema Nacional de Información de la Calidad del Aire (National Air Quality Information System, SINAICA) provides access to a database with data from every sensing station in Mexico (INECC, 2017). However, air quality has not been consistently monitored over the years. In this context, the use of atmospheric models to analyze the transport of pollutants and their interaction with meteorological events has been very useful to understand these realities. In Mexico City, local atmospheric conditions have been simulated by the WRF/Chem model, and a high correlation of meteorological parameters and NO_2 was found when compared to observations (Zhang et al., 2009).

In 2008, 40% of NO_x emissions in Mexico were estimated to come from mobile sources (SEMARNAT, 2013). The growth of cities has favored an increase in NO_2 emissions. Measured emissions in 2013 and emissions projected to 2030 indicate that levels of this pollutant will continue to occupy the third place (9.84×10^5 Mg year^{-1} in 2013 and 5.08×10^5 Mg year^{-1} in 2030) behind CO (3.28×10^6 Mg year^{-1} in 2013 and 2.16×10^6 Mg year^{-1} in 2030) and CO_2 (1.48×10^8 Mg year^{-1} in 2013 and 2.47×10^8 Mg year^{-1} in 2030) (INECC, 2014).

The measured atmospheric NO_2 concentrations in several Mexican states are barely under the limit set by official Mexican standards. Although most large Mexican cities have an air quality monitoring system, the monitoring networks of several cities such as Guadalajara, Tabasco, and San Luis Potosí, among others, need to be updated due to maintenance problems and the quality of measurements according to a report carried out in 2009 (INE, 2011).

In particular, the municipality of San Luis Potosi has a high level of NO_x emissions (18%), equaling 22 614.56 Mg in 2011 (SEGAM, 2013), largely due

to vehicle combustion processes. In the span of 20 years (1990 to 2010), the population grew 32% (IMPLAN, 2016). At the same time, the vehicle density in the metropolitan zone of San Luis Potosí-Soledad de Graciano Sánchez increased 33% from 2005 to 2015 (INEGI, 2017).

The increase in vehicles and fossil-fuel-based industrial activities in the city of San Luis Potosí (INEGI, 2014) has correspondingly led to an increase in mobile and point sources of NO_2. Therefore, the objective of the present study is to quantify NO_2 in the atmosphere of San Luis Potosi using a spectroscopy technique and to analyze and explain its variability in relation to the behavior of several meteorological parameters obtained from the WRF model.

2. Methodology

2.1 Theoretical model

The DOAS technique (Noxon, 1975; Solomon and Schmeltekopf, 1987; Platt and Stutz, 2008) utilizes the structured absorption of many trace gases in the UV spectral region. It relies on the application of the Beer-Lambert law to the atmosphere considering a limited range of wavelengths. This law states that the radiant intensity traversing a homogeneous medium decreases exponentially with the product of the extinction coefficient, the number density and path length. The Beer-Lambert law applied to the atmosphere is written as follows (Danckaert et al., 2017):

$$\boldsymbol{I}(\lambda) = \boldsymbol{I_O}(\lambda) exp\left(-\sum_{i=0}^{n} \sigma_i(\lambda) c_i \right) \quad (1)$$

where $I(\lambda)$ is the measured spectrum after extinction in the atmosphere; $I_o(\lambda)$ is the spectrum at the top of the atmosphere, without extinction; $\sigma_i(\lambda)$ (in cm^2 $molecule^{-1}$) is the absorption cross section of the i-th species, which is wavelength dependent; and c_i (in cm^2 $molecule^{-2}$) is the column density of the i-th species defined by the concentration integrated along the light path in the atmosphere (Platt and Stutz, 2008).

From Eq. (1), we define the optical density as:

$$\tau = \ln\left(\frac{\boldsymbol{I_O}(\lambda)}{\boldsymbol{I}(\lambda)} \right) = \sum_{i=0}^{n} \sigma_i(\lambda) c_i \,. \quad (2)$$

The basic idea of the DOAS technique is to separate the broad and narrow spectral structures of the measured spectra in order to isolate the narrow structures associated to the different chemical species contained in the atmosphere. To perform that procedure, it is necessary to know the absorption cross section of each species in the spectral range of interest (taken from literature) and fitting the measured spectra by using Eq. (2) and determine numerically the optical density and the column density of each species. In the procedure, we have assumed that the absorption cross sections are independent of temperature and pressure. Thus we use the concept of Slant Column Density (SCD) to refer to c_i (Danckaert et al., 2017).

2.2 Experimental set-up and NO_2 measurement via DOAS

The telescope used to measure DOAS is composed by a lens with a 2.54-cm diameter and a focal length f = 50 mm coupled to an optical fiber with a diameter d = 1000 mm (Fig. 1). With these parameters, the field of view (FOV) of our instrument defined as d/f (Frins et al., 2006) gives a value of approximately 1.1º. The light captured by the optical fiber was guided to a digital UV-Vis lightweight spectrometer (B&W Tek, model BRC641E) with Czerny-Turner configuration and a one-dimensional charge-coupled device (CCD) array (2048 pixels) with a resolution of 0.3 nm full width at half maximum (FWHW) and a spectral range from 198 to 450 nm. The integration time was 100 ms to prevent saturation throughout the day. The acquisition program averages 100 spectra and saves the

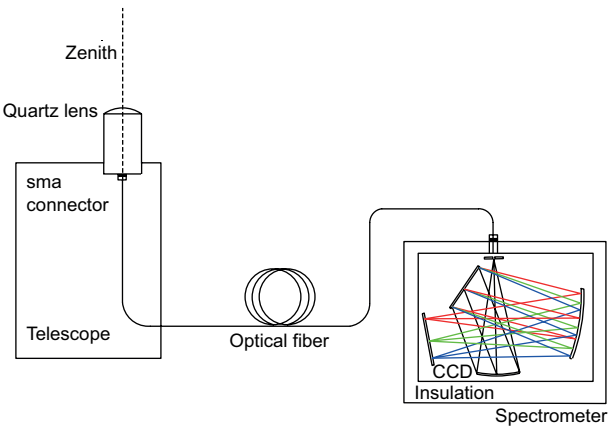

Fig. 1. Schematic diagram of the setup used to perform the DOAS measurements.

resultant spectrum. The system is cooled at constant temperature of 18 ºC by a thermoelectric regulator to minimize the dark current (Arellano et al., 2016).

To determine the NO_2 composition, we have analyzed spectra in the range from 360 to 420 nm. In this region, the spectral composition of NO_2 has been fully identified (Rublev et al., 2003). Two spectra were considered for DOAS analysis: one (reference spectrum) acquired at zenith (around noon) and the other measured also at zenith but temporarily displaced from the reference. The latter spectrum is normalized by dividing the reference spectrum in order to discriminate slow spectral information structures.

3. Data analysis

3.1 Study area

The City of San Luis Potosí is located in central northern Mexico (22º N, 100º W) at 1860 masl (INE, 2011). The prevailing climate is dry and semi-dry. The statewide average annual temperature is 21 ºC. The average minimum temperature is −5 ºC, occurring in January, and the average maximum temperature is around 38 ºC from May to July. The rainy season spans the summer from June to September, and the average precipitation is about 300 mm annually (Rivera, 2014).

From July to August 2015, a sampling campaign was carried out at the Instituto de Investigación en Comunicación Óptica (Research Institute for Optic Communications, IICO). In Figure 2, the locations of IICO and the meteorological station (northeast of the sampling point) are shown. Data from the meteorological station were used to validate the data obtained by using WRF. Subsequently, WRF was used to obtain the meteorological data at the IICO site where NO_2 was measured.

3.2 Spectral analysis

The spectra were measured statically between 8:00 and 17:00 LT (UT-6). The electronic offset induced by the CCD dark current of our system was obtained by

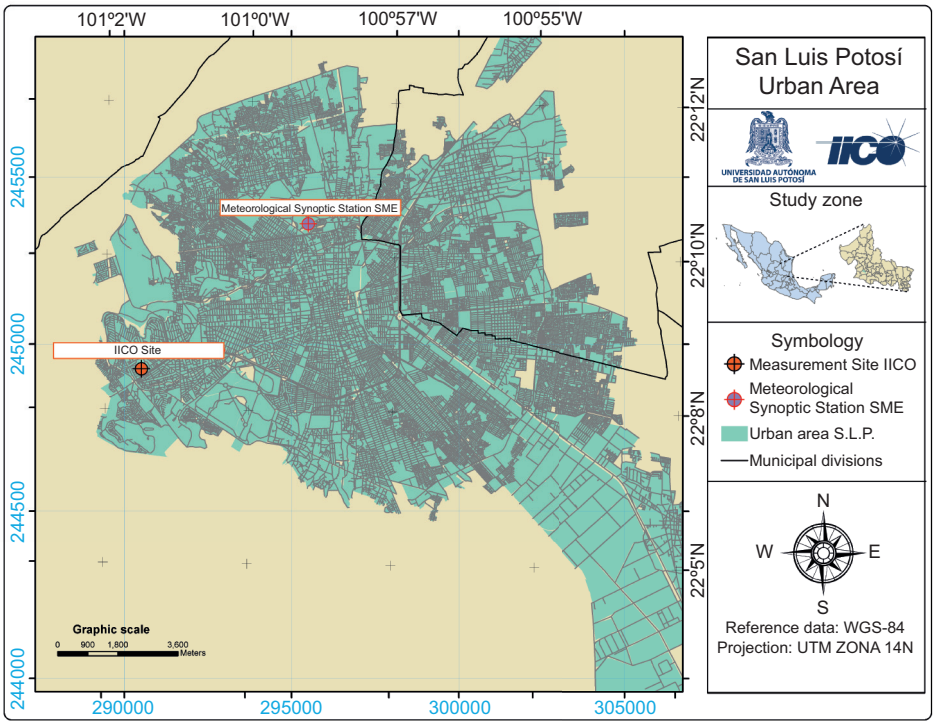

Fig. 2. Map of the city of San Luis Potosí metropolitan area. Red dots indicate the location of the DOAS instruments and the meteorological station.

obstructing the light entrance of the spectrometer and measuring the output signal of the CCD array. This procedure was performed every day at the beginning of each set of measurements. The offset signal was numerically subtracted from the reference spectrum (taken at noon) and from every spectrum taken during the day. After that, the spectra were divided to normalize them. A low-pass filter was used to separate the broad and narrow spectral bands (Rivera et al., 2013). By using Eq. (1), the experiments and the cross-section of the gases of interest, we evaluated numerically the SCDs of each species.

The numerical approach was performed by using the QDOAS software (Danckaert et al., 2017) and high-resolution differential cross-sections in the spectral range from 360 to 420 nm of NO_2 (Vandaele et al., 1998), O_3 at 221 and 241 K (Burrows et al., 1999), oxygen dimer (O_4) (Hermans et al., 1999) and a Ring spectrum generated at 273 K (Kurucz, 1995).

3.3 Urban atmospheric conditions generated by the WRF model

Atmospheric conditions were simulated by the WRF model due to the lack of meteorological data at the location of the experiment. The WRF model is a non-hydrostatic, numerical and three-dimensional model that uses sigma levels in physical equations to determine a vertical distribution in which the dynamic of the atmosphere is studied under different physical parameterization schemes (Skamarock et al., 2008). It was developed by Pennsylvania State University and the National Center for Atmospheric Research (NCAR), among others (Guichard et al., 2003; Skamarock et al., 2005). The global weather data fed to the model were obtained from the National Center for Environmental Prediction (NCEP) and had a spatial resolution of 100 × 100 km with time intervals of 6 h for every input variable considered (Kalnay et al., 1996).

Physical variables are also included in the WSM 6-class graupel scheme, such as condensation, precipitation, and latent heat (Lim and Hong, 2005). The Rapid Radiation Transfer Model (RRTM) was incorporated to consider the influence of long-wave radiation and the Dudhia (1989) scheme to consider the influence of short-wave radiation.

The model was calculated for three surfaces: the largest was 1500 × 1500 m, the other two were nested domains of 500 × 500 and 250 × 250 m. The experiment began on July 27, 2015 but the modeling was carried out from three days before (July 24) to stabilize the model. The simulation ended on August 16, 2015. Meteorological data were extracted from the WRF model for the same location where NO_2 was measured (Fig. 2).

The analyzed parameters were solar radiation, temperature, wind speed, and relative humidity because of their importance in the creation and destruction of NO_2, as emphasized by Pineda-Martínez et al. (2012).

The only place where a meteorological synoptic station (MSE) may be found is to the north of the city, for which the WRF results were validated in that location by comparing the meteorological data measured with those generated by the model. Subsequently, the NO_2 levels measured south of the city were compared with the meteorological data calculated (Fig. 2) to analyze the relation between these parameters. The BIAS and root mean square error (RMSE), whose reliability was confirmed by Pineda-Martinez et al. (2012, 2014), were calculated.

4. Results and discussion

4.1 WRF simulations

Figure 3 shows some examples of the meteorological station data and the corresponding simulations obtained by WRF for wind speed, temperature, and relative humidity for July 31, July 29 and July 29, respectively. The model predicts well the general behavior for wind speed, temperature and relative humidity, respectively in the period from 8:00 to 13:00 LT. The modeling nested domain was carried out on a 250 × 250 m grid.

The statistical confidence of wind simulations decreases at resolutions lower than 3 km, especially for higher run periods (Pineda-Martínez et al., 2012). In this case, the reliability of the WRF model in predicting wind speed was highest on July 31, with a RMSE of 0.91 and a BIAS of −0.54 m s^{-1}, whereas the lowest confidence was recorded on August 1, with a RMSE of 2.05 and a BIAS of −1.67 m s^{-1}. It is difficult to analyze urban conditions at this resolution because, near the surface, there is influence from convective processes, turbulent effects of air flow, and buildings (Pineda-Martínez et al., 2014). The behavior of wind

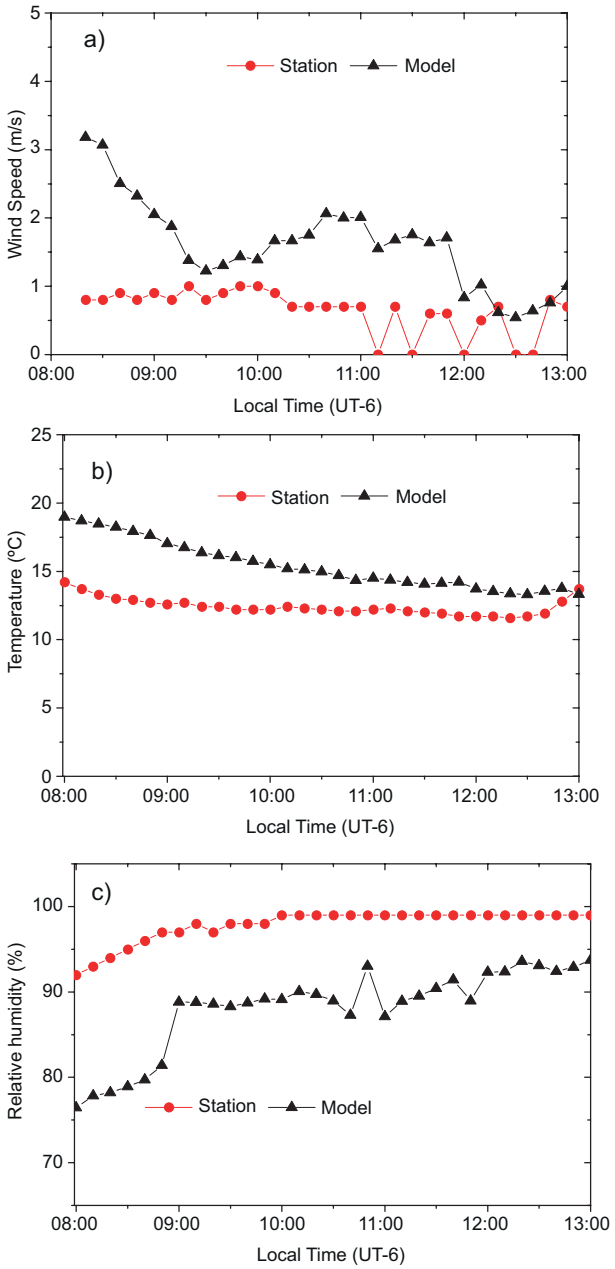

Fig. 3. Comparison between data from the meteorological station and data generated by the WRF model for: (a) wind speed on July 31, (b) temperature on July 29, and (c) relative humidity on July 29.

speed differed daily due to local conditions, whereas the behavior of temperature showed some common patterns.

Relative humidity explains the aforementioned case. For example, at almost the same time when temperature was underestimated by the model with respect to the meteorological station (13:00 h), relative humidity was overestimated (Fig. 3c). This could result from the influence of the mountains surrounding the urban area and the trade winds from the eastern Sierra Madre (Pineda-Martinez et al., 2014), causing a mountain breeze effect in the afternoon.

However, the model can explain meteorological behavior with good confidence in the morning until approximately 13:00 LT. After this time, the recorded relative humidity tended to decrease, while the recorded temperature tended to increase. The model, on the other hand, indicated water vapor saturation (high relative humidity reaching 100%) and temperature decrease (Fig. 3b, c).

4.2 NO_2 optical density

To illustrate the performance of the technique we show in Figure 4, optical density was obtained for: (a) NO_2, (b) O_3, (c) O_4, and (d) the ring spectrum. The spectra were measured on August 1, 2015 at 10:52 LT, with 1-min time integration in the wavelength range from 360 to 420 nm. Black spectra are the contributions to the optical density of each species and the red lines are the fitted spectra. The numerical approach was obtained by using QDOAS software. We found that in this range the optical track of NO_2 was well resolved. Note that the contribution of O_3 and O_4 in this range is not evident. The contribution of the ring spectrum (Fig. 4d) is clear and it is dominated by sharp structures below 400 nm. To have further evidence of the performance of our technique, we show in Figure 5 the contribution of O_3 and O_4 bands in the range from 340 to 370 nm, where the presence of these gases is evident.

4.3 NO_2 column behavior and interaction with meteorological parameters

Black circles in Figure 6 show the SCD of NO_2 obtained by QDOAS on (a) July 28, (c) July 29, and (e) July 31 between 8:00 and 16:00 LT (UT-6). All the SCD curves increase their value during the first 2 h and reach their maximum at (a) 11:00, (c) 9:45 and (e) 9:30 LT. After that, the curves start to decrease. The maximum SCD value for each day is (a) 2.5×10^{16}, (c) 2.0×10^{16} and (e) 2.1×10^{16} molecules cm^{-2}. These SCD values of NO_2 are below those measured at other cities: for example, in Mexico City, the NO_2

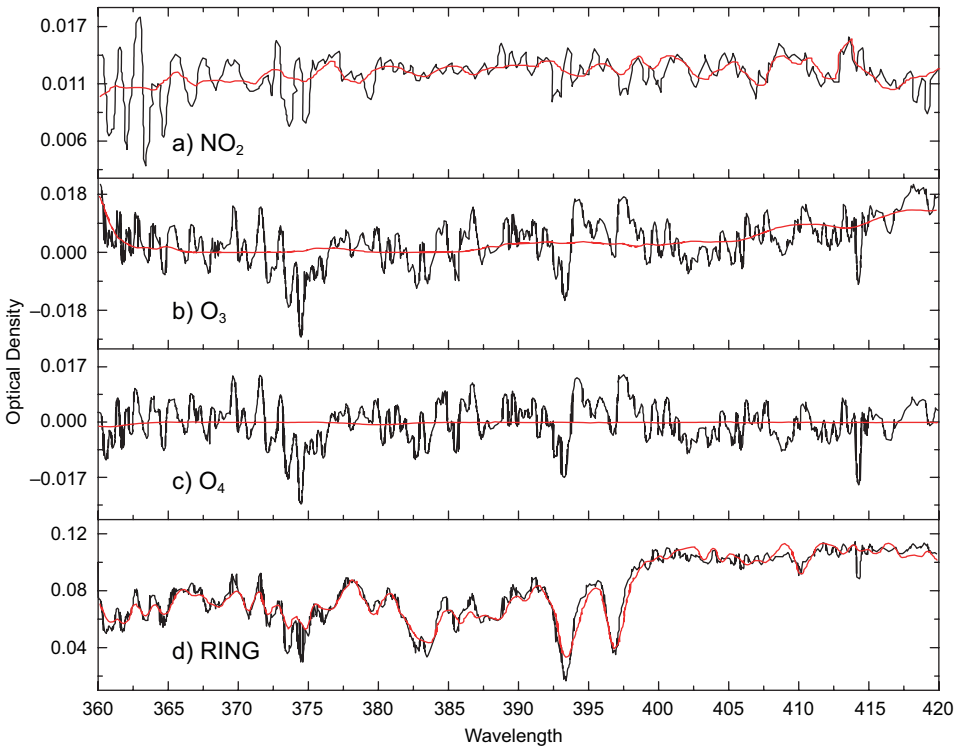

Fig. 4. Optical activity measured on August 1, 2015 at 10:52 LT with 1-min time integration for (a) NO_2, (b) O_3, (c) O_4, and (d) the ring spectrum for the wavelength range from 360 to 420 nm. Black lines are the spectra obtained numerically from the measured spectrum, the absorption cross section of the gases of interest and Eq. (2). The red lines are the fitted spectra. The numerical approach was obtained by using the QDOAS software.

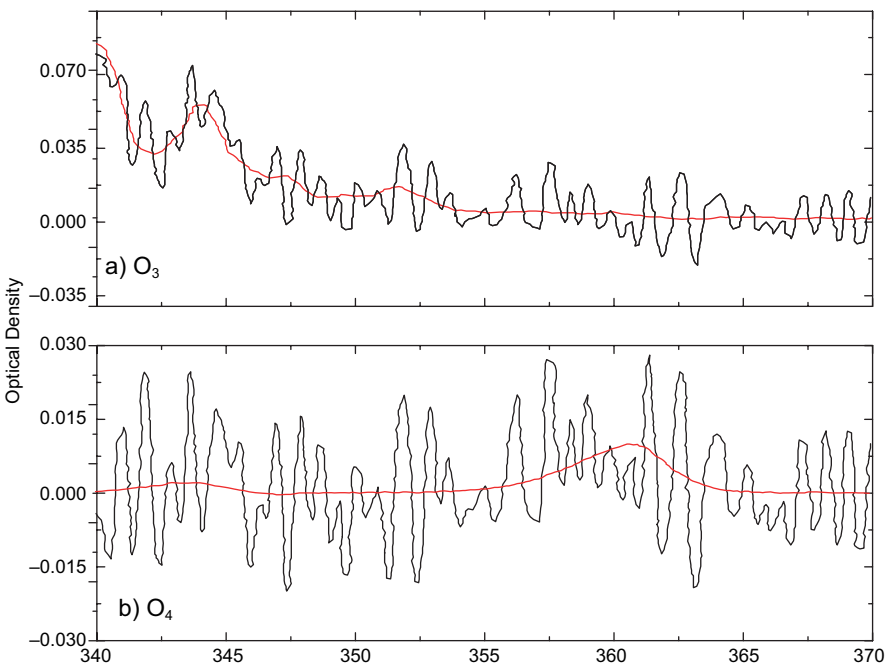

Fig. 5. O_3 and O_4 bands in the range from 340 to 370 nm.

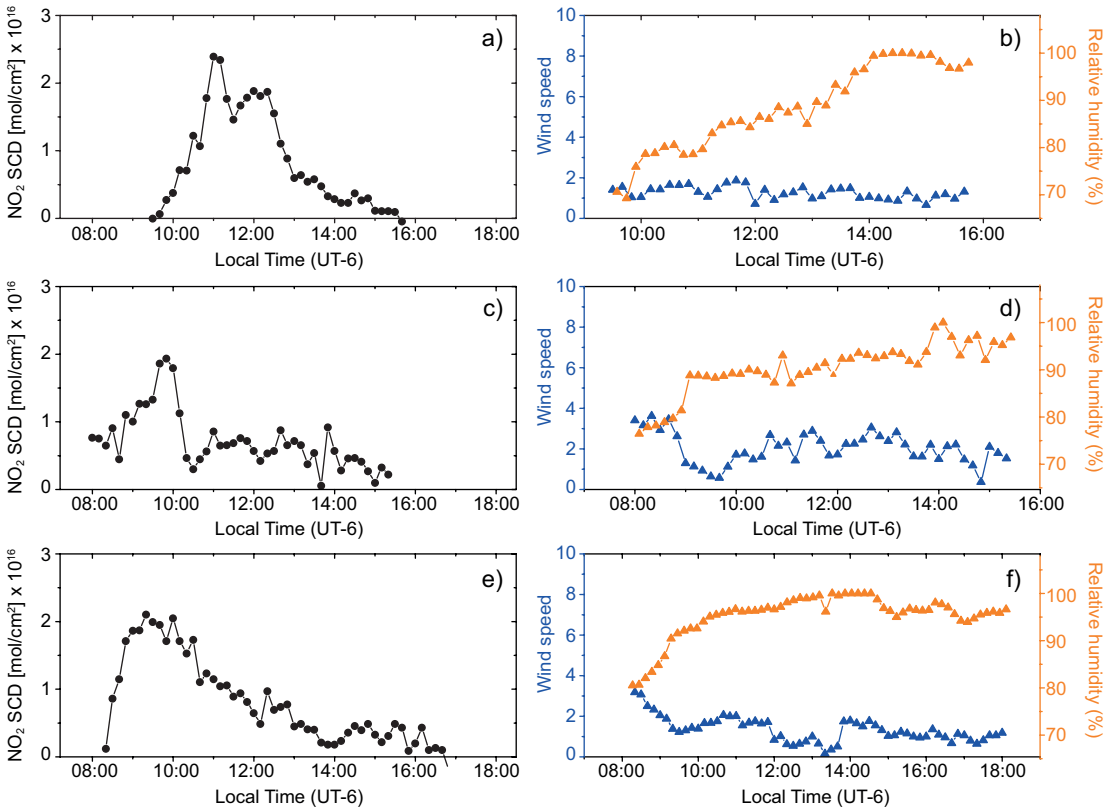

Fig. 6. SCD of NO_2 measured on (a) July 28, (c) July 29, and (e) July 31, and relationship between wind speed and relative humidity estimated by WRF on (b) July 28, (d) July 29, and (f) July 31.

column oscillates around 2.0×10^{17} molecules cm^{-2} according to the measurements of Rivera et al. (2013).

In order to correlate the SCD behavior with meteorological parameters, we also show in Figure 6 meteorological data generated by the WRF model for the same days and hours of the SCDs data. Figure 6 includes wind speed (blue triangles) and relative humidity (orange triangles). The highest values of the NO_2 column (around 2.0×10^{16} molecules cm^{-2}) were found around (a) 11:00, (c) 9:30 and (e) 9:00 LT. These peaks were possibly influenced by minimum wind speed around 9:30 LT, or by human activities.

The decrease of the NO_2 column can also be associated to the increase of relative humidity during the day. The levels of concentration for air pollution depend on the levels of emissions but also meteorological conditions as wind speed, temperature, rainfall, and relative humidity, which have influence in the formation of secondary pollutants and atmospheric dispersion (Crutzen, 1979; Countess et al., 1981, Habeebullah et al., 2015).

5. Conclusions

The presence of NO_2 in the atmosphere was characterized for the first time using remote sensing techniques (DOAS) in San Luis PotosÍ, Mexico. Additionally, meteorological data such as wind speed, relative humidity, and surface temperature were estimated by the WRF. The parameters estimated by the WRF model were reliable from 8.:00 to 12:30 lt. Later in the day, it was necessary to consider data from a meteorological station to describe the variability of NO_2.

The highest sensitivity of the technique for the detection of NO_2 was obtained for the wavelength range from 360 to 420 nm. The method utilized

herein represents an alternative for monitoring atmospheric gases in the city of San Luis Potosí. It is suitable for understanding the distribution of gases at specific sites (based on surface measurements) as well as across regions. This method can be extended to analyze other gases such as SO_2, O_3, and BrO at different wavelengths in the UV rage.

Acknowledgments

J.H.T and L.F.L.M. contributed equally to this work. The authors thank Caroline Fayt and Michel van Roozendael from the Belgian Institute for Space Aeronomy for the QDOAS software, the Coordinación General del Servicio Meteorológico Nacional (CGSMN) of the Comisión Nacional del Agua for providing meteorological data, and J. Rodríguez, M. Zúñiga, L. E. Guevara-Macías, E. Ontiveros and F. Ramírez-Jacobo for their skillful technical support. J.H.T. acknowledges CONACYT (CVU 334690) for the scholarship that supported his Ph.D. studies.

References

Arellano J, Krüger A, Rivera C, Stremme W, Friedrich MM, Bezanilla A, Grutter M. 2016. The Max Doas network in Mexico City to measure atmospheric pollutants. Atmósfera 29: 157-167. https://doi.org/10.20937/ATM.2016.29.02.05

Bernard F, Cazaunau M, Grosselin B, Zhou B, Zheng J, Liang P, Zhang Y, Ye X, Daële V, Mu Y, Zhang R, Chen J, Mellouki A. 2015. Measurements of nitrous acid (HONO) in urban area of Shanghai, China. Environmental Science and Pollution Research 23: 5818-5829. https://doi.org/10.1007/s11356-015-5797-4

Burrows JP, Richter A, Dehn A, Deters B, Himmelmann S, Voigt S, Orphal J. 1999. Atmospheric remote-sensing reference data from GOME: Part 2. Temperature-dependent absorption cross-sections of O_3 in the 231-794 nm range. Journal of Quantitative Spectroscopy and Radiative Transfer 61: 509-517. https://doi.org/10.1016/S0022-4073(98)00037-5

Countess RJ, Wolff GT, Whitbeck MR. 1981. The effect of temperature on ozone formation in the propene/nitrogen dioxide/air system. Journal of Environmental Science and Health, Part A: Environmental Science and Engineering 16: 1-8. https://doi.org/10.1080/10934528109374958

Crutzen PJ. 1979. The role of NO and NO_2 in the chemistry of the troposphere and stratosphere. Annual Review of Earth and Planetary Sciences 7: 443-72. https://doi.org/10.1146/annurev.ea.07.050179.002303

Dudhia J. 1989. Numerical study of convection observed during the winter monsoon experiment using a mesoscale two-dimensional model. Journal of the Atmospheric Sciences 46: 3077-3107. https://doi.org/10.1175/1520-0469(1989)046<3077:NSOCOD>2.0.CO;2

Danckaert T, Fayt C, Van M. 2017. QDOAS software user manual. Royal Belgian Institute for Space Aeronomy. Available at: http://uv-vis.aeronomie.be/software/QDOAS/QDOAS_manual.pdf (accessed on December 12, 2017).

Frins E, Bobrowski N, Platt U, Wagner T. 2006. Tomographic multiaxis-differential optical absorption spectroscopy observations of Sun-illuminated targets: A technique providing well-defined absorption paths in the boundary layer. Applied Optics 45: 6227-6240. https://doi.org/10.1364/AO.45.006227

Frins E, Ibrahim O, Casaballe N, Osorio M, Arismendi F, Wagner T, Platt U. 2011. Ground based measurements of SO_2 and NO_2 emissions from the oil refinery "La Teja" in Montevideo city. Journal of Physics: Conference Series 274: 012083. https://doi.org/10.1088/1742-6596/274/1/012083

Gaffney JS, Marley NA. 2003. Atmospheric chemistry and air pollution. Scientific World Journal 3: 497142. https://doi.org/10.1100/tsw.2003.18

Gotoh T. 1993. Relation between heat islands and NO_2 pollution in some Japanese cities. Atmospheric Environment. Part B. Urban Atmosphere 27: 121-128 https://doi.org/10.1016/0957-1272(93)90051-7

Grutter M, Basaldud R, Rivera C, Harig R, Junkerman W, Caetano E, Delgado H. 2008. SO_2 emissions from Popocatépetl volcano: Emission rates and plume imaging using optical remote sensing techniques. Atmospheric Chemistry and Physics 8: 6655-6663. https://doi.org/10.5194/acp-8-6655-2008

Guichard F, Parsons D, Dudhia J, Bresch J. 2003. Evaluating mesoscale model predictions of clouds and radiation with SGP ARM data over a seasonal timescale. Monthly Weather Review 131: 926-944. https://doi.org/10.1175/1520-0493(2003)131<0926:EMMPOC>2.0.CO;2

Habeebullah T, Munir S, Awad A, Morsy E, Seroji A, Mohammed A. 2015. The interaction between air quality

and meteorological factors in an arid Environment of Makkah, Saudi Arabia. International Journal of Environmental Science and Development 6: 576-580. https://doi.org/10.7763/IJESD.2015.V6.660

Henschel S, Tertre A, Atkinson R, Querol X, Pandolfi M, Zeka A, Haluza D, Analitis A, Katsouyanni K, Bouland C, Pascal M, Medina S, Goodman P. 2015. Trends of nitrogen oxides in ambient air in nine European cities between 1999 and 2010. Atmospheric Environment 117: 234-241. https://doi.org/10.1016/j.atmosenv.2015.07.013

Hermans C, Vandaele A, Carleer M. 1999. Absorption cross-sections of atmospheric constituents: NO_2, O_2, and H_2O. Environmental Science and Pollution Research 6: 151-158. https://doi.org/10.1007/BF02987620

Hou S, Tong S, Ge M, An J. 2016. Comparison of atmospheric nitrous acid during severe haze and clean periods in Beijing, China. Atmospheric Environment 124: 199-206. https://doi.org/10.1016/j.atmosenv.2015.06.023

IMPLAN. 2016. Plan Municipal de Desarrollo de San Luis Potosí 2015-2018. Available at: http://www.cefimslp.gob.mx/documentos/PMD/2015-2018/30_PMD_2015-2018_SAN_LUIS_POTOSI.pdf (accessed on October 31, 2017).

INE. 2011. Cuarto almanaque de datos y tendencias de la calidad del aire en 20 ciudades mexicanas (2000-2009). Instituto Nacional de Ecología, Mexico. Available at: https://sinaica.inecc.gob.mx/archivo/informes/Cuarto%20almanque%20de%20datos%20y%20tendencias%20de%20la%20calidad%20de%20aire%20en%2020%20ciudades%20mexicanas.pdf (accessed on April 21, 2017).

INECC. 2014. Elaboración del Inventario Nacional de Emisiones de Fuentes Móviles para México 2013 y proyección 2030 mediante el uso del modelo Motor Vehicle Emission Simulator (MOVES). Instituto Nacional de Ecología y Cambio Climático, Mexico. Available at: https://www.gob.mx/cms/uploads/attachment/file/197011/2014_CGCSA_Inventario_de_emisiones_usando_MOVES_parte_1.pdf (accessed on April 21, 2017).

INECC. 2017. Sistema Nacional de Información de la Calidad del Aire, SINAICA. Instituto Nacional de Ecología y Cambio Climático, Mexico. Available at: http://sinaica.inecc.gob.mx/ (accessed on March 3, 2017).

INEGI. 2014. Censos económicos 2014. San Luis Potosí. Instituto Nacional de Estadística y Geografía, Mexico. Available at: https://www.inegi.org.mx/contenidos/programas/ce/2014/doc/minimonografias/mslp_ce2014.pdf (accessed on April 21, 2017).

INEGI. 2017. Vehículos de motor registrados en circulación. Consulta interactiva de datos. México. Instituto Nacional de Estadística y Geografía, Mexico. Available at: https://www.inegi.org.mx/programas/vehiculosmotor (accessed on April 21, 2017).

Kalnay E, Kanamitsu M, Kistler R, Collins W, Deaven D, Gandin L, Iredell M, Saha S, White G, Woollen J, Zhu Y, Chelliah M, Ebisuzaki W, Higgins W, Janowiak J, Mo KC, Ropelewski C, Wang J, Leetmaa A, Reynolds R, Jenne R, Joseph D. 1996. The NCEP/NCAR 40-year reanalysis project. Bulletin of the American Meteorological Society 77: 437-471. https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2

Lazaridis M. 2011. First principles of meteorology and air pollution. Springer Dordrecht, 362 pp. (Environmental Pollution, 19). https://doi.org/10.1007/978-94-007-0162-5

Kurucz R.L. 1995. The solar spectrum: Atlases and line identifications. In: Laboratory and astronomical high-resolution spectra (Eds.). Astronomical Society of the Pacific Conference Series 81: 17.

Lim J, Hong S, 2005. Effects of bulk ice microphysics on the simulated monsoonal precipitation over east Asia. Journal of Geophysical Research: Atmospheres 110: D24201. https://doi.org/10.1029/2005JD006166

Mapoma H, Tenthani C, Tsakama M, Kosamu I. 2014. Air quality assessment of carbon monoxide, nitrogen dioxide and sulfure dioxide levels in Blantyre, Malawi: A statistical approach to a stationary environmental monitoring station. African Journal of Environmental Science and Technology 86: 330-343. https://doi.org/10.5897/AJEST2014.1696

Melamed M, Basaldud R, Steinbrecher R, Emeis S, Ruiz L, Grutter M. 2009. Detection of pollution transport events southeast of Mexico City using ground-based visible spectroscopy measurements of nitrogen dioxide. Atmospheric Chemistry and Physics 9: 4827-4840. https://doi.org/10.5194/acp-9-4827-2009

Noxon J. 1975. Nitrogen dioxide in the stratosphere and troposphere measured by ground-based absorption spectroscopy. Science 189: 547-549. https://doi.org/10.1126/science.189.4202.547

Pineda-Martínez L, Carbajal N, Campos A, Noyola C, Aragón A. 2012. Numerical research of extreme wind-induced dust transport in a semi-arid human-impacted

region of Mexico. Atmospheric Environment 45: 4652-4660. https://doi.org/10.1016/j.atmosenv.2011.05.056

Pineda-Martínez L, Carbajal N, Campos A, Aragón A. 2014. Dispersion of atmospheric coarse particulate matter in the San Luis Potosí, Mexico, urban area. Atmósfera, 27: 5-19. https://doi.org/10.1016/S0187-6236(14)71097-5

Platt U, Perner D. 1980. Direct measurements of atmospheric CH_2O, HNO_2, O_3, NO_2 and SO_2 by differential optical absorption in the near. Journal of Geophysical Research 85: 7453-7458. https://doi.org/10.1029/JC085iC12p07453

Platt U, Stutz J. 2008. Differential optical absorption spectroscopy (DOAS) principle and applications. Springer Heidelberg, 597 pp. https://doi.org/10.1007/978-3-540-75776-4

Rivera C, Barrera H, Grutter M, Zavala M, Galle B, Bei N, Li G, Molina L. 2013. NO_2 fluxes from Tijuana using a mobile mini-DOAS during Cal-Mex 2010, Atmospheric Environment 70: 532-539. https://doi.org/10.1016/j.atmosenv.2012.12.026.

Rivera J. 2014. Análisis climático de la ciudad de San Luis potosí para la determinación de variables de confort térmico en edificaciones. B.Sc. thesis in Environmentas Engineering. Universidad Autónoma de San Luis Potosí, Mexico.

Rublev A, Chubarova N, Trotsenko A, Gorchakov G. 2003. Determination of NO_2 column amounts from AERONET data. Izvestiya Atmospheric and Ocean Physics 40: 62-77.

SEDEMA. 2017. Base de datos. Red Automática de Monitoreo Atmosférico (RAMA). Gobierno de la Ciudad de México. Available at: http://www.aire.cdmx.gob.mx/default.php?opc=%27aKBh%27 (accessed on March 3, 2017).

SEGAM. 2013. Inventario de emisiones a la atmósfera. Gobierno del Estado de San Luis Potosí, México. Available at: https://slp.gob.mx/segam/Documentos%20compartidos/ESTUDIOS%20PROGRAMAS%20Y%20PROYECTOS/InventarioEstataldeEmisiones_SLP-2011.pdf (accessed on April 21, 2017).

SEMARNAT. 2013. Calidad del aire: una práctica de vida. Secretaría de Medio Ambiente y Recursos Naturales, Mexico. Available at: http://biblioteca.semarnat.gob.mx/janium/Documentos/Ciga/Libros2013/CD001593.pdf (accessed on April 21, 2017).

Sillman S. 1999. The relation between ozone, NO_x and hydrocarbons in urban and polluted rural environments. Atmospheric Environment 33: 1821-1845. https://doi.org/10.1016/S1352-2310(98)00345-8

Skamarock W, Klemp J, Dudhia J, Gill D, Barker D, Wang W, Powers J. 2005. A description of the advanced research WRF version 2. NCAR technical note NCAR/TN-468+STR. 8 pp. https://doi.org/10.5065/D68S4MVH

Skamarock W, Klemp J, Dudhia J, Gill D, Barker D, Duda M, Powers J. 2008. A Description of the Advanced Research WRF version 3 (No. NCAR/TN-475+STR). University Corporation for Atmospheric Research. https://doi.org/10.5065/D68S4MVH

Solomon S, Schmeltekopf A. 1987. On the interpretation of zenith sky absorption measurements. Journal of Geophysical Research 92: 8311-8319. https://doi.org/10.1029/JD092iD07p08311

Spinei E, Cede A, Herman J, Mount G, Eloranta E, Morley B, Baidar S, Dix B, Ortega L, Koenig T, Volkamer R. 2015. Ground-based direct-sun DOAS and airborne MAX-DOAS measurements of the collision-induced oxygen complex, O_2O_2, absorption with significant pressure and temperature differences. Atmospheric Measurement Techniques 8, 793-809 https://doi.org/10.5194/amt-8-793-2015

Tyler N, Bohórquez J, Ramírez C. 2013. Marco teórico de contaminación atmosférica en Colombia. University College London/Universidad de los Andes. Colombia. Available at: https://prosperityfund.uniandes.edu.co/site/wp-content/uploads/Caracterizaci%C3%B3n-de-la-contaminaci%C3%B3n-atmosf%C3%A9rica-en-Colombia1.pdf (accessed on October 31, 2019).

Vandaele A, Hermans C, Simon P, Carleer M, Colin R, Fally S, Mérienne M, Jenouvrier A, Coquart B. 1998. Measurements of the NO_2 absorption cross-section from 42 000 cm^{-1} to 10 000 cm^{-1} (238-1000 nm) at 220 K and 294 K, Journal of Quantitative Spectroscopy and Radiative Transfer 59: 171-184. https://doi.org/10.1016/S0022-4073(97)00168-4

WHO. 2016. WHO Global Urban Ambient Air Pollution Database (update 2016). World Health Organization, Geneva. Available at: http://www.who.int/phe/health_topics/outdoorair/databases/cities/en/ (accessed on February 24, 2017).

Zhang Y, Dubey M, Olsen S, Zheng J, Zhang R. 2009. Comparisons of WRF/Chem simulations in Mexico City with ground-based RAMA measurements during the 2006-MILAGRO. Atmospheric Chemistry and Physics 9: 3777-3798. https://doi.org/10.5194/acp-9-3777-2009

Atmósfera 36(2), 329-341 (2023)
https://doi.org/10.20937/ATM.53056

Patterns related to pollutant concentrations in the Metropolitan Area of Belo Horizonte, Brazil

Luiza Maria Marcos Cerqueira MENDES[1], Vanessa Silveira Barreto CARVALHO[1]*, Fabrina Bolzan MARTINS[1] and Taciana Toledo de Almeida ALBUQUERQUE[2]

[1] *Instituto de Recursos Naturais, Universidade Federal de Itajubá, Itajubá, 37500-903, Minas Gerais, Brazil.*
[2] *Departamento de Engenharia e Sanitária Ambiental, Universidade Federal de Minas Gerais, Belo Horizonte, 31270-010, Minas Gerais, Brazil.*
*Corresponding author; email: vanessa.silveira@unifei.edu.br

Received: March 23, 2021; accepted: August 20, 2021

RESUMEN

La contaminación del aire por actividades humanas e industriales ha sido una gran preocupación en los últimos años. Entre los diversos contaminantes que se encuentran en la atmósfera, el material particulado (PM) y el ozono (O_3) muestran ocurrencias significativas, con altas concentraciones en los centros urbanos frecuentemente asociadas con problemas ambientales y de salud pública. Por lo tanto, este trabajo utiliza la técnica de análisis de varianza (ANOVA) y la prueba de Tukey para investigar los patrones relacionados con la variabilidad de las concentraciones máximas diarias de O_3 y las concentraciones medias diarias de MP con diámetro menor a 10 μm (PM_{10}), registradas entre 2007 y 2012 en seis sitios de la región metropolitana de Belo Horizonte, Brasil. Para ello, los datos se analizaron mediante ANOVA dispuestos en un esquema factorial (6 × 4 × 2) con cuatro repeticiones por tratamiento, seguido de la prueba de Tukey. En el ANOVA y la prueba de Tukey, el primer factor (A) representa las seis estaciones de monitoreo de la calidad del aire, el segundo (B) representa las estaciones y el tercero (C) las mediciones realizadas durante los días hábiles y fines de semana. Los patrones de variabilidad estacional muestran concentraciones más altas de O_3 en la primavera y de PM_{10} en el invierno. Los valores fueron 22.9 y 35.32% superiores a los promedios anuales de concentraciones de O_3 y PM_{10}, respectivamente. Los valores medios para los días laborables y los fines de semana mostraron patrones diferentes para los dos contaminantes. Las concentraciones de PM_{10} fueron 11% más altas durante los días laborables en comparación con los fines de semana. El efecto de fin de semana de O_3 se encontró sólo en una de las estaciones. Los perfiles de emisiones vehiculares e industriales se han identificado como un factor potencial que conduce a estos resultados.

ABSTRACT

Air pollution from human and industrial activities has been a major concern in recent years. Among the various pollutants found in the atmosphere, particulate matter (PM) and ozone (O_3) show significant occurrences, with high concentrations in urban centers frequently associated with environmental and public health problems. Therefore, this study uses the analysis of variance (ANOVA) technique and Tukey's test to investigate patterns related to the variability of maximum daily O_3 concentrations and mean daily concentrations of PM with a diameter less than 10 μm (PM_{10}), registered between 2007 and 2012 through six sites in the Metropolitan Area of Belo Horizonte, Brazil. To this end, the data were analyzed using ANOVA arranged in a factorial scheme (6 × 4 × 2) with four repetitions per treatment, followed by Tukey's test. In the ANOVA and Tukey's test, the first factor (A) represents the six air quality monitoring stations, the second (B) represents the seasons, and the third (C), the measurements carried out during working days and weekends. Seasonal variability patterns show higher concentrations of O_3 in spring and of PM_{10} in winter. The values were 22.9 and 35.32% higher than the annual averages of O_3 and PM_{10} concentrations, respectively. The mean values for working days and weekends

showed different patterns for the two pollutants. PM_{10} concentrations were 11% higher during working days when compared to weekends. The O_3 weekend effect was found only in one of the stations. The profiles of vehicular and industrial emissions have been identified as a potential factor leading to these results.

Keywords: ANOVA, Ozone, Particulate Matter.

1. Introduction

The Metropolitan Area of Belo Horizonte (MABH), located to the mid-west of Minas Gerais state, covers 33 municipalities (Gouveia et al., 2019) with a population of 5961 815 inhabitants, which represents about 40% of the population of the entire state (IBGE, 2018), and has a vehicle fleet of nearly three million (DENATRAN, 2018). The MABH is in a region with enormous mining potential, including the Iron Quadrangle. This characteristic, combined with the population cluster, the availability of skilled labor, and the proximity to a consumer market, developed numerous industrial activities with high polluting potential (Radicchi, 2012), negatively impacting air quality (Miranda et al., 2011; FEAM, 2016; Gouveia et al., 2019; Peláez et al., 2020). In addition to high emission rates, air quality in the MABH is impacted by topography, which affects the distribution of rain, temperature variation, and wind speed and direction (Prudente et al., 2006; Santos et al., 2019).

The MABH registered an increase in concentrations of pollutants over the last two decades (Miranda et al., 2011; Pacheco et al., 2017; Peláez et al., 2020), and according to the Fundação Estadual do Meio Ambiente (State Environmental Foundation, FEAM) (FEAM, 2016), the main problem facing the MABH refers to high concentrations of particulate matter with a diameter less than 10 μm (PM_{10}) and ozone (O_3). For example, Pacheco et al. (2017) showed an increase in the average annual concentration of PM_{10} from 21.5 μg m^{-3} in 2005 to 31.3 μg m^{-3} in 2010. Sicard (2021) showed worldwide O_3 levels decreased in rural areas and increased in urban areas with an average increase for Brazil, between 2005 and 2014, of 0.56 ppb per year (approximately 1.1 μg m^{-3}). Paláez et al. (2020) identified frequent O_3 (8 h average concentrations) exceedances of the Brazilian National Air Quality Standards (NAQS) established by Resolution No. 491 (CONAMA, 2018) as safe for human health. The current NAQS in Brazil adopts the model of successive targets recommended by the WHO (2006) for reducing air pollution and ensuring the well-being and health of the population. Brazil is currently in the first stage; however, there is no timetable established for the implementation of the next targets nor indication of how they will be achieved.

Air quality depends directly on the emission of pollutants and is subjected to weather variability, which plays an important role in the transport, transformation, and dispersion of pollutants into the atmosphere. Several studies have been carried out worldwide, including Brazil, to relate ambient pollutant concentrations and their variability to local weather conditions. As an example, Sánchez-Ccoyllo and Andrade (2002) studied the influence of weather conditions on the behavior of pollutants in the Metropolitan Area of São Paulo (MASP) and found that the high relative humidity and the occurrence of cold fronts favors the dispersion and dilution of pollutants in the area. On the other hand, strong thermal-inversion conditions in the low atmosphere were associated with high pollutant concentrations at the surface. Similar results were found by Carvalho et al. (2015) and Santos et al. (2016), showing higher concentrations of most pollutants in the MASP during winter, when low rainfall rates, low winds, the predominance of the South Atlantic Subtropical High (SASH), and thermal inversions in the lower atmosphere inhibit dispersal of pollutants. However, higher O_3 concentrations were registered during spring and summer, when clear sky conditions, high temperatures, low relative humidity, and no precipitation dominate (Carvalho et al., 2012, 2015; Santos et al., 2018).

It is important to point out that several statistical techniques have been used to analyze the patterns associated with the concentration of pollutants, as well as to verify the seasonal and weekday-weekend influence (Mansouri et al. 2011; Aenab et al., 2013; Mohamad et al., 2015; Govender and Sivakumar, 2020; Carvalho et al., 2020). Among those techniques, the Analysis of variance (ANOVA) allows investigating

the influence of preponderant factors on pollutant concentrations and whether there is a relationship between them (Storck et al., 2011; Aenab et al., 2013; Mohamad et al., 2015; Carvalho et al., 2020; Tyubee et al., 2020). Within the preponderant factors, the following can be listed: (*i*) seasonal effects (Carvalho et al., 2015; Santos et al., 2016; Tyubee et al., 2020); (*ii*) effect of the day of the week, related to weekdays vs. weekends (Brönnimann and Neu, 1997; Carvalho et al., 2020); and (*iii*) monitoring sites (Carvalho et al., 2020). In other words, ANOVA allows grouping these factors (*i*, *ii*, and *iii*) with similar patterns of pollutant concentrations, in addition to assessing the individual influence of each factor on the observed patterns. Nevertheless, despite the potential and possibilities of ANOVA, studies applying this technique to air pollution data are still scarce. Some examples of studies that use the ANOVA technique are: (*i*) Aenab et al. (2013) evaluated Baghdad's air quality by analysis of heavy metals; (*ii*) Fang et al. (2017) tested mean concentration variance differences for metallic elements (PM, Hg(p), Mn, Fe, Zn, Cr, Cu, and Pb) through the seasons at Taiwan; (*iii*) Tyubee et al. (2020) assessed the influence of factors such as land use/land cover and climate have on the spatial and seasonal variation in pollutants concentrations in Gboko, Nigeria; and (*iv*) for Brazil, the most recent work was performed by Carvalho et al. (2020) to evaluate O_3 concentration behavior and variability throughout the State of São Paulo.

Although some studies have been carried out for the MABH (Miranda et al., 2011; Pacheco et al., 2017; Peláez et al., 2020), none looked in-depth at the patterns associated with PM_{10} and O_3 concentrations, nor applied the ANOVA technique. We should add that this type of analysis provides effective information for developing public policies to control air pollution. Therefore, this study aims to verify, using the ANOVA technique, the patterns associated with PM_{10} and O_3 concentrations considering the following effects: seasonal variability, weekday-weekend concentration patterns, and similarities and differences of MABH air quality monitoring sites. Since O_3 is a secondary pollutant formed from photochemical reactions involving nitrogen oxides (NO_x) and volatile organic compounds (VOCs), NO_x patterns were also analyzed to complement the analyses. VOCs data were not available for the region.

2. Methodology

2.1 Data source

To carry out this study, 1-h mean concentrations of O_3 and PM_{10} monitored by FEAM between 2007 and 2012 were used. Overall, six automatic monitoring stations, whose location, mean (Mean), and maximum (Max) values within the analyzed period are listed in Table I. The spatial distribution can be observed in Figure 1.

From the 1-h mean data, the daily maximum O_3 concentrations and the daily mean concentration of PM_{10} were obtained for each site. At least 75% of valid data was required to calculate this data. Then, the data were classified according to the seasons

Table I. Location, parameters, and mean and maximum values of the pollutants (PM_{10} and O_3) in $\mu g\ m^{-3}$, monitored by the network of air quality monitoring stations in MABH.

Municipality	Monitoring stations	Monitored parameters			
		PM_{10}		O_3	
		Mean	Maximum	Mean	Maximum
Belo Horizonte	Praça Rui Barbosa (RB)	32.4	613.0	29.7	228.8
	Bairro Jardim das Alterosas (JA)	40.3	659.5	29.0	239.4
Betim	Bairro Petrovale (BP)	33.9	627.0	38.4	284.1
	Centro Administrativo (Safran)	50.9	688.2	31.0	313.1
Ibirité	Bairro Cascata (BC)	23.2	637.0	38.5	314.4
	Bairro Piratininga (Ibiritermo)	29.0	354.0	43.68	257.0

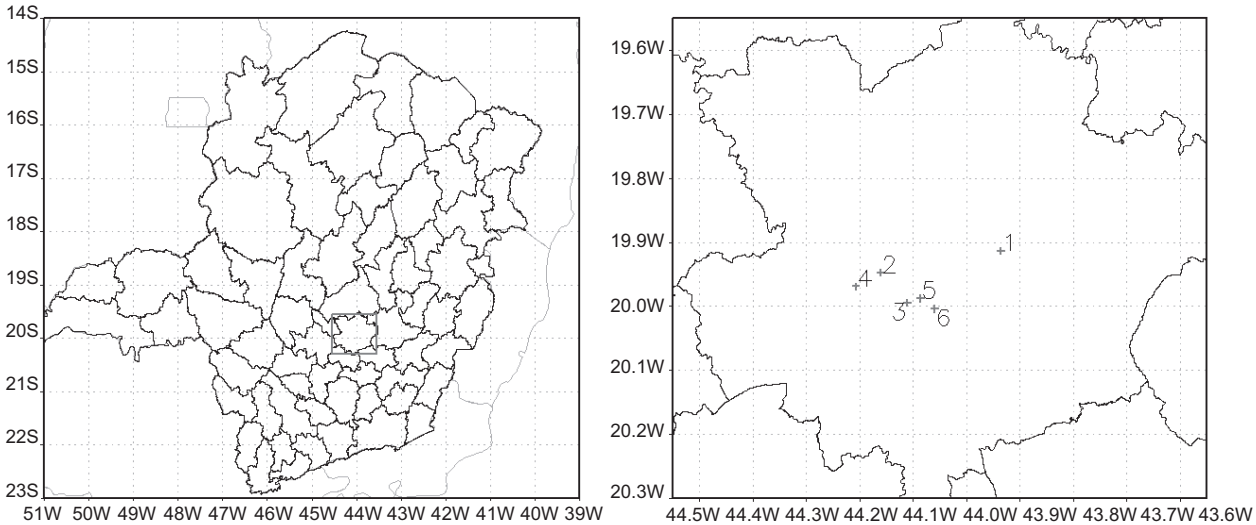

Fig. 1. State of Minas Gerais, Brazil (left) and spatial distribution of the air quality monitoring stations through Belo Horizonte Metropolitan Area (right), where: (1) Rui Barbosa, (2) Jardim Alterosa, (3) Bairro Petrovale, (4) Safran, (5) Cascata, and (6) Ibiritermo.

and day of the week corresponding to each of these registers. As a complement to the analysis related to the patterns of O_3 observed in MABH, the 1-h mean data for nitrogen dioxide (NO_2), nitrogen monoxide (NO), and the ratio between the concentrations of NO_2/NO were also used. The NO_2/NO ratio was used by Carvalho (2010) as an indicator of O_3 potential production in a region once its troposphere formation depends on the NO_2 photodissociation and that the NO acts on the consumption of the O_3 molecules. So, higher ratios are expected to be associated with higher concentration values of O_3.

2.2 Analysis of variance

Before proceeding to the ANOVA, the data of daily maximum and mean concentrations of the pollutants (O_3 and PM_{10}) were submitted to verify the assumption of normality and homogeneity of variance with the Shapiro-Wilk test ($\alpha = 0.05$). This procedure was performed to verify whether the data follow the normal distribution and have homogeneous variance ($p \geq 0.05$) or if the data do not follow the normal distribution and have heterogeneous variance ($p < 0.05$) (Mohamad et al., 2015). After this procedure, the data were analyzed using the three-factor ANOVA, organized in a factorial scheme ($4 \times 2 \times 6$) with four replicates per treatment. This procedure was adopted to evaluate the effect of the sources of variation (seasonal, day of the week, and monitoring site) at a 5% probability (Mohamad et al., 2015; Carvalho et al., 2020; Tyubee et al., 2020). The null hypothesis for ANOVA is that the factors are independent of each other, i.e., there is no interaction between them ($p \geq 0.05$) against the alternative hypothesis that the factors are dependent, meaning that there is an interaction between the sources of variation ($p < 0.05$). In this study, factor 1 refers to the seasons (four levels of factor 1: summer, autumn, winter, and spring); factor 2 refers to the weekday effect (2 levels of factor 2: weekday = data grouped from Monday to Friday; and weekend = data grouped from Saturday and Sunday); and factor 3 refers to the monitoring stations (six levels of factor 3: each of the six monitoring sites: Rui Barbosa (RB), Jardim das Alterosas (JA), Bairro Petrovale (BP), Safran, Bairro Cascata (BC), and Ibiritermo. Each repetition was considered the mean value of the maximum concentrations of O_3, and the mean value of the average concentrations of PM_{10}, totaling 576 data for each pollutant. Each treatment was the combination of the data among the three factors. Finally, a comparison of means was performed for the maximum concentration values of O_3 and the average concentration of PM_{10} using Tukey's test ($\alpha = 0.05$). All the tests were performed using the

Sisvar software (Ferreira, 2011) and followed the methodology analogous to Carvalho et al. (2020).

To complement the analysis of O_3 patterns, the same procedure (Shapiro-Wilk, ANOVA, and Tukey's test) was performed with the daily mean concentrations of NO_x and NO_2/NO ratio.

3. Results and discussion

3.1 Patterns associated with ozone concentration and its precursors

Using the Shapiro-Wilk test ($\alpha = 0.05$), the daily maximum O_3 concentrations, it was found that the daily mean NO and NO_2 concentrations and the daily mean NO_2/NO ratio violated the assumption of normality and homogeneity of variance. Therefore, they were transformed by (Ln(x)), following the recommendations of Storck et al. (2011) and Carvalho et al. (2020).

Regarding the ANOVA, daily maximum O_3 concentrations showed a significant interaction between the sources of variation: (*i*) seasons × monitoring stations, and (*ii*) day of the week × monitoring stations. Also, the factors monitoring stations (sites) and seasons were significant ($p < 0.05$) (Table II). This behavior means there is a differentiated pattern of maximum O_3 concentrations between monitoring stations, seasons, and days of the week. In other words, depending on the site, there are higher concentrations of O_3 in a given season, and there will be variations between weekdays and weekends. This result agrees with those presented by Carvalho et al. (2020), in which the ANOVA technique was applied to O_3 concentration data in the state of São Paulo. Also, the factors that contributed most to explain the pattern associated with O_3 concentrations were (in order of magnitude): season (F test = 169.5), monitoring station (F test = 81.8), the interaction between the levels of factors: season and monitoring station (test F = 7.1), and the interaction between the levels of factors: monitoring station and day of the week (F test = 3.1).

Once the interactions between (*i*) seasons × sites, and (*ii*) day of the week × sites were verified, it was necessary to analyze which monitoring stations differ and if/how the seasonal and weekend effects occur (Tables III and IV). For this purpose, the combined effects of each level of both factors were unfolded through Tukey's test. It is important to highlight that the coefficient of variation of 3.56% (less than 10%) indicates high precision in the analysis (Storck et al., 2011). Considering the results of the Tukey test

Table II. Analysis of variance of maximum daily O_3 concentrations ($\mu g\ m^{-3}$).

SV	DF	MS	F-test	P-value
Seasons*Week*Sites	15	0.0036	0.157	0.9999
Seasons*Week	3	0.0452	1.975	0.1167
Seasons*Sites	15	0.1623	7,086	0.0000*
Week*Sites	5	0.0698	3.046	0.0101*
Seasons	3	3.8820	169.463	0.0000*
Week	1	0.0872	3.805	0.0516
Sites	5	0.0229	81.772	0.0000*
Error	528	0.0229		
Total	575			
CV (%) =	3.56			
Overall mean	4.2551	Number of data:	576	

*Interaction between variation factors; **significant values by the F-test ($\alpha = 0.05$).

SV: sources of variation; Seasons: seasons of the year; Week: day of the week; Sites: monitoring stations; DF: degrees of freedom; MS: mean square; Fc: test F, given by the ratios of MS to the MS of the error; CV: coefficient of variation, given by equation ($cv = \pm 100 \times \sigma / \text{mean}$); P: probability value of the F-test.

Table III. Comparison of means for the average maximum O_3 concentrations (μg m^{-3}) considering the interaction between the Monitoring Stations (6 levels) and Seasons (4 levels).

Monitoring stations	Season			
	Summer	Autumn	Winter	Spring
Alterosa	63.35 Bb	49.79 ABa	65.26 ABb	85.24 ABc
Cascata	69.95 Ba	71.13 Ca	76.32 Ca	93.58 BCb
Ibiritermo	69.94 Ba	64.03 Ca	70.76 BCa	92.06 BCb
Petrovale	85.60 Ca	81.76 Da	77.72 Ca	97.16 Cb
Rui Barbosa	48.75 Aa	46.53 Aa	62.88 Ab	76.20 Ac
Safran	67.18 Bb	55.98 Ba	81.09 Cc	91.60 BCd

Note: Different uppercase letters in the column (monitoring stations) and different lowercase letters in the row (seasons) differ by the Tukey test at 5% probability. Letters A to C and a to c were set in ascending order of O_3 concentrations.

Table IV. Comparison of means for the average maximum O_3 concentrations (μg m^{-3}) considering the interaction between day of the week (two levels) and monitoring stations (six levels).

Monitoring stations	Day of the week			
	Monday-Friday		Saturday-Sunday	
Alterosa	64.69	Ba	67.13	Aa
Cascata	78.28	Da	77.21	Ba
Ibiritermo	75.00	CDa	73.40	Ba
Petrovale	85.86	Ea	85.26	Ca
Rui Barbosa	54.98	Aa	62.20	Ab
Safran	72.12	Ca	75.81	Ba

Note: Different uppercase letters in column (monitoring stations) and different lowercase letters in the row (seasons) differ by the Tukey test at 5% probability. Letters A to E and a or b were set in ascending order of O_3 concentrations.

($\alpha = 0.05$) for the interaction between the sources of variation seasons and sites (Table III), it is verified that the highest mean of maximum O_3 concentration was registered during spring (mean: 89.30 μg m^{-3}), significantly differing for the other seasons in all monitoring stations, followed by winter (72.34 μg m^{-3}), summer (67.46 μg m^{-3}), and autumn (61.54 μg m^{-3}). This pattern of higher concentrations during spring was also verified by Carvalho et al. (2015, 2020) for the state of São Paulo. In general, during spring, higher ozone concentration values are observed, mainly because of the combination of increased incidence of solar radiation and lower cloud cover (compared to summer), particularly in the afternoon (Fig. 2). This combination favors photochemical reactions responsible for forming the O_3 molecule (Carvalho et al., 2015). Furthermore, the highest mean values of the maximum daily O_3 concentrations were registered at Petrovale (97.16 μg m^{-3}) and Cascata (93.58 μg m^{-3}), both located near the Gabriel Passos refinery, a region influenced by industrial and vehicular sources (FEAM, 2016). It is also observed that, except for the Cascata and Petrovale stations, the lowest mean values relative to maximum concentrations of O_3 were recorded during autumn, a similar pattern observed by Alvim (2013) for the main O_3 precursors in the Metropolitan Area of São Paulo. Alvim (2013) showed that more compounds contribute to the formation of O_3 in the winter than in autumn, mainly due to higher VOC concentrations in the atmosphere. Although the ratio between VOC and NO_x concentrations for the MABH is unknown, by applying ANOVA for NO (Table V), it was observed that mean concentrations for all monitoring stations were higher in winter (17.46 μg m^{-3}), followed by autumn (12.55 μg m^{-3}), summer (7.63 μg m^{-3}), and spring (7.45 μg m^{-3}). Higher concentrations of NO during winter would result in lower concentrations of O_3 in that season. However, only for JA, RB, and Safran there was a pattern opposite to the expected, indicating higher concentrations of O_3 in the winter when compared to

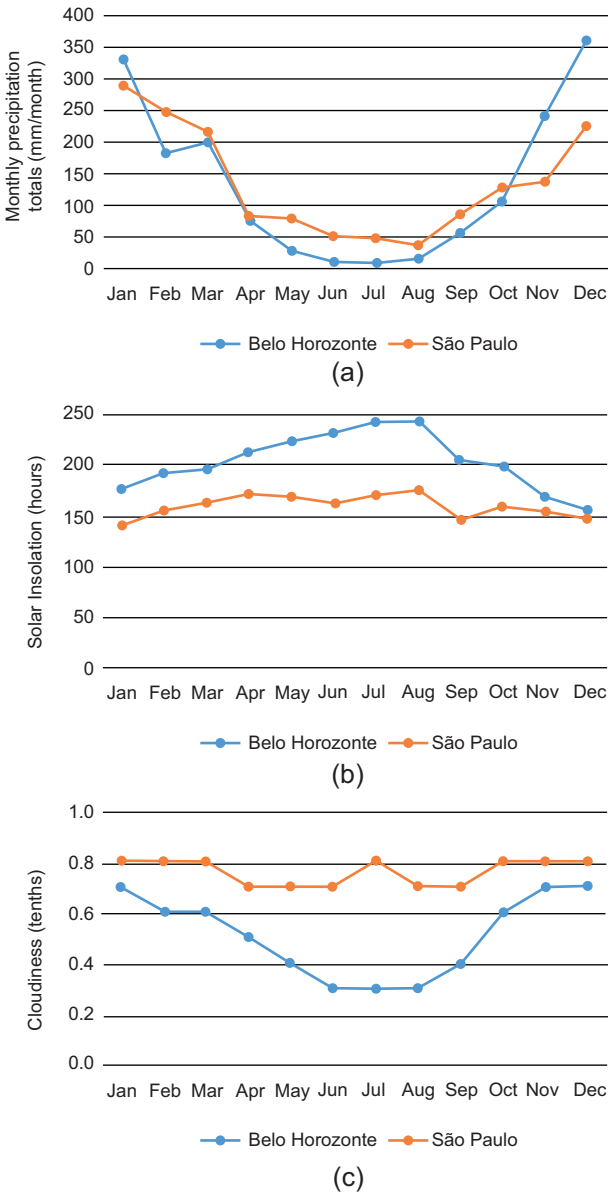

Fig. 2. Climatological normals (1981-2010) referring to the stations Belo Horizonte and São Paulo (Mirante de Santana) for (a) monthly precipitation totals (mm), (b) solar insolation (hours), and (c) cloudiness (tenths). Source: https://portal.inmet.gov.br/normais

autumn. The profile with higher mean concentrations in winter when compared to summer differs from that observed by Carvalho et al. (2015) for the MASP, probably due to the different climatic conditions registered between the two regions. As shown in Figure 2, during winter the climatological normals

Table V. Comparison of means for average NO concentrations (μg m^{-3}), NO_2/NO ratio, and O_3 concentrations (μg m^{-3}) between monitoring stations.

Monitoring stations	NO	NO_2/NO	O_3
Alterosa	—	—	65.91 B
Cascata	5.74 B	4.71 AB	77.75 C
Ibiritermo	4.82 A	6.49 D	74.20 C
Petrovale	8.21 C	5.31 BC	85.56 D
Rui Barbosa	16.48 D	4.21 A	58.59 A
Safran	20.79 D	6.54 C	73.96 C

Different uppercase letters in column (monitoring stations) and different lowercase letters in the row (seasons) differ by the Tukey test at 5% probability. Letters A to D were set in ascending order of NO, NO_2/NO ratio, and O_3 concentrations.

(1981-2010) for the Belo Horizonte station show lower precipitation rates, higher solar insolation values, and lower cloudiness values, factors that contribute to the formation of O_3.

It is noteworthy that the BC and BP stations showed slightly lower values in summer and winter, respectively, but they did not differ from those found in summer. The lowest means were, in general, registered in the RB (46.53 μg m^{-3}) and JA (49.79 μg m^{-3}) monitoring stations. This fact may be associated with vehicular traffic, since both stations are in regions significantly influenced by vehicle emissions.

Considering the interaction between the factors site and day of the week (Table IV), it is observed that, except for the RB station, the other monitoring stations did not present significant differences between the concentrations on weekdays and weekends, confirming, therefore, the weekend effect only at the Rui Barbosa station. Except for RB, the others did not follow the pattern observed by Silva Junior et al. (2009) and Carvalho et al. (2015, 2020) for MASP, and Carvalho (2006) for the Metropolitan Area of Rio de Janeiro, where they found higher concentrations of O_3 during the weekend. As noted for RB, higher concentrations of O_3 during weekends may be related to the increase of the ratio between NO_2 and NO concentrations (NO_2/NO), which occurs due to a decrease in vehicle traffic during the weekend. This decreases the emission of NO, a compound that acts

in the consumption of the O_3 molecule ($O_3 + NO \rightarrow NO_2 + O_2$. The weekend effect was also observed in studies conducted in France (Pont and Fontan, 2001), Turkey (Im et al., 2013), United States (Wolff et al., 2013), and Canada (Huryn and Gough, 2014). Sicard et al. (2020) showed a significant downward trend (–0.26 per year) for the weekend effect in Canada, United States, Japan, Germany and Italy and an upward trend (+1.15 ppb per year [approximately +2.2 $\mu g\ m^{-3}$]) in urban stations in France, South Korea, and the United Kingdom. It is important to point out that, between the six monitoring stations, only Rui Barbosa is located within the Greater Belo Horizonte, being the only site predominantly influenced by vehicle emissions. Sites predominantly influenced by vehicular emissions in the state of São Paulo also presented a significant weekend effect (Carvalho et al., 2020).

The highest concentration values of O_3 were observed in Cascata and Petrovale, located in the MABH industrial area. This result can be associated with lower NO levels, combined with lower vehicle traffic on these two stations (Table V). In all stations, mean NO concentrations showed higher values during weekdays (13.29 $\mu g\ m^{-3}$) compared with weekends (9.22 $\mu g\ m^{-3}$). Considering all the MABH, this represents an average reduction of 30% of weekend concentrations compared to working days. Additionally, considering only the RB station, an average reduction of 28.4% of NO was observed on weekends. The results of this study are similar to those found by Carvalho et al. (2015) for MASP, and Qin et al. (2004) for California (USA), with an average of NO_x reductions of 32 and 37% over the weekend, respectively.

Considering that NO acts on the consumption of the O_3 molecule, it is expected that in areas where NO levels are higher, O_3 levels will be lower, as in Rui Barbosa. This hypothesis is supported by the mean concentrations of NO found in all sites (Table V). Analyzing the average behavior of the ratio NO_2/NO, when compared to the average daily maximum O_3 concentrations, it can be noted that, in general, higher values of the ratio between nitrogen oxides (NO_2/NO) were followed by higher concentration values of O_3, matching the expected pattern, since the increase in O_3 concentrations is directly associated with the increase of the NO_2/NO ratio.

Additionally, the monitoring stations where no differences were found between O_3 concentrations during weekdays and weekends (Alterosa, Cascata, Ibirifero, Petrovale, and Safran) are close to industrial units, which supports the hypothesis that the vehicle emissions profile of Rui Barbosa station plays an important role in the occurrence of the weekend effect (Table IV).

When analyzing mean O_3 concentrations at the monitoring stations (Table V), it is observed that the highest concentration occurs at Petrovale, regardless of the factors season and day of the week, unlike the rest of the stations. Comparatively, the Rui Barbosa station presents the lowest O_3 concentrations (Tables IV and V). This result corroborates the results obtained by Mendes (2018), who identified that Petrovale station had the highest number of exceedances of 100 $\mu g\ m^{-3}$ (suggested by the WHO as a safe value for the population) in an 8-h average. The study also found that although Rui Barbosa also exceeded this threshold in all years (except for 2009), it presented the lowest number of exceedances when compared with the other sites.

3.2 Patterns related to the concentration of PM_{10}

A violation of the assumption of normality and homogeneity of variance for the data of daily mean concentrations of PM_{10} was found by the Shapiro-Wilk test ($p < 0.05$). Therefore, the data were transformed into Ln(x) according to the recommendation by Storck et al. (2011) and Carvalho et al. (2020).

According to ANOVA, the factors that most contributed to explaining the pattern associated with PM_{10} concentrations were (in order of magnitude): season (F test = 169.5), monitoring stations (F test = 157.3), day of the week (F test = 43.5), and interaction between the levels of factors season and monitoring station (F test = 3.7) (Table VI).

Since there was a significant interaction between the sources of variation seasons × monitoring stations ($p < 0.05$) (Table VI), it was necessary to unfold the combined effects of each level of both factors (Table VII). Besides, it was necessary to verify the individual behavior of the sources of variation (*i*) sites, (*ii*) seasons, and (*iii*) days of the week.

In general, the highest values of PM_{10} daily mean concentrations were registered in winter in all sites, with an average of 45.78 $\mu g\ m^{-3}$, followed by spring

Table VI. Analysis of variance of daily average PM_{10} concentrations (μg m^{-3}).

SV	DF	MS	F-test	P-value
Seasons*Week*Site	15	0.0139	0.205	0.9995
Seasons*Week	3	0.1044	1.538	0.2036
Seasons*Site	15	0.2538	3.739	0.0000**
Week*Site	5	0.0480	0.707	0.6183
Seasons	3	11.5036	169.481	0.0000**
Week	1	2.9518	43.488	0.0000**
Site	5	10.6806	157.355	0.0000**
Error	528	0.0679		
Total	575			
CV (%)	7.65			
Overall mean	3.4059	Number of data:	576	

*Interaction between variation factors; **significant values by the F-test ($\alpha = 0.05$). SV: sources of variation; Seasons: seasons of the year; Week: day of the week; Sites: monitoring stations; DF: degrees of freedom; MS: mean square; F-test: F-test value, given by the MS ratios; CV: coefficient of variation, given by equation CV= $\pm 100 \times \sigma$/mean; P: probability value of the F-test.

Table VII. Comparison of means for average PM_{10} concentrations (μg m^{-3}) considering the interaction between the monitoring stations (six levels) and seasons (four levels).

Monitoring stations	Seasons			
	Summer	Autumn	Winter	Spring
Alterosa	28.67 Da	44.99 Cc	62.41 Cd	37.15 Bb
Cascata	16.16 Ba	20.94 Ab	29.08 Ac	26.78 Ac
Ibiritermo	13.28 Aa	19.53 Ab	28.20 Ac	22.29 Ab
Petrovale	21.51 Ca	30.47 Bb	43.22 Bc	35.25 Bb
Rui Barbosa	25.91 CDa	30.22 Bab	37.40 Bc	35.05 Bbc
Safran	29.86 Da	48.18 Cb	74.35 Cc	51.00 Cb

Different uppercase letters in the column (monitoring stations) and different lowercase letters in the row (seasons) differ from each other by the Tukey test at 5% probability. Letters A to D and a to d were set in ascending order of particulate matter concentrations.

(34.59 μg m^{-3}), autumn (32.39 μg m^{-3}), and summer (22.56 μg m^{-3}). Similar results were found in several regions in Brazil, such as São Paulo and Rio de Janeiro (Castanho and Artaxo, 2001; Miranda et al., 2002, 2011; Carvalho et al., 2015; Santos et al., 2016).

Additionally, it was possible to verify, using Tukey's test, that the highest mean PM_{10} concentrations recorded during winter (Table VII) deffered from the other seasons in all monitoring stations, except for Cascata and Rui Barbosa, with no significant differences between winter and spring. Unfavorable atmospheric conditions of dispersion justify the pattern of occurrence of higher PM_{10} concentrations during winter and early spring, when the SASH position is closest to the continent (Santos et al., 2016), reducing cloudiness and precipitation in the MABH. Besides, the low wind speed also impairs the dispersion of pollutants during winter.

The highest mean value of PM_{10} concentrations was observed for Safran (50.85 μg m^{-3}), followed by

Alterosa (43.30 μg m^{-3}), Petrovale (32.61 μg m^{-3}), Rui Barbosa (32.14 μg m^{-3}), Cascata (23.24 μg m^{-3}), and Ibiritermo (20.82 μg m^{-3}) (Table VIII). These results corroborate those reported by the atmospheric emissions inventory for RMBH in 2003 (FEAM, 2003) since the highest concentrations of PM_{10} are observed in the vicinity of the Alterosa and Safran stations.

Table VIII. Comparison of means for average PM_{10} concentrations (μg m^{-3}) considering as source of variation the Monitoring Stations

Monitoring Stations	Concentrations
Alterosa	43.30 D
Cascata	23.24 B
Ibiritermo	20.82 A
Petrovale	32.61 C
Rui Barbosa	32.14 C
Safran	50.85 E

Different uppercase letters in the column (monitoring stations) differ by the Tukey test at 5% probability. Letters A to E were set in ascending order of particulate matter concentrations.

The sites Safran, Alterosa, and Rui Barbosa are influenced by the heavy vehicles traffic, which increases PM_{10} concentrations due to emissions and to road dust resuspension, which is an important PM_{10} source (Pant and Harrison, 2013). In addition to the local impact of vehicular emissions on air quality, atmospheric pollutants can also be transported to other regions, impacting areas without heavy traffic (Santos et al., 2016). Tavares et al. (2010) attributed the poor air quality in the central area of the MABH to traffic of diesel-powered vehicles (mainly buses) as the main PM emission source. The intense industrial activity related to the steel sector, namely, non-metallic minerals (cement and lime), oil, and the automobile industry in the vicinity of the MABH, was also identified as one of the main sources affecting PM_{10} concentrations (Brum, 2010). Through elemental composition analysis of PM_{10}, Moura (2016) also noted that mining activity at various locations of the MABH has a negative impact on the region's air quality.

Significant differences on PM_{10} concentrations through weekdays and weekends were also shown. The highest concentration of PM_{10} was observed during weekdays (35.77 μg m^{-3}) and the lowest in weekends (31.89 μg m^{-3}), with an average percentage reduction of 11% at the latter. Lower concentrations of PM_{10} at weekends were also found in the MASP (Carvalho et al., 2015) and Southern California (Qin et al., 2004), where average reductions of 15 and 14% were observed, respectively, regarding weekdays. In general, during weekdays (Monday to Friday), there is a greater flow of vehicles in large urban centers, which explains the higher concentrations of PM_{10}. According to the official local emissions inventory, this result is expected, since the largest source of PM_{10} in Belo Horizonte are vehicles (71%).

It is important to point out that the absence of valid data at several points in the data series and limited access to data undermine the available information regarding air quality in the MABH. Thus, we identified the importance of future improvements in the current air quality monitoring network of the MABH, including VOCs sampling, which is vital for the O_3 formation process. Also, it is necessary to perform regular maintenance of the existing monitoring network, besides its expansion, since the current stations are mainly located in traffic corridors and industrial areas, covering a small and unrepresentative number of municipalities of the MABH. Resizing the monitoring network, including the installation of stations in areas exposed to different types of emission, would also be important for a better diagnosis of air quality in the state of Minas Gerais.

4. Conclusions

The application of the ANOVA technique to data of daily maximum concentrations of O_3 and daily mean concentrations of PM_{10} registered between 2007 and 2012 in six monitoring stations of the MABH, was efficient to identify patterns associated to the variability of both pollutants' concentrations. It was possible to verify significant differences in the behavior pattern of the pollutants between the considered sources of variation (seasons, day of week, and monitoring stations). The O_3 concentration pattern was higher in the spring in all monitoring stations, while the PM_{10} concentration was higher during winter. There was a

higher concentration of PM_{10} during weekdays when compared to weekends. Regarding O_3, the day of the week effect was identified only in the Rui Barbosa station, which is predominantly influenced by vehicle emissions. The other stations are located in industrial plants and showed no significant differences between the concentrations observed on working days and weekends. This shows that the vehicle emissions profile, which decreases during weekends, plays an important role in the weekend effect pattern associated O_3 concentrations in the MABH.

Acknowledgments

To the Coordination of Improvement of Higher Education Personnel – CAPES (process number 1577748) for financial support and FEAM for the data, which allowed this study to be performed.

References

Aenab AM, Singh SK, Lafta AJ. 2013. Critical assessment of air pollution by ANOVA test and human health effects. Atmospheric Environment 71: 84-91. https://doi.org/10.1016/j.atmosenv.2013.01.039

Alvim DS. 2013. Estudo dos principais precursores de ozônio na Região Metropolitana de São Paulo. Ph.D. thesis. Instituto de Pesquisas Energéticas e Nucleares, Universidade de São Paulo.

Brönnimann S, Neu U. 1997. Weekend-weekday differences of near surface ozone concentrations in Switzerland for different meteorological conditions. Atmospheric Environment 31: 1127-1135. https://doi.org/10.1016/S1352-2310(96)00311-1

Brum DR. 2010. Estudo da composição química do material particulado fino ($MP_{2,5}$) em Porto Alegre e Belo Horizonte. M.Sc. thesis. Universidade de São Paulo, Brazil.

Carvalho VSB. 2006. Metodologia da qualidade do ar no que tange as concentrações de Ozônio e dos Óxidos de Nitrogênio na Região Metropolitana do Rio de Janeiro**.** M.Sc. thesis. Universidade Federal do Rio de Janeiro, Brazil.

Carvalho VSB. 2010. O impacto das megacidades na qualidade do ar em escala local e regional: os casos das Regiões Metropolitanas de São Paulo e do Rio de Janeiro. São Paulo, 2010. Ph.D. thesis. Universidade de São Paulo, Brazil.

Carvalho VSB, Freitas ED, Mazzoli CR, Andrade MF. 2012. Avaliação da influência de condições meteorológicas na ocorrência e manutenção de um episódio prolongado com altas concentrações de ozônio sobre a Região Metropolitana de São Paulo. Revista Brasileira de Meteorologia 27: 463-474. https://doi.org/10.1590/S0102-77862012000400009

Carvalho VSB, Freitas ED, Martins LD, Martins JA, Mazzoli CR, Andrade MF. 2015. Air quality status and trends over the Metropolitan Area of São Paulo, Brazil as a result of emission control policies. Environmental Science and Policy 47: 68-79. https://doi.org/10.1016/j.envsci.2014.11.001

Carvalho VSB, Martins FB, Silveira WW, Campos B, Simões JB. 2020. Variance analysis applied to ground-level ozone concentrations in the state of São Paulo, Brazil. Brazilian Journal of Chemical Engineering 37: 505-513. https://doi.org/10.1007/s43153-020-00045-7

Castanho ADA, Artaxo P. 2001. Wintertime and summertime São Paulo aerosol source apportionment study. Atmospheric Environment 35: 4889-4902. https://doi.org/10.1016/S1352-2310(01)00357-0

CONAMA. (2018). Resolução CONAMA 491 de 19 de novembro. Ministério do Meio Ambiente/Conselho Nacional do Meio Ambiente, Brasília. Available at: https://www.in.gov.br/web/guest/materia/-/asset_publisher/Kujrw0TZC2Mb/content/id/51058895/do1-2018-11-21-resolucao-n-491-de-19-de-novembro-de-2018-51058603 (accessed on September 22, 2022).

DENATRAN. 2017. Frota de veículos 2017. Departamento Nacional de Trânsito, Brazil. Available at: http://www.denatran.gov.br/estatistica/610-frota-2017 (accessed on February 22, 2018).

Fang GC, Lo CT, Zhuang YJ, Cho MH, Huang CY, Xiao YF, Tsai KH. 2017. Seasonal variations and sources study by way of back trajectories and ANOVA for ambient air pollutants (particulates and metallic elements) within a mixed area at Longjing, central Taiwan: 1-year observation. Environmental Geochemistry and Health 39: 99-108. https://doi.org/10.1007/s10653-016-9810-8

Ferreira DF. 2011. Sisvar: A computer statistical analysis system. Ciência e Agrotecnologia. Lavras 35: 1039-1042. https://doi.org/10.1590/s1413-70542011000600001

FEAM. 2003. Inventário de Fontes Emissoras de Poluentes Atmosféricos. Estudos de Dispersão Atmosférica e

Projeto de Rede Otimizada para a Região Metropolitana de Belo Horizonte/2003. Fundação Estadual do Meio Ambiente. Available at: http://www.feam.br/qualidade-do-ar/relatorios-artigos-e-publicacoes (accessed on: October 1, 2020).

FEAM. 2016. Relatório Anual de Qualidade do Ar – Ano Base 2015. Fundação Estadual do Meio Ambiente. Available at: http://www.feam.br/qualidade-do-ar/relatorios-artigos-e-publicacoes (accessed on October 1, 2020).

Gouveia N, Leon AP, Junger W, Lins JF, Freitas CU. 2019. Poluição do ar e impactos na saúde na Região Metropolitana de Belo Horizonte – Minas Gerais, Brasil. Ciência & Saúde Coletiva 24: 3773-3781. https://doi.org/10.1590/1413-812320182410.29432017

Govender P, Sivakumar V. 2020. Application of k-mean and hierarchical clustering techniques for analysis of air pollution: A review (1980-2019). Atmospheric Pollution Research 11: 40-56. https://doi.org/10.1016/j.apr.2019.09.009

Huryn SM, Gough WA. 2014. Impact of urbanization on the ozone weekday/weekend effect in Southern Ontario, Canada. Urban Climate 8: 11-20. https://doi.org/10.1016/j.uclim.2014.03.005

Im U, Incecik S, Guler M, Tek A, Topcu S, Unal YS, Yenigun O, Kindap T, Odman MT, Tayanc M. 2013. Analysis of surface ozone and nitrogen oxides at urban, semi-rural and rural sites in Istambul, Turkey. Science of the Total Environment 443: 920-931. https://doi.org/10.1016/j.scitotenv.2012.11.048

IBGE. 2018. População. Instituto Brasileiro de Geografia e Estatística. Available at: http://www.ibge.gov.br/home/estatistica/populacao (accessed on July 18, 2018).

Mansouri B, Hoshyari E, Mansouri A. 2011. Study on ambient concentrations of air quality parameters (O_3, SO_2, CO and PM_{10}) in different months in Shiraz city, Iran. International Journal of Environmental Sciences 1: 1440-1447. https://doi.org/10.6088/ijessi.00107020005

Mendes LMMC. 2018. Diagnóstico da qualidade do ar na Região Metropolitana de Belo Horizonte, MG. M.Sc. tesis. Universidade Federal de Itajubá, Brazil.

Miranda RM, Andrade MF, Worobiec A, Van Grieken R. 2002. Characterization of aerosol particles in São Paulo Metropolitan area. Atmospheric Environment 36: 345-352. https://doi.org/10.1016/S1352-2310(01)00363-6

Miranda RM, Andrade MF, Fornaro A, Astolfo R, Andre PA, Saldiva P. 2011. Urban air pollution: A representative survey of $PM_{2,5}$ mass concentrations in six Brazilian cities. Air Quality, Atmosphere & Health 5: 63-77. https://doi.org/10.1007/s11869-010-0124-1

Mohamad ND, Ash'aari H, Othman M. 2015. Preliminary assessment of air pollutant sources identification at selected monitoring stations in Klang Valley, Malaysia. Procedia Environmental Sciences 30: 121-126. https://doi.org/10.1016/j.proenv.2015.10.021

Moura IFS. 2016. Avaliação de MP_{10} na Região Metropolitana de Belo Horizonte. M.Sc. thesis, Universidade Federal de Minas Gerais, Brazil.

Pacheco MT, Parmigiani MMM, Andrade MF, Kumar P. 2017. A review of emissions and concentrations of particulate matter in the three major Metropolitan Areas of Brazil. Journal of Transport and Health 4: 53-72. https://doi.org/10.1016/j.jth.2017.01.008

Pant P, Harrison RM. 2013. Estimation of the contribution of road traffic emissions to particulate matter concentrations from field measurements: A review. Atmospheric Environment 77: 78-97. https://doi.org/10.1016/j.atmosenv.2013.04.028

Peláez LM, Santos JM, Albuquerque TTA, Reis Jr NC, Andreão WL, Andrade MF. 2020. Air quality status and trends over large cities in South America. Environmental Science & Policy 114: 422-435. https://doi.org/10.1016/j.envsci.2020.09.009

Pont V, Fontan J. 2001. Comparison between weekend and weekday ozone concentration in large cities in France. Atmospheric Environment 35: 1527-1535. https://doi.org/10.1016/S1352-2310(00)00308-3

Prudente CN, Raia A, Reis RJ. 2006. Comportamento do Vento na Região Metropolitana de Belo Horizonte. In: Anais do XIV Congresso Brasileiro de Meteorologia, Florianópolis.

Qin Y, Tonnesen GS, Wang Z. 2004. Weekend/weekday differences of ozone, NO_x, CO, VOCs, PM_{10} and the light scatter during ozone season in southern California. Atmospheric Environment 38: 3069-3087. https://doi.org/10.1016/j.atmosenv.2004.01.035

Radicchi ALA. 2012. A poluição na bacia aérea da região metropolitana de Belo Horizonte e sua repercussão na saúde da população. Revista Brasileira de Estudos de População 29: 195-198. https://doi.org/10.1590/S0102-30982012000100013

Sanchez-Ccoyllo OR, Andrade MF. 2002. The influence of meteorological conditions on the behavior of pollutants concentrations in São Paulo, Brazil. Environmental Pollution 116: 257-263. https://doi.org/10.1016/S0269-7491(01)00129-4

Santos TC, Carvalho VSB, Reboita MS. 2016. Avaliação da influência das condições meteorológicas em dias com altas concentrações de material particulado na Região Metropolitana do Rio de Janeiro. Engenharia Sanitária e Ambiental 21: 307-313. https://doi.org/10.1590/s1413-41522016139269

Santos TC, Reboita MS, Carvalho VSB. 2018. Investigação da Relação entre Variáveis Atmosféricas e a Concentração de MP_{10} e O_3 no Estado de São Paulo. Revista Brasileira de Meteorologia 33: 631-645, 2018. https://doi.org/10.1590/0102-7786334006

Santos FS, Pinto JA, Maciel FM, Horta FS, Albuquerque TTA, Andrade MF. 2019. Avaliação da influência das condições meteorológicas na concentração de material particulado fino ($MP_{2,5}$) em Belo Horizonte, MG. Engenharia Sanitaria e Ambiental 24: 371-381. https://doi.org/10.1590/s1413-41522019174045

Silva Júnior RS, Oliveira MGL, Andrade MF. 2009. Weekend/weekday differences in concentrations of ozone, NOx and non-methane hydrocarbon in the Metropolitan Area of São Paulo. Revista Brasileira de Meteorologia 24: 100-110. https://doi.org/10.1590/S0102-77862009000100010

Sicard P, Paoletti E, Agathokleous E, Araminienė V, Proietti C, Coulibaly F, De Marco A. 2020. Ozone weekend effect in cities: Deep insights for urban air pollution control. Environmental Research 191: 110193. https://doi.org/10.1016/j.envres.2020.110193.

Sicard P. 2021. Ground-level ozone over time: An observation-based global overview. Current Opinion in Environmental Science & Health 19: 100226. https://doi.org/10.1016/j.coesh.2020.100226

Storck L, Lopes SJ, Estefanel V, Garcia DC. 2011. Experimentação vegetal. 3rd ed. Editora da Universidade Federal de Santa Maria, Rio Grande do Sul, Brazil.

Tavares FVF, Barreto AA, Dutra EG, Jacomino VMF. 2010. Estudo do processo de dispersão de emissões veiculares em uma microrregião de Belo Horizonte (MG) utilizando simulação numérica. Engenharia Sanitária e Ambiental 15: 315-324. https://doi.org/10.1590/S1413-41522010000400004

Tyubee BT, Yiyeh GH, Onah MA. 2020. Assessment of air pollution in Gboko, Benue State, Nigeria. Urban Studies and Public Administration 3: 38-58. https://doi.org/10.22158/uspa.v3n4p38

Wolff GT, Kahlbaum DF, Heuss JM. 2013. The vanishing ozone weekday/weekend effect. Journal of the Air Waste Management Association 63: 292-299. https://doi.org/10.1080/10962247.2012.749312

WHO. 2006. Air quality guidelines. Global update 2005: Particulate matter, ozone, nitrogen dioxide, and sulfur dioxide. World Health Organization, Geneva. Available at: https://apps.who.int/iris/handle/10665/69477 (accessed on September 9, 2021).

Atmósfera 36(2), 343-354 (2023)
https://doi.org/10.20937/ATM.53052

Air pollution and mobility in the Mexico City Metropolitan Area in times of COVID-19

J. Eduardo VERA-VALDÉS[1,2]* and C. Vladimir RODRÍGUEZ-CABALLERO[2,3]

[1] *Department of Mathematical Sciences, Aalborg University, Skjernvej 4A, Aalborg, 9210, Nordjylland, Denmark.*
[2] *Center for Research in Econometric Analysis of Time Series, Fuglesangs Allé 4, Aarhus, 8210, Midtjylland, Denmark.*
[3] *Departmento de Estadística, Instituto Teconólogico Autónomo de México, Río Hondo 1, 01080 Ciudad de México, México.*
*Corresponding author; email: eduardo@math.aau.dk

Received: March 7, 2021; Accepted: July 27, 2021

RESUMEN

Este artículo analiza la relación entre COVID-19, contaminación atmosférica y movilidad en el transporte público de la Zona Metropolitana de la Ciudad de México (ZMCM). Analizamos si las restricciones a la actividad económica introducidas para mitigar los contagios por COVID-19 están asociadas con un cambio estructural en los niveles de contaminación atmosférica y movilidad en el transporte público. Nuestros resultados muestran que la movilidad en el transporte público se redujo significativamente dadas las recomendaciones gubernamentales. No obstante, la reducción en movilidad no fue acompañada de una reducción en contaminación atmosférica. Más aún, las pruebas de Granger-causalidad muestran que la relación de precedencia entre movilidad en transporte público y contaminación atmosférica desapareció como consecuencia de las restricciones. Por lo tanto, nuestros resultados sugieren que la contaminación atmosférica en la ZMCM se asocia primordialmente a actividad industrial y movilidad en transporte privado. En este sentido, el gobierno debería redoblar sus esfuerzos para implementar políticas públicas dirigidas a reducir contaminación industrial y el uso del automóvil.

ABSTRACT

This paper analyzes the relation between COVID-19, air pollution, and public transport mobility in the Mexico City Metropolitan Area (MCMA). We test if the restrictions to economic activity introduced to mitigate the spread of COVID-19 are associated with a structural change in air pollution levels and public transport mobility. Our results show that mobility in public transportation was significantly reduced following the government's recommendations. Nonetheless, we show that the reduction in mobility was not accompanied by a reduction in air pollution. Furthermore, Granger-causality tests show that the precedence relation between public transport mobility and air pollution disappeared as a product of the restrictions. Thus, our results suggest that air pollution in the MCMA seems primarily driven by industry and private car usage. In this regard, the government should redouble its efforts to develop policies to reduce industrial pollution and private car usage.

Keywords*:* pandemic, structural change, Granger-causality, particle matters, public transport.

1. Introduction

The COVID-19 pandemic is one of the most severe health crises in recent memory. The official death toll around the world surpassed one million as of September 29, 2020. Considering reporting problems in some countries and that the pandemic is still not under control, the actual death toll may not be known for several years.

Countries worldwide have imposed restrictions on economic activity to slow the rate of infection. Most of the restrictions can be motivated by the early results from the rate of infection in Wuhan, China (Kraemer et al., 2020; Prem et al., 2020). The restrictions on economic activity resulted in mass unemployment and reductions to the GDP worldwide. If the current pandemic follows similar dynamics as previous ones, the economic effects may be felt even in the long run (Rodríguez-Caballero and Vera-Valdés, 2020). In this context, assessing the effect of economic restrictions on public transport mobility and air pollution emissions is of great importance.

Most governments have imposed restrictions on public transport mobility throughout the COVID-19 pandemic. For example, Badr et al. (2020) and Cartenì et al. (2020) document the restrictions in the USA and Italy, respectively. These mobility limits may introduce a structural change in the global dynamic of public transport systems. As in other large cities, the local government in the Mexico City Metropolitan Area (MCMA) has imposed restrictions on the city's public mobility. The MCMA is an interesting case due to its high population density and the high number of workers in the informal sector. Therefore, it is relevant to formally study whether MCMA's restrictions cause a statistically significant reduction in passengers in the most used public transport systems: the subway system (Metro) and the bus rapid transit system (Metrobús).

In connection with the study of possible structural changes in public transport mobility, it is crucial to test if the government restrictions also result in lower air pollution levels. The evidence on the effect that restrictions have on pollution levels across the world is mixed. Significant reductions in nitrogen dioxide (NO_2) are encountered in, among others, Brazil, India, and Spain (Baldasano, 2020; Shehzad et al., 2020; Nakada and Urban, 2020). However, Adams (2020) finds that $PM_{2.5}$ (inhalable particles with diameters of 2.5 μm and smaller) levels do not change in response to a region-wide state of emergency in Ontario, Canada. Meanwhile, Berman and Ebisu (2020) find slight declines in $PM_{2.5}$ levels in the USA, but the results differ significantly between urban and non-urban counties. The authors argue that the different effects of economic restrictions between NO_2 and $PM_{2.5}$ may be explained by the fact that multiple non-transportation sources, including emissions from food industries and biomass burning, contribute to $PM_{2.5}$ levels. In this regard, they argue for more research on the impacts of the COVID-19 pandemic on industrial sourced pollutants. Moreover, Wang et al. (2020) find that severe air pollution events still occurred in most North China Plain areas even after all avoidable activities in China were prohibited on January 23, 2020.

This paper contributes to the literature by testing the effects of social distancing restrictions on public transport mobility and air pollution in the MCMA. Furthermore, we use the Granger-causality test to show that the precedence relation between public transport mobility and air pollution vanished during the restrictions.

This article proceeds as follows. The following section presents the data used in this study. Section 3 analyzes if the restrictions introduced due to COVID-19 result in structural changes in air pollution levels and mobility in the MCMA, while section 4 presents results from Granger-causality tests between mobility and air pollution in times of COVID-19. Section 5 concludes.

2. Data

The data come from the Portal de Datos Abiertos de la CDMX (Mexico City's data repository). We gather data on air pollution (PM_{10}, $PM_{2.5}$, and SO_2) levels at all stations and the number of passengers at all Metro and Metrobús stations. The data span from January 1, 2017, to July 31, 2020, and it presents several missing observations and some outliers that we clean first.

Outliers are detected in some of the Metro lines. A few observations (no more than 10 in total) show a thousand-fold increase compared to the rest. We attribute these differences to errors in capturing the data. We remove the outliers and impute them using observations in close proximity. It is worth pointing out that the small proportion of imputed outliers do not qualitatively alter the results.

Missing data are reported for some of the air pollution measuring stations. The missing values seem to randomly occur for some days. To correct the missing values, we use the vast amount of information to construct daily indexes for the air pollution measured in the MCMA. The index construction is motivated by

the strong correlation across air pollution measuring stations (Fig. 1). In this regard, missing observations are smoothed out by the construction of the index.

Furthermore, the data show some seasonal patterns.

For the mobility indexes, weekends and holidays show a clear seasonal pattern with a significant decrease in users. We control the seasonality by using data on nearby dates using linear imputation.

For the air pollution indices, the data show some natural seasonal patterns related to the weather. Therefore, we control the seasonality by using monthly dummy variables as is standard in the literature.

3. Structural changes due to COVID-19

The Mexican government established a Jornada Nacional de Sana Distancia (National Campaign of Social Distancing, NCSD), on March 23, 2020 (Secretaría de Salud, 2020). The plan established four measures to mitigate the effects of COVID-19 on the general population: (a) personal hygiene recommendations; (b) suspension of activities deemed non-essential; (c) postponement of mass gathering events (more than 5000 participants); and (c) guidelines for care of the elderly. The plan was heralded by "Susana Distancia", a fictitious heroine promoting social distancing. The preventive measures ended on May 30, 2020.

The goal of the plan was to impose social distancing measures and slow the spread of the virus. This section uses NCSD as a natural experiment to test if the restrictions introduced structural changes in pollution and public transport mobility.

As a first step, we study the trend mechanism of the series. We employ a broad range of unit root tests: the Augmented Dickey-Fuller (ADF) (Dickey and Fuller, 1979), the Phillips-Perron (PP) (Phillips and Perron, 1988), the Dickey-Fuller Generalized Least Squares (DG-GLS) (Elliott et al., 1996), and the Ng-Perron (Ng and Perron, 1995). In the unit root literature, it is well known that these tests suffer from a loss of power in the presence of structural breaks under the alternative hypothesis. As previously argued, we consider that the restrictions imposed due to COVID-19 provoked an exogenous break as in Perron (1989). Nonetheless, as a robustness exercise, we use unit root tests that allow for endogenous breaks (those not imposed by the practitioner). Therefore, we employ the tests of Zivot and Andrews (1992) (ZA92), that allows for a break under the alternative; Perron (1997), (P97), that allows for structural breaks under both the null and the alternative, and Kapetanios (2005) (K05), which allows for up to three breaks under the alternative.

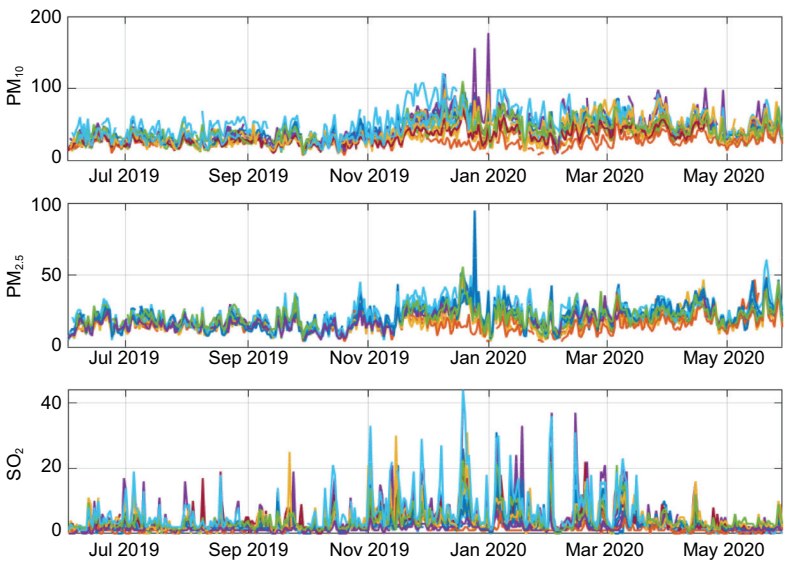

Fig. 1. Air pollution measurements in all stations in the MCMA.

Table I displays the results from the seven unit-root tests considered. As seen, we reject the null hypothesis of unit root processes in our variables. Note that the ADF and Ng-Perron tests fail to reject the null, possibly due to a loss of power due to the break. Nevertheless, note that the last four tests reject the possible unit root involved. Breaks in ZA92, P97, and K05 tests are located in the neighborhood of March 23, 2020. This date matches the origin of the NCSD.

Moreover, given that aggregation is used to construct the indexes, we estimate the fractional difference parameter for the series (Granger, 1980; Haldrup and Vera-Valdés, 2017). We use semiparametric estimators in the frequency domain to avoid the effect of the mean's specification to affect the results (Geweke and Porter-Hudak, 1983; Künsch, 1987; Shimotsu and Phillips, 2005). Results from the long memory estimates are presented in Table II. All tests find the data to be in the stationary range, well below the unit root scenario. Note that all stationarity tests consider the subperiod between January 1, 2017, and December 31, 2019, to avoid spurious results due to the possible structural change (Martínez-Rivera et al., 2012).

Once we guarantee that our data is stationary, we consider the following specification to test for a structural change:

$$y_t = \alpha_0 + \beta_0 t + \alpha_1 DU_t + \beta_1 DT_t + \varepsilon_t, \quad (1)$$

Table I. Unit root tests without constant term for pollutants, Metrobús, and Metro using full-sample data.

Variable	ADF	PP	DF–GLS	Ng–Perron	ZA92	P97	K05
PM_{10}	–13.31***	–17.65***	–4.28***	–11.07**	–16.72***	–11.06***	–14.31***
PM_{25}	–13.70***	–18.74***	–2.95***	–7.84**	–17.30***	–14.75***	–14.69***
SO_2	–20.29***	–23.18***	–5.05***	–14.50***	–21.67***	–21.46***	–21.49***
Metrobús	–2.07	–2.74*	–1.32**	–4.12	–10.32***	–9.11***	–9.09***
Metro	–3.35**	–13.14***	–3.04***	–13.33**	–17.50***	–11.85***	–14.38***

Notes: Lags in ADF and DF-GLS with Schwarz information criteria. Model with constant in PP. Model with intercept in ZA92 with two lags. P97 test considering model A. *, **, and *** denote rejection of the null hypothesis (unit root) at 10, 5, and 1%, respectively.

Table II. Long memory estimates, confidence intervals are shown below. Standard T1/2 bandwidth where T is the sample size.

Variable	GPH	LW	ELW
Metro	0.199 [–0.021-0.419]	0.234 [0.063-0.405]	0.271 [0.100-0.442]
Metrobús	0.643 [0.423-0.863]	0.632 [0.461-0.803]	0.660 [0.483-0.831]
PM10	0.408 [0.188-0.628]	0.378 [0.207-0.549]	0.419 [0.248-0.590]
PM2.5	0.347 [0.127-0.567]	0.358 [0.187-0.529]	0.402 [0.231-0.573]
SO_2	0.184 [–0.036-0.404]	0.174 [0.003-0.345]	0.201 [0.030-0.372]

GPH: Geweke and Porter-Hudak (1983) long memory estimators; LW: Künsch (1987) long memory estimators; ELW: Shimotsu and Phillips (2005) long memory estimators.(1983)

where y_t is the air pollution or mobility measure, and $t = [1,2,\ldots,T]'$, with T the sample size. Furthermore, DU and DT are dummy variables that model the possible structural change due to NCSD. That is, $DU = [0,\ldots,0,1,\ldots,1]'$, and $DT = [0,\ldots,0,1,2,\ldots,T_1]'$, where the non-zero elements start on March 23, 2020, and T_1 is the size of the subsample after that date. We test for a change in level if $\alpha_1 \neq 0$, and for a change in both level and trend if $\alpha_1 \neq 0$ and $\beta_1 \neq 0$.

The test for structural change proceeds as follows:

- Estimate the unrestricted model (Eq. [1]), and recover the unrestricted residual sum of squares, $URSS$, given by $URSS = \Sigma e_t^2$, where e_t are the residuals from estimating Eq. (1).
- Estimate the restricted model (Eq. [1]), with $\alpha_1 = 0$ and $\beta_1 = 0$, or $\beta_1 = 0$, and recover the restricted residual sum of squares, $RRSS$. The restricted sum of squares is given by $URSS = \Sigma e_t^2$, where e_t are the residuals from estimating Eq. (1) imposing $\alpha_1 = 0$ and $\beta_1 = 0$, or $\beta_1 = 0$.
- Compute the test statistic for the null hypothesis of no structural change by

$$F = \frac{\frac{RRSS - URSS}{r}}{\frac{URSS}{T-k}}, \qquad (2)$$

where T is the sample size, k is the number of parameters in the unrestricted model, and r is the number of restrictions.
- The test statistic follows an F distribution with r and $T - k$ degrees of freedom.

The structural change test assumes that the date of the break is known. As argued above, the restrictions due to COVID-19 are considered exogenous with a precise start date. Thus, the assumptions of the F-test are satisfied. Nonetheless, as a robustness exercise, we use the method developed by Bai and Perron (1998) to estimate the date of the break endogenously.

3.1 Mobility data

Figure 2 presents the mobility indexes for Metro and Metrobús. The data range from January 1, 2017, to July 31, 2020. The shaded region contains the period considered in NCSD. Also plotted are the estimates from the linear model in Eq. (1). We allow for both a change in level and a change in level and trend at the start of the NCSD. As can be seen from the figure, the dynamics of the mobility indices change significantly due to NCSD.

Table III presents the estimates from Eq. (1) allowing for a change in level and a change in level and trend and the structural change test results. The table presents some interesting findings. First, note

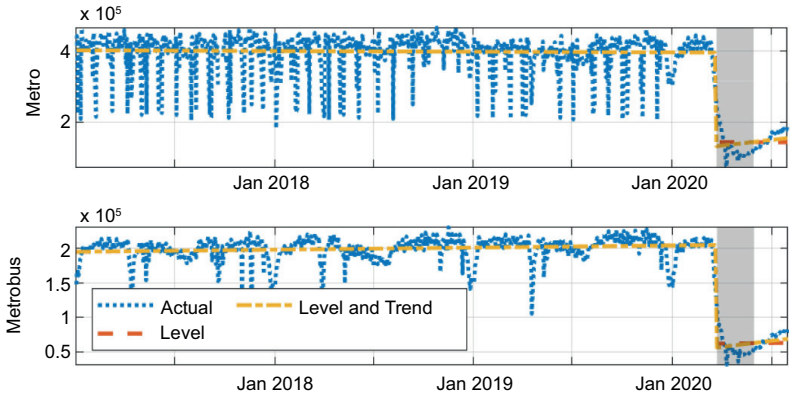

Fig. 2. Mobility indices in the Mexico City Metropolitan Area. The figure shows actual values (dotted blue) along with fitted values from the linear models with a change in level (dashed orange) and change in level and trend (dashed-dotted yellow). NCSD is shown in the shaded area.

Table III. Unrestricted equation estimation and test for structural change.

Variable	Change in level				Change in level and trend				
	α_0	β_0	α_1	F	α_0	β_0	α_1	β_1	F
Metro	$4(10^5)$***	–5.386	$-3(10^5)$***	2086***	$4(10^5)$***	–5.682	$-3(10^5)$***	215*	1046***
Metrobús	$2(10^5)$***	42.5***	$-2(10^5)$***	7006***	$2(10^5)$***	42.4***	$-2(10^5)$***	69.3*	3510***
PM_{10}	4.412***	–0.01***	–1.322	1.101	4.428***	–0.01***	–2.681	0.021	0.849
$PM_{2.5}$	1.806***	–0.00***	–1.431*	3.149*	1.805***	–0.00***	–1.384	–0.001	1.574
SO_2	1.027***	–0.00***	–0.028	0.006	1.029***	–0.00***	–0.157	0.002	0.039

*, **, and *** denote rejection of the null hypothesis at 10, 5, and 1%, respectively.

the different results regarding the trend coefficient, β_0. There is no significant trend in the number of Metro users, while a significant but small positive trend in Metrobús users over the last three years. The results suggest that more people started using public transit systems in the MCMA in the last few years.

Second, note the statistically significant decrease in the level of public transport users associated with NCSD. These results are in line with those from Badr et al. (2020) and Cartenì et al. (2020) for the USA and Italy. For the MCMA, the structural change is quite significant. The number of users more than halved during NCSD. That is, most users seem to have followed the government's recommendations and avoided the public transport system. Nonetheless, given the lack of data on the number of private cars and their number of passengers, we cannot extrapolate this result to state that people remained at home during NCSD. Furthermore, as a robustness exercise, we test all Metro and Metrobús lines individually for a structural change (Table IV and Fig. 3). The results from the robustness exercise are in line with the ones for the indices.

Regarding the method to estimate the break endogenously, it finds the break date on March 21, 2020, with NCSD contained in the confidence interval. That is, the date of the break estimated endogenously coincides with the start of NCSD.

Table IV. Structural change test for individual Metro and Metrobús lines and number of cyclists at several reporting stations.

Mobility	F_{level}	F_{trend}
Metro Line 1	1839***	930***
Metro Line 2	1729***	865***
Metro Line 3	1030***	515***
Metro Line 4	1382***	691***
Metro Line 5	934***	467***
Metro Line 6	945***	471***
Metro Line 7	953***	476***
Metro Line 8	1523***	762***
Metro Line 9	760***	380***
Metro Line A	559***	280***
Metro Line B	1878***	943***
Metro Line 12	1134***	533***
Metrobús Line 1	5429***	2716***
Metrobús Line 2	2947***	1471***
Metrobús Line 3	5646***	2824***
Metrobús Line 4	4993***	2616***
Metrobús Line 5	4469***	2232***
Metrobús Line 6	3446***	1720***

*, **, and *** denote rejection of the null (no structural change) at 10, 5, and 1%, respectively.

3.2 Pollution data

Figure 4 presents the air pollution indices. The figure shows PM_{10}, $PM_{2.5}$, and SO_2 levels from January 1, 2017, to July 31, 2020. The shaded region contains the period considered in NCSD. Also plotted are the estimates from the linear model in Eq. (1). We allow for both a change in level and a change in level and trend at the start of the NCSD. As shown in the figure, the dynamics of air pollution do not significantly change due to NCSD.

Furthermore, Table III presents the estimates from Eq. (1) allowing for a change in level and a change in level and trend and the structural change test results. The table presents some interesting findings.

First, the estimates show a significant decreasing trend for all pollutants across the period considered.

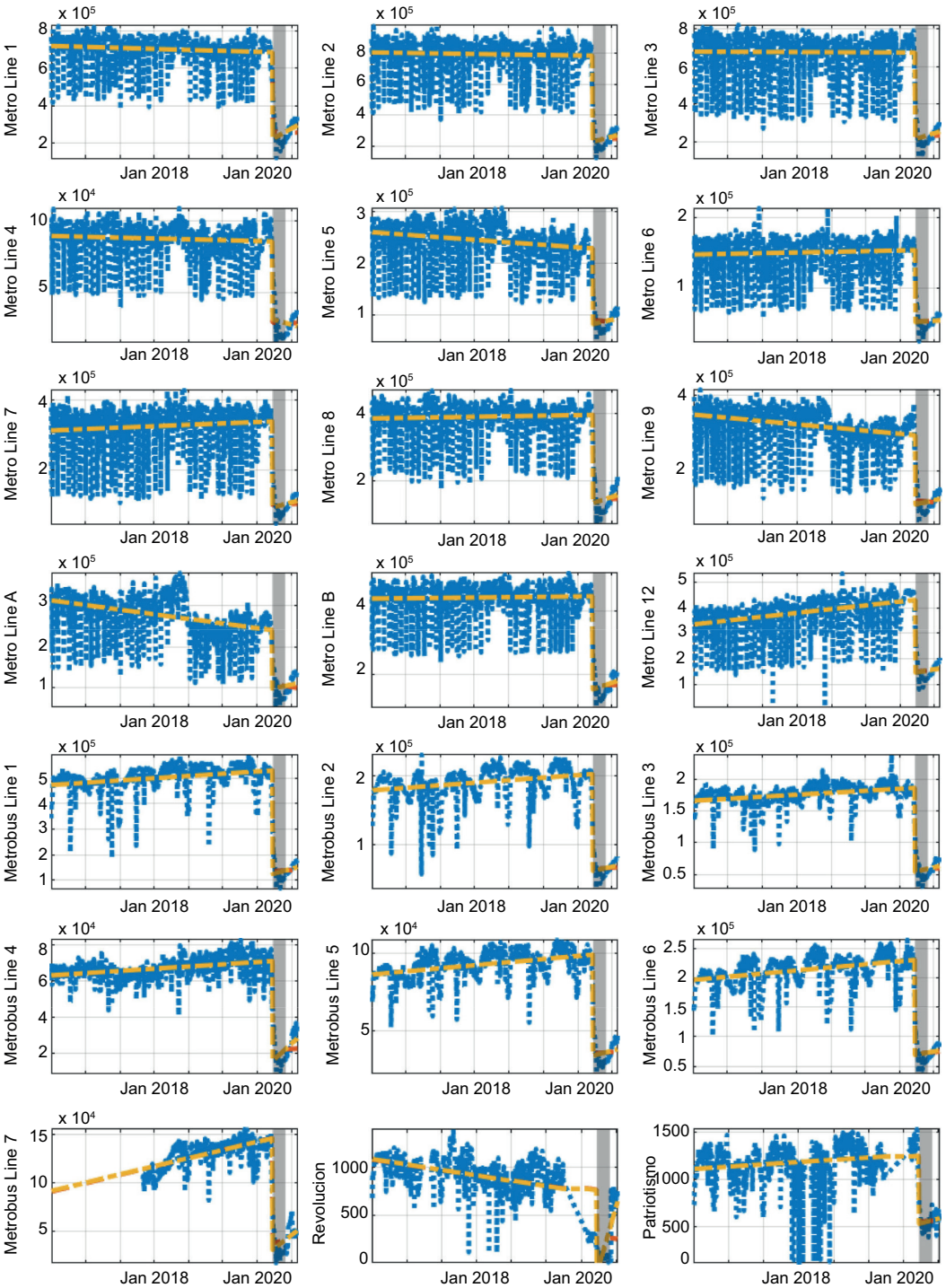

Fig. 3. Mobility in the MCMA. The figure shows actual values (dotted blue) along with fitted values from the linear model with a change in level (dashed orange) and change in level and trend (dashed-dotted yellow). NCSD is shown in the shaded area.

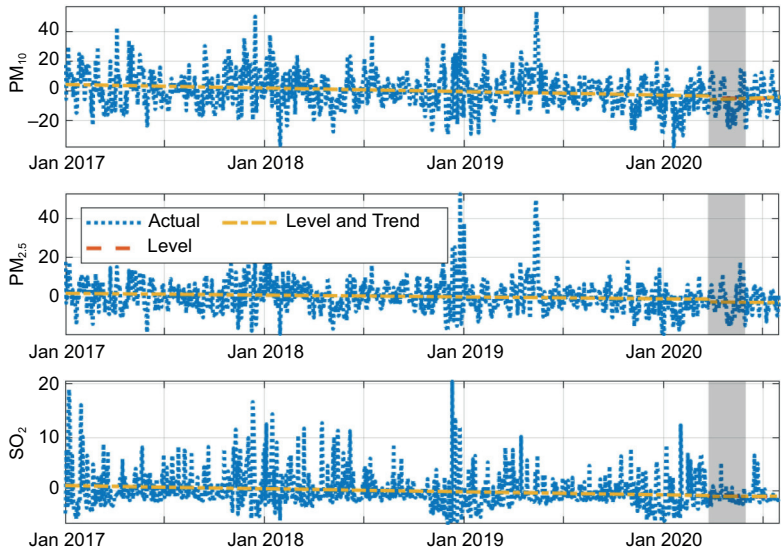

Fig. 4. Pollution indices in the Mexico City Metropolitan Area. The figure shows actual values (dotted blue) along with fitted values from the linear model with a change in level (dashed orange) and change in level and trend (dashed-dotted yellow). NCSD is shown in the shaded area.

Nonetheless, the estimates from the trend parameter are relatively small. Air pollutant levels have been decreasing through the years, but the decrease seems to be occurring at a slow pace.

Second, note that the null of no structural change is not rejected for both tests. The restrictions imposed by NCSD do not seem to be associated with a lower level of air pollution. These results are in line with the ones reported by Adams (2020) for Ontario, Canada. The author finds no significant reduction in $PM_{2.5}$ resulting from restrictions imposed due to COVID-19. Moreover, Wang et al. (2020) find that severe air pollution events still occurred in most of the North China Plain areas, even after all avoidable activities in China were prohibited on January 23, 2020.

Third, NCSD can be considered a natural experiment regarding public transport usage on air pollution. The lack of structural change in air pollution during NCSD coupled with the significant decrease in the mobility indices point to a non-significant effect of the number of users of the public transport system on pollution. As argued before, this may relate to a higher number of private cars during NCSD. Thus, these results suggest that tackling air pollution in the MCMA requires specific policies to reduce private car usage, particularly in light of the positive willingness to pay for clean air by inhabitants of the MCMA (Rodríguez-Sánchez, 2014; Filippini and Martínez-Cruz, 2016; Fontenla et al., 2019).

Finally, regarding the method to estimate the date of the break endogenously, the method does not find a break in 2020. Thus, our results are robust to an endogenous specification of the date of the break.

To properly assess the relationship between public transport and air pollution, the following section uses the Granger-causality test to assess if there exists a relation of precedence between them. Furthermore, we test if there is a change in this relationship after NCSD.

4. Granger-causality

In this section, we test the type of relation that exists between public transport mobility and air pollution indices. We use the concept of causality developed by Granger (1969). Although sometimes misrepresented in the literature, the test evaluates if a variable x has explanatory power on the variable y in the sense that x precedes y. We interpret this precedence as changes in variable x being related to changes in variable y.

Note that this does not necessarily denote a causal relation, given that a third variable could be driving both x and y. Nonetheless, the literature has settled on denoting this type of test as Granger-causality tests.

The test for Granger causality proceeds as follows:

- Estimate the unrestricted model given by

$$y_t = \alpha_0 + \sum_{i=1}^{k} \alpha_0 y_{t-i} + \sum_{i=1}^{m} \beta_i x_{t-i} + \varepsilon_t, \quad (3)$$

  where k, m are the number of lags included in the regression. In applied work, $k = m$ is common. From the estimation, we recover the residual sum of squares, *URSS*. Our analysis considers specifications with the same number of lags for both variables from the previous day and two days before.
- Estimate the restricted model given by

$$y_t = \alpha_0 + \sum_{i=1}^{k} \alpha_0 y_{t-i} + \varepsilon_t \quad (4)$$

  and recover the residual sum of squares, .
- Compute the test statistic for the null hypothesis of no structural change by

$$F = \frac{\frac{RRSS - URSS}{m}}{\frac{URSS}{T-k-m-1}}, \quad (5)$$

  where T is the sample size, k is the number of parameters in the unrestricted model, and m is the number of restrictions.
- The test statistic follows an F distribution with m and $T - k - m - 1$ degrees of freedom.

Intuitively, the test for Granger-causality assesses if the extra information contained in the additional variable helps explain the dynamics of the dependent variable better than the information contained in the lags of the dependent variable alone. This additional explanatory power is denoted in the literature as a precedence relation.

Granger-causality has been shown to produce spurious results (rejection of the null when the null is true) when the data follow processes with structural breaks or unit root processes (Ventosa-Santaulària and Vera-Valdés, 2008; Rodríguez-Caballero and Ventosa-Santaulària, 2014). Thus, our methodology relies on testing for Granger-causality before NCSD and contrast the results against estimation in the period after NCSD to avoid spurious results.

Table V presents the results from the Granger-causality test for the period before NCSD. The table shows that Metrobús Granger-causes air pollution in terms of PM_{10} and SO_2. Thus, there is statistical evidence that Metrobús usage changes are associated with PM_{10} and SO_2 air pollution changes. Nonetheless, recall that we cannot conclude that changes in Metrobús usage cause changes in air pollution in the typical sense, given that a third common factor for both could be the main driver behind both dynamics. In this context, more Metrobús users could be associated with more economic activity and more cars on the road.

Table V. Test for public transport Granger-causing air pollution in the periods before and after NCSD. The tests consider specifications including lags from the previous day, *GC*(1), and two days before, *GC*(2).

Variable-period	PM_{10}		$PM_{2.5}$		SO_2	
	GC(1)	*GC*(2)	*GC*(1)	*GC*(2)	*GC*(1)	*GC*(2)
Metro pre-NCSD	0.269	0.169	0.170	0.201	0.873	0.691
Metro post-NCSD	1.315	1.470	0.680	0.506	2.170	0.667
Metrobús pre-NCSD	3.448*	3.324**	0.477	0.915	4.090**	2.860*
Metrobús post-NCSD	1.829	1.816	0.803	0.536	2.602	0.867

*, **, and *** denote rejection of the null hypothesis (no Granger-causality) at 10, 5, and 1%, respectively.

To evaluate the effect that NCSD had on the precedence relation between public transport mobility and air pollution, Table V presents the results from the Granger-causality test for the post-NCSD period. The table shows that Granger-causality between public transport mobility variables and PM_{10} and SO_2 disappeared during NCSD. That is, changes in mobility indices do not precede changes in air pollution indices. In this regard, we argue that other sources of air pollution like industry and private car usage may be the major contributors to air pollution in the MCMA.

Overall, the results from the Granger-causality analysis support the notion that the link between public transport users and air pollution was temporarily broken during NCSD. The reduction in public transport users during NCSD was not accompanied by a reduction in air pollution.

5. Conclusions

This paper analyzes the relation between COVID-19, air pollution exposure, and mobility in the MCMA.

We test if the Mexican Government's economic and social restrictions to mitigate the spread of the virus produced a structural change in air pollution and mobility in the MCMA. Our results show that mobility in public transportation was significantly reduced following the government's recommendations. We find that mobility in public transit systems in the MCMA decreased by more than 65%. Thus, our results suggest that a large share of the inhabitants of the MCMA stopped using public transit during this period.

In connection with the structural change in mobility, we analyze if the restrictions resulted in lower air pollution in the MCMA. Our results show an overall decreasing trend in pollution levels in the MCMA throughout the years. Nonetheless, no statistically significant change is detected as a result of economic restrictions imposed due to COVID-19. That is, air pollution levels and trends were not affected as a product of the economic restrictions.

Furthermore, we use the Granger-causality test to analyze the existence of a precedence relation between public transport users and air pollution. Our results show that before the emergence of COVID-19, changes in public transport users were associated with changes in air pollution. Nonetheless, the precedence relation between public transport mobility and air pollution disappeared following the restrictions. These results suggest that additional factors as private car usage or industrial pollution may be more significant factors behind changes in air pollution.

The results from this analysis could help in designing policies aimed to reduce pollution levels in the MCMA. Structural changes in mobility in the public system do not seem to be associated with changes in air pollution levels. In this regard, our results suggest that tackling air pollution requires policies aimed explicitly at reducing industrial pollution and private car usage.

Acknowledgments

The authors would like to thank the anonymous referees for their valuable comments and suggestions. The paper has improved significantly because of them. All remaining errors are ours.

References

Adams MD. 2020. Air pollution in Ontario, Canada during the COVID-19 state of emergency. Science of The Total Environment 742: 140516. https://doi.org/10.1016/j.scitotenv.2020.140516

Badr HS, Du H, Marshall M, Dong E, Squire MM, Gardner LM. 2020. Association between mobility patterns and COVID-19 transmission in the USA: A mathematical modelling study. The Lancet Infectious Diseases 20: 1247-1254. https://doi.org/10.1016/S1473-3099(20)30553-3

Bai J, Perron P. 1998. Estimating and testing linear models with multiple structural changes. Econometrica, 66: 47-78. https://doi.org/10.2307/2998540

Baldasano JM. 2020. COVID-19 lockdown effects on air quality by NO_2 in the cities of Barcelona and Madrid (Spain). Science of The Total Environment 741: 140353. https://doi.org/10.1016/j.scitotenv.2020.140353

Berman JD, Ebisu K. 2020. Changes in U.S. air pollution during the COVID-19 pandemic. Science of The Total Environment 739: 139864. https://doi.org/10.1016/j.scitotenv.2020.139864

Cartenì A, di Francesco L, Martino M. 2020. How mobility habits influenced the spread of the COVID-19 pandemic: Results from the Italian case study. Science

of The Total Environment 741: 140489. https://doi.org/10.1016/j.scitotenv.2020.140489

Dickey DA, Fuller WA. 1979. Distribution of the estimators for autoregressive time series with a unit root. Journal of the American Statistical Association 74: 427-431. https://doi.org/10.2307/2286348

Elliott G, Rothenberg TJ, Stock H. 1996. Efficient tests for an autoregressive unit root. Econometrica 64: 813-836. https://doi.org/10.3386/t0130

Filippini M, Martínez-Cruz AL. 2016. Impact of environmental and social attitudes, and family concerns on willingness to pay for improved air quality: A contingent valuation application in Mexico City. Latin American Economic Review 25: 7. https://doi.org/10.1007/s40503-016-0037-y

Fontenla M, ben Goodwin M, González F. 2019. Pollution and the choice of where to work and live within Mexico City. Latin American Economic Review 28: 11. https://doi.org/10.1186/s40503-019-0072-6

Geweke J, Porter-Hudak S. 1983. The estimation and application of long memory time series models. Journal of Time Series Analysis 4: 221-238. https://doi.org/10.1111/j.1467-9892.1983.tb00371.x

Granger CWJ. 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37: 424-438. https://doi.org/10.2307/1912791

Granger CWJ. 1980. Long memory relationships and the aggregation of dynamic models. Journal of Econometrics 14: 227-238. https://doi.org/10.1016/0304-4076(80)90092-5

Haldrup N, Vera-Valdés JE. 2017. Long memory, fractional integration, and cross-sectional aggregation. Journal of Econometrics 199: 1-11. https://doi.org/10.1016/j.jeconom.2017.03.001

Kapetanios G. 2005. Unit-root testing against the alternative hypothesis of up to m structural breaks. Journal of Time Series Analysis 26: 123-133. https://doi.org/10.1111/j.1467-9892.2005.00393.x

Kraemer MUG, Yang CH, Gutierrez B, Wu CH, Klein B, Pigott DM, du Plessis L, Faria NR, Li R, Hanage WP, Brownstein JS, Layan M, Vespignani A, Tian H, Dye C, Pybus OG, Scarpino SV. 2020. The effect of human mobility and control measures on the COVID-19 epidemic in China. Science 368: 493-497. https://doi.org/10.1126/science.abb4218

Künsch H. 1987. Statistical aspects of self-similar processes. Proceedings of the First World Congress of the Bernoulli Society, 67-74.

Martínez-Rivera B, Ventosa-Santaulària D, Vera-Valdés JE. 2012. Spurious forecasts? Journal of Forecasting 31: 245-259 . https://doi.org/10.1002/for.1219

Nakada LYK, Urban RC. 2020. COVID-19 pandemic: Impacts on the air quality during the partial lockdown in São Paulo state, Brazil. Science of The Total Environment 730: 139087. https://doi.org/10.1016/j.scitotenv.2020.139087

Ng S, Perron P. 1995. Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag. Journal of the American Statistical Association 90: 268-281. https://doi.org/10.1080/01621459.1995.10476510

Perron P. 1989. The great crash, the oil price shock, and the unit root hypothesis. Econometrica 57: 1361-1401. https://doi.org/10.2307/1913712

Phillips PCB, Perron P. 1988, Testing for a Unit root in time series regression. Biometrika 75: 335-346. https://doi.org/10.1093/biomet/75.2.335

Prem K, Liu Y, Russell TW, Kucharski AJ, Eggo RM, Davies N, Flasche S, Clifford S, Pearson CAB, Munday JD. 2020. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: A modelling study. The Lancet Public Health 5: e61-e270. https://doi.org/10.1016/S2468-2667(20)30073-6

Rodríguez-Caballero CV, Ventosa-Santaulària D. 2014. Granger causality and unit roots. Journal of Statistical and Econometric Methods 3: 97-114.

Rodríguez-Caballero CV, Vera-Valdés JE. 2020. Long-lasting economic effects of pandemics: Evidence on growth and unemployment. Econometrics 8: 1-16. https://doi.org/10.3390/econometrics8030037

Rodríguez-Sánchez JI. 2014. Do Mexicans care about air pollution? Latin American Economic Review 23: 9. https://doi.org/10.1007/s40503-014-0009-z

Secretaría de Salud. 2020. Jornada Nacional de Sana Distancia. Available at: https://www.gob.mx/cms/uploads/attachment/file/541687/Jornada_Nacional_de_Sana_Distancia.pdf

Shehzad K, Sarfraz M, Shah SGM. 2020. The impact of COVID-19 as a necessary evil on air pollution in India during the lockdown. Environmental Pollution 266: 115080. https://doi.org/10.1016/j.envpol.2020.115080

Shimotsu K, Phillips PCB. 2005. Exact local whittle estimation of fractional integration. The Annals of Statistics 33: 1890-1933. https://doi.org/10.1214/009053605000000309

Ventosa-Santaulària D, Vera-Valdés JE. 2008. Granger-causality in the presence of structural breaks. Economics Bulletin 3: 1-14.Wang P, Chen K, Zhu S, Wang P, Zhang H. 2020. Severe air pollution events not avoided by reduced anthropogenic activities during COVID-19 outbreak. Resources, Conservation and Recycling 158: 104814. https://doi.org/10.1016/j.resconrec.2020.104814

ISSN 0187-6236

VNIVERSIDAD NACIONAL
AVTONOMA DE
MEXICO