University of São Paulo
Institute of Astronomy, Geophysics and Atmospheric Sciences
Department of Atmospheric Sciences

Gabriel Martins Palma Perez

Improving the quantitative precipitation forecast: a deep learning approach

São Paulo
2018

Gabriel Martins Palma Perez

Improving the quantitative precipitation forecast: a deep learning approach

Dissertation submitted in partial fulfillment of the requirements for the degree of Master of Sciences in the Institute of Astronomy, Geophysics and Atmospheric Sciences.

Major field: Meteorology
Advisor: Dr. Maria Assunção Faus da Silva Dias

São Paulo
2018

Dedicated to my family, girlfriend and cats. Specially the cats.

Acknowledgements

Firstly, I would like to thank my parents, for the unconditional love and support dedicated to me throughout my academic path. Their encouragement and patience are the keystones over which this research was built.

I also would like to express profound gratitude to my advisor, Dr. Maria Assunção, for guiding me since my undergraduate years. She always endorsed me to follow my own path, but steered me in the right direction when it was necessary. Her competence as a researcher is a role model for younger generations to follow.

I would like offer my thanks to Dr. Pedro Dias and all my colleagues and friends, in particular Dr. Elisa Sena, Victória Peli and Paola Bueno for the valuable insights and support. A special thanks to my friend Leonardo Kamigauti for first introducing me to artificial intelligence and motivating this project.

This project would have been impossible without CNPq financial support (grant no. 131341/2016-2).

I would like to offer an special thanks to my girlfriend Magnólia. Her encouragement, support, patience and love were fundamental in the past two years of my career.

This dissertation was written in LaTeX with the class IAGTESE, for IAG thesis and dissertations.

> "*We can only see a short distance ahead, but we can see plenty there that needs to be done*"
>
> Alan Turing, in Computing Machinery and Intelligence

Abstract

An accurate and reliable precipitation forecast is long recognized as one of the most urgent societal demands to meteorology. Particularly in densely populated areas, such as the East region of the State of São Paulo (ESP), intense rainfall events often lead to significant human and material losses. To address this demand, this study proposes an approach based on recent developments in the field of artificial intelligence. The results are grouped into four main parts. The first part concerns the evaluation of CHIRPS (Climate Hazards group InfraRed Precipitation with Station data) daily rainfall dataset in the area of study. When compared against purely in-situ observations, CHIRPS presents a root mean squared error of 7 mm/day and a coefficient of determination in the order of 40%. Considering the high spatial variability of rainfall, the evaluation is considered satisfactory. The second part, based on reanalysis data, investigates atmospheric patterns related with extreme precipitation events. This discussion supports the assembly of explanatory datasets and underpins physical interpretations of the deep neural networks (DNNs) solution.The third and fourth parts present the results of the DNNs precipitation forecast in the city of São Paulo and in ESP, respectively. Analyses of Probability of Detection (PoD) and False Alarm Rate (FAR) metrics reveal that the DNN forecast performance is unequivocally superior to the GFS 24-hour precipitation forecast. In the fourth part, the deep autoencoder is presented as an adequate non-linear dimensionality reduction tool that allows a spatial precipitation forecast with a small number of DNNs. The most relevant predictors are evaluated with one-hot and two-hot vectors stimulations, providing a quantification of the most relevant atmospheric patterns for predicting rainfall. A complementary discussion compares the top-ranked predictors with the patterns identified in the second part. This

comparison shows that, after the training process, the DNNs behave in agreement with the local atmospheric dynamics, thus narrowing the gap between machine learning and physics. The central conclusion of this study is that the proposed methodology provides trustworthy predictions with substantially higher accuracy when compared to model outputs, highlighting the applicability of modern artificial intelligence implementations in weather forecasting.

Resumo

Uma previsão de precipitação confiável e acurada é uma das demandas mais urgentes da sociedade para a meteorologia. Áreas com alta densidade populacional, como a região Leste do Estado de São Paulo (ESP), costumam sofrer perdas humanas e materias em eventos extremos de chuva. Para atender a essa demanda, esse estudo propõe uma abordagem baseada em técnicas modernas de inteligência artificial. Os resultados estão divididos em quatro partes principais. A primeira parte trata da avaliação da precipitação diária do CHIRPS (Climate Hazards group InfraRed Precipitation with Station data) na área de estudo. Quando comparado com observações de superfície, o CHIRPS apresenta erro quadrático médio na ordem de 7 mm/dia e coeficiente de determinação de 40%. Levando em consideração a alta variabilidade espacial da chuva, os resultados foram considerados satisfatórios. A segunda parte, através do uso de dados de reanálise, investiga os padrões atmosféricos relacionados com a ocorrência de eventos extremos. Essa discussão serve para dar suporte à definição de um conjunto de variáveis preditoras e para fundamentar as interpretações físicas das soluções encontradas pelas redes neurais. A terceira e quarta partes apresentam, respectivamente, os resultados da previsão de precipitaçao na cidade de São Paulo e no ESP. Análises das métricas de Probabilidade de Detecção (PoD) e Razão de Alarme Falso (FAR) mostram que o desempenho das redes neurais é significativamente superior a previsão de chuva acumulada em 24 horas do GFS. Na quarta parte, o deep autoencoder é apresentado como um método não-linear de redução de dimensionalidade que permite a previsão espacial de chuva com um número reduzido de redes neurais. As variáveis preditoras mais relevantes foram analisadas com o input de vetores one-hot e two-hot, permitindo a quantificação dos padrões mais relevantes para a previsão de chuva. Uma

discussão complementar compara os preditores mais relevantes com os padrões identificados na segunda parte. Essa comparação mostra que, após a etapa de treinamento, as redes neurais se comportam em acordo com a dinâmica atmosférica local. Esse resultado tem a proposta de aproximar o aprendizado de máquina e a física. A conclusão central do presente estudo é que a metodologia proposta fornece previsões confiáveis e com acurácia significativamente superior se comparada à saída de modelos numéricos, enfatizando a aplicabilidade de técnicas modernas de inteligência artificial na previsão de tempo.

List of Figures

List of Tables

Contents

Chapter 1

Introduction

Precipitation is a complex atmospheric phenomenon that involves numerous processes of a wide variety of scales. Microscale variables such as the concentration of cloud condensation nuclei are key elements to the formation of rain droplets in saturated environments. An unsaturated air-parcel originating in the planetary boundary layer can become saturated through the process of adiabatic cooling when uplifted to the lifting condensation level. This uplift can be caused by several forcings such as mesoscale circulations, significant topographic features, large-scale dynamics etc.

This complexity makes any attempt to accurately predict rainfall become a very demanding task. In the context of Numerical Weather Prediction (NWP), the rainfall rates either rely on parameterizations schemes or cloud resolving subroutines with very high computational cost. As a consequence, the precipitation forecast is often misleading and quantitatively unrealistic. Despite many advances in the last decades, the rainfall forecast is still insufficient to attend society practical demands. For example, Hamill (2014) shows that operational global and regional models were unable to predict an extreme precipitation event that caused severe human and material losses in Colorado, USA. Clark et al. (2009) analyzed ensemble forecasts of both parameterized and convection-allowing models. Their results show that convection-allowing models produce better precipitation forecasts, but still many improvements are necessary, specially for intense rainfall events.

Considering the population density and economic importance of the East Region of the State of São Paulo (ESP), an accurate and reliable precipitation forecast is of vital importance to mitigate the impacts of intense rainfall events. Specially in the summer, such events are frequent (Silva Dias et al., 2013) and their economic impact is significant (Haddad and Teixeira, 2015). In order to address this demand, this project employs state-of-the-art deep learning techniques to improve the 24-hour precipitation forecast in ESP.

1.1 Large-scale and mesoscale precipitation mechanisms in Southeastern Brazil

The precipitation regime in Brazil presents substantial spatial and temporal variability that can be partially attributed to large-scale systems. Figure 1.1 shows the annual precipitation averaged between 1981 and 2016 from CHIRPS (Climate Hazards group InfraRed Precipitation with Station data, further details in Methodology section) and Figure 1.2 shows the seasonally averaged precipitation fields in South America (SA).

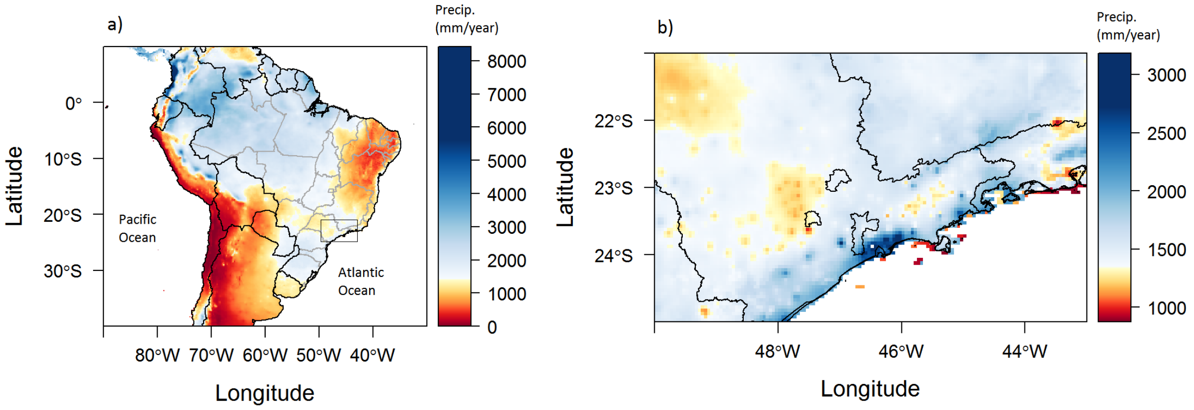


Figure 1.1: Annual average precipitation (1981-2016) from CHIRPS in South America and b) on the study area (b). The study area is bounded by a rectangle in (a).

During the austral summer, a northwest-southeast configuration shows as a dominant precipitation feature in Brazil (Figure 1.2a). This precipitation pattern distinguishes the South Atlantic Convergence Zone, SACZ (Carvalho et al., 2002), which is an important component of the South American Monsoon System, SAMS (Horel et al., 1989; da Silva and de Carvalho, 2007). The SACZ influences the total precipitation as well as the occurrence of extreme events in Southeastern Brazil, SEB (Carvalho et al., 2002).

Cold fronts and mesoscale convective systems (MCS) are also important precipitation mechanisms in Southern Brazil (SB) and SEB. Cavalcanti and Kousky (2009), through composite analysis, show that cold front passage have an expressive influence on the precipitation in SEB and SB. This influence is particularly evident during austral winter, when a local maximum is located across the states of Rio Grande do Sul, Santa Catarina and Paraná (Figure 1.2c). The occurrence of MCSs during summer, autumn and spring shifts this local maximum towards the south of Paraguay and northeast of Argentina. The high MCS activity observed in this region (Velasco and Fritsch, 1987; Durkee and Mote, 2010) influences the precipitation totals in SEB and SB specially during summer, autumn and fall (Figures 1.2a, 1.2b and 1.2d).

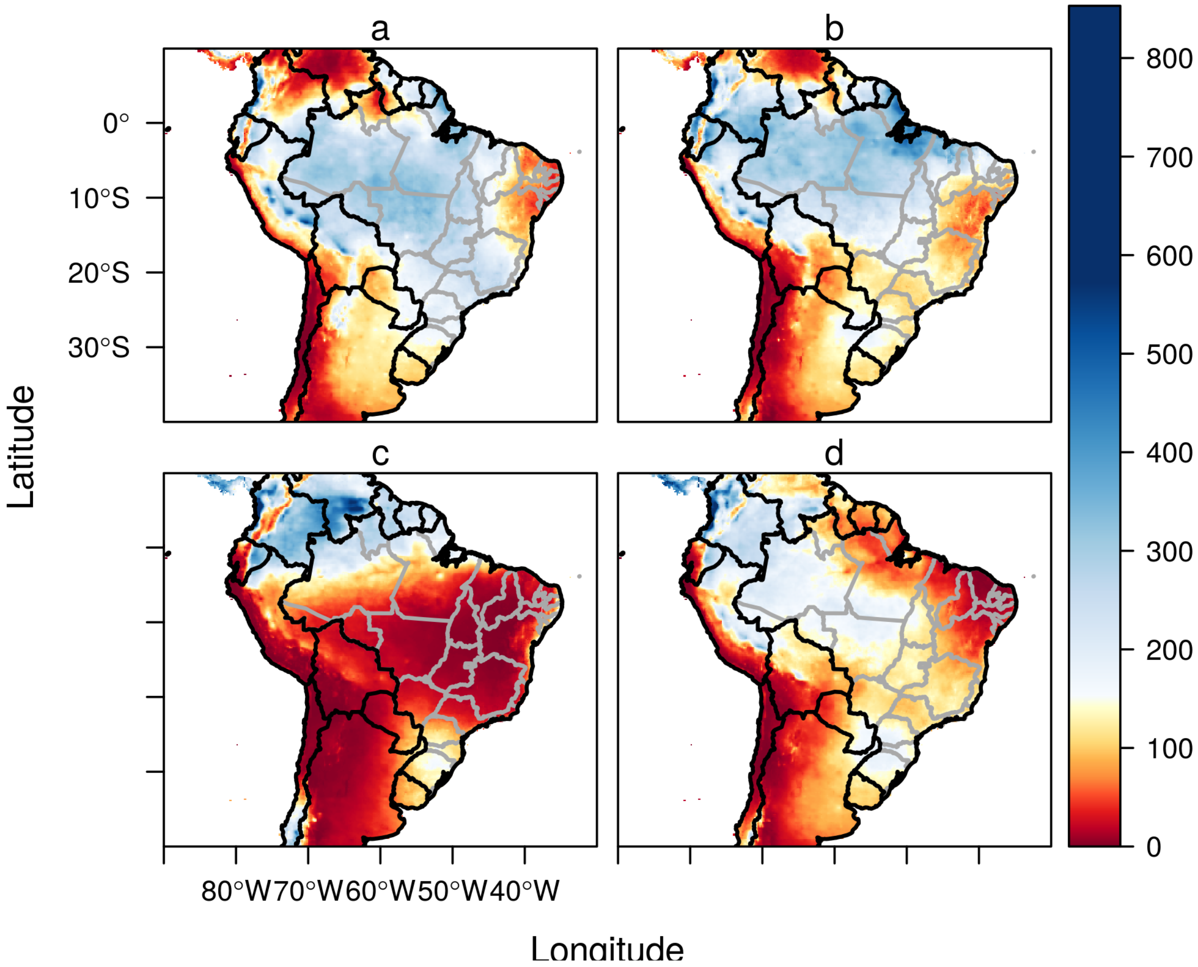


Figure 1.2: Monthly averaged precipitation (mm/month, 1981 - 2016) from CHIRPS in South America during austral summer (a), autumn (b), winter (c) and spring (d).

The sea surface temperature (SST) is also related locally and remotely with precipitation mechanisms in SEB. Locally, it influences the intensity and position of SACZ (Chaves and Nobre, 2004; Jorgetti et al., 2014) and the sea-breeze passage (Perez and Silva Dias, 2017). Remotely, disturbances in SST trigger wave trains propagation that influence the precipitation regime in SEB. Climatic oscillations such as Atlantic Multidecadal Oscillation (AMO) and El-Nino Southern Oscilation (ENSO) have been related with precipitation patterns (de MBS Xavier et al., 1995; Prado, 2010) and sea breeze circulation (Perez and Silva Dias, 2017) in the state of São Paulo.

1.2 *Local circulations and urban characteristics of the East region of the State of São Paulo*

Local circulations also play a major role in the development of precipitating clouds. These circulations are generated by surface heterogeneities, such as the presence of urban areas, a bodies of water or substantial changes in topography. Figure 1.3 shows the topography of the State of São Paulo, located in SEB. The topographic features and proximity with Atlantic Ocean are ingredients to the development of local circulations, such as mountain-valley circulation and sea/land breeze.

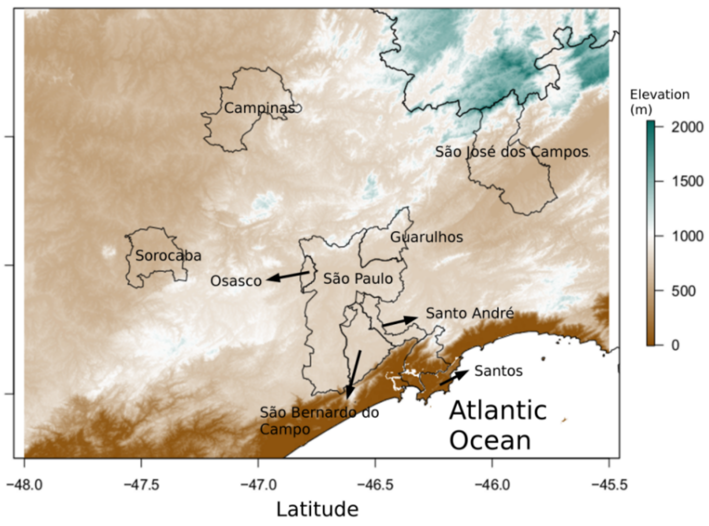


Figure 1.3: Topography of the East region of the State of São Paulo and the 9 most populous municipalities of the State.

ESP concentrates the nine most populous cities of the State, including the city of São Paulo – the most populous city in Southern Hemisphere with approximately 12 million inhabitants (Brazilian Institute of Geography and Statistics 2015 population estimates).

The contrasting land use caused by urbanization (see Figure 1.4) induces urban heat-island circulations that influence the precipitation regime. Silva Dias et al. (1995) and Freitas et al. (2007), through numerical simulations, demonstrated the association between the urban heat-island and the formation of severe rainfall events in São Paulo.

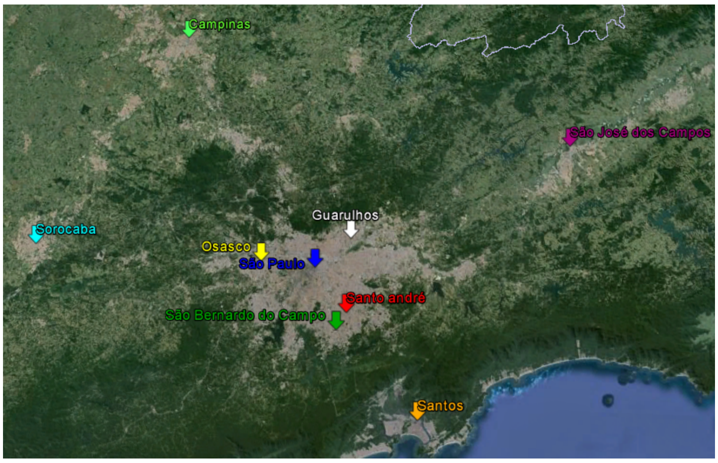


Figure 1.4: Satellite view of the East region of the State of São Paulo identifying the 9 most populous municipalities of the State. Source: Google Earth.

1.3 Statistical methods to improve precipitation forecast

Obtaining a realistic rainfall prediction is a long standing issue in weather forecasting. In 1960, the National Weather Service of the United States of America recognized the quantitative precipitation forecast (QPF) as a matter of high social and economic value (Olson et al., 1995). Since then, many studies have aimed to improve the probabilistic and quantitative precipitation forecast. Many of these studies employ post-processing techniques in NWP ensembles to crate probability density functions (Hamill et al., 2004; Sloughter et al., 2007). In the case of Hamill et al. (2004), the author trains a logistic regression model to post-process ensemble reforecast data. Logistic regression is an early class of artificial neural network (ANN), equivalent to the single-layer perceptron, discussed in the next section.

Other authors have employed ANNs as a tool to forecast rainfall. Ramirez et al. (2005) show that an ANN trained with atmospheric fields from the ETA model provides better rainfall predictions than multiple linear regression in São Paulo. Nasseri et al. (2008) combined genetic programming and ANNs to improve the precipitation forecast in

Sidney, Australia. Anochi and Velho (2016) propose the use of the Multi-Particle Collision Algorithm to optimize the topology of ANNs for climate predictions. ANNs also have been used to downscale and remove the bias of hydroclimatic projections of IPCC models in the Han River, South Korea (Kang and Moon, 2017).

Given the limitations of operational NWP models, the aforementioned studies show that statistical methods are powerful tools to provide a potentially more accurate rainfall forecast. The next section presents a detailed discussion ranging from early artificial neural networks implementations until recent deep learning developments.

1.4 Objectives and dissertation outline

The availability of massive datasets and computational resources, allied to new methodologies, are responsible for the recent success of deep learning applications in science and engineering. Deep learning is now employed in progressively more complex systems, such as: handwriting and speech recognition, chat-bots, books and movies recommendation, self-driving vehicles, medical diagnosis and even developing original music and paintings.

Precipitation is perhaps the most critical meteorological variable for human activities and the greatest limitation in numerical weather prediction. The East Region of the State of São Paulo is a densely populated and economically relevant area in Brazil that would greatly benefit from an improved rainfall forecast. For this reason, the general aim of this research is to employ recent and successful machine learning techniques to develop a more reliable and accurate precipitation forecast methodology in this region. A secondary aim of this project is to improve the understanding of how deep neural networks learn the underlying physical mechanisms of rainfall. To achieve these goals, this monograph is divided in the following parts:

- Chapter 1 - The first chapter is dedicated to supply to the reader a background of the main rainfall related meteorological systems and geographical characteristics of the study area.
- Chapter 2 - This chapter aims to provide to the reader with no machine learning

previous knowledge a review of the main developments that culminated into the present field of deep learning;

- Chapter 3 - This chapter presents a brief overview of the datasets employed in this project. Each following chapter provides further details about particular usages of the data;

- Chapter 4 - The third chapter presents an evaluation of CHIRPS (the precipitation dataset chosen as target variable) against observational precipitation data;

- Chapter 5 - The fourth chapter presents atmospheric patterns associated with extreme precipitation events;

- Chapter 6 - This chapter presents the methodology, discussion and results about the DNNs implementation for precipitation forecast in the city of São Paulo. This chapter also quantifies the most relevant predictors and discusses the physical validity of the solution found during the training process;

- Chapter 7 - This chapter presents the deep learning framework and results of the spatial rainfall forecast in ESP. Deep auto-encoders are presented as an appropriate non-linear dimensionality reduction tool for encoding high-resolution precipitation data;

- Chapter 8 - This final chapter presents a brief summary of the results, final remarks and discusses potentially relevant future studies.

Chapter 2

Artificial Neural Networks preliminaries

ANN is a class of statistical models in the machine learning framework with plenty applications in science and engineering. Some groundbreaking advances in the last decade have greatly improved ANNs capabilities and popularity and now they are part of many practical aspects of society (LeCun et al., 2015). The most canonical example of ANN is the Multilayer Perceptron (MLP), which consists of a non-linear mapping of a target function into some input space.

The MLP originates from early models of the biological brain. Rosenblatt (1958), based on previous studies about the brain functioning, developed the perceptron (Figure 2.1): a model that could optimize its parameters (w and b) to approximate a function y given a set of explanatory variables $x_1, x_2, ..., x_n$ (Equation 2.1). Despite using a non-linear activation function g (traditionally the logistic function given by Equation 2.2), the perceptron is a linear model in the context of classification problems since it draws a simple linear boundary between two categories.

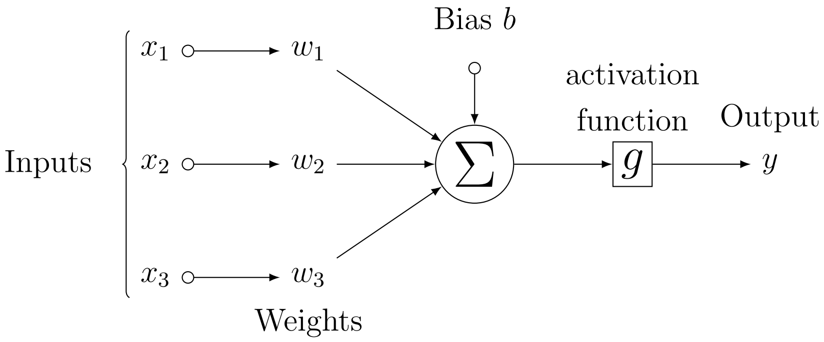


Figure 2.1: Diagram of a perceptron. In this example, the input vector consists of three variables x_1, x_2 and x_3 that are multiplied by a weight vector of the same size and added to a bias b. The summation of the result is passed to the activation function g to yield an output y.

$$y = g(\sum_{x=1}^{n} w_i x_i + b) \tag{2.1}$$

$$g(x) = \frac{1}{1 + e^x} \tag{2.2}$$

The MLP is Rossenblatt's perceptron non-linear extension. It consists of multiple interconnected layers with a number of perceptrons (hereafter “neurons” or “nodes”) in each layer (Figure 2.2). The input layer and the output layer are commonly called “visible” layers, while the layers in between are called “hidden” layers. It has been shown (Cybenko, 1989; Hornik et al., 1989) that a single hidden-layer MLP with a finite number of neurons and a logistic activation function can approximate any continuous function arbitrarily well. However, there is no guarantee if it is feasible to optimize a single-layer (i.e., “shallow”) network with a large set of weights (Vidal et al., 2017).

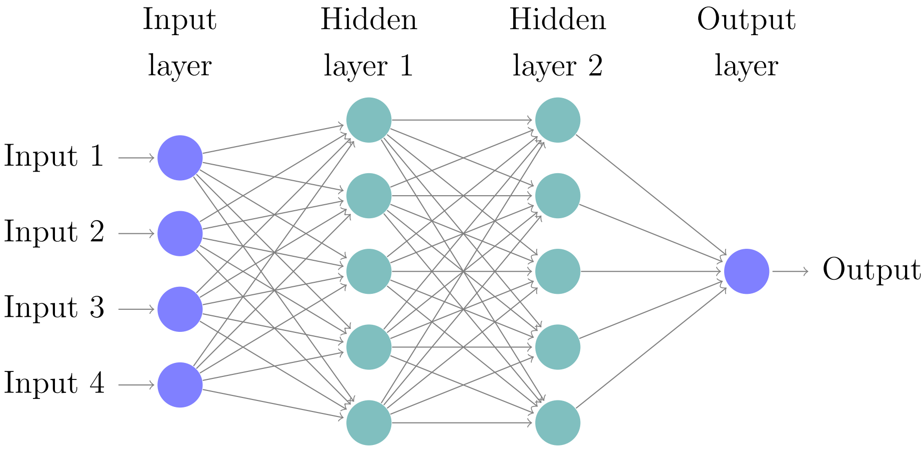


Figure 2.2: Diagram of a Multilayer Perceptron with four variables in the input layer, two hidden layers and one output variable. The biases are omitted.

Deeper architectures (i.e., several hidden layers) are often more practicable and perform better than shallow ones (Vidal et al., 2017; Lin et al., 2017). The increased number of hidden layers allows the MLP to create an improved latent representation of the input data that optimizes the output predictions (LeCun et al., 2015). Each hidden layer has the ability to disentangle information (Achille and Soatto, 2017) and highlight features of the raw input data that are more significant to the prediction or classification task (LeCun et al., 2015). Lin et al. (2017) further argue that hierarchical physical processes (Markov chains) are better represented in deeper architectures. However, not until 2006 (Hinton et al., 2006) it has become possible to train a deep MLP successfully.

A single-layer perceptron is traditionally trained by the Stochastic Gradient Descent, SGD (Robbins and Monro, 1951; Kiefer and Wolfowitz, 1952). The SGD simply evaluates some error function for a batch of input samples and update the weights and biases towards the error minimum. Several authors (Werbos, 1974; LeCun, 1985; Rumelhart et al., 1986) have shown that SGD could be employed to train MLPs with the assistance of the chain-rule of derivatives to propagate the gradients through the hidden layers. This method is called “error-backpropagation” or, simply, “backpropagation”.

In principle, the backpropagation could train a MLP of any depth. However, its performance dramatically decreases when training deeper architectures (usually more than two or three hidden layers). This phenomenon, called “vanishing gradient problem”, occurs when the SGD is applied through several layers with sigmoidal activations (such as Equation 2.2), culminating in poor optimization of the bottom layers (i.e., the layers subsequent to the input). Goodfellow et al. (2016) provide a detailed discussion of this issue. The vanishing gradient problem, therefore, limited the domain of ANNs to simpler applications, resulting in a decline of interest in artificial neural networks in the decade of 1990.

Following the work of Hinton et al. (2006), numerous authors provided different approaches to facilitate the training of deep neural networks (DNNs). Glorot et al. (2011) introduced the Rectifier Linear Unit (ReLU, Equation 2.3) as a more efficient and biologically plausible activation function. In contrast to the logistic function, ReLUs are not affected by the vanishing gradient problem. Note that, despite being linear for positive values, a summation of ReLU layers can approximate any non-linear continuous function (Glorot et al., 2011). Figure 2.3 compares ReLU with the logistic function.

$$y = \max\left(0, x\right) \tag{2.3}$$

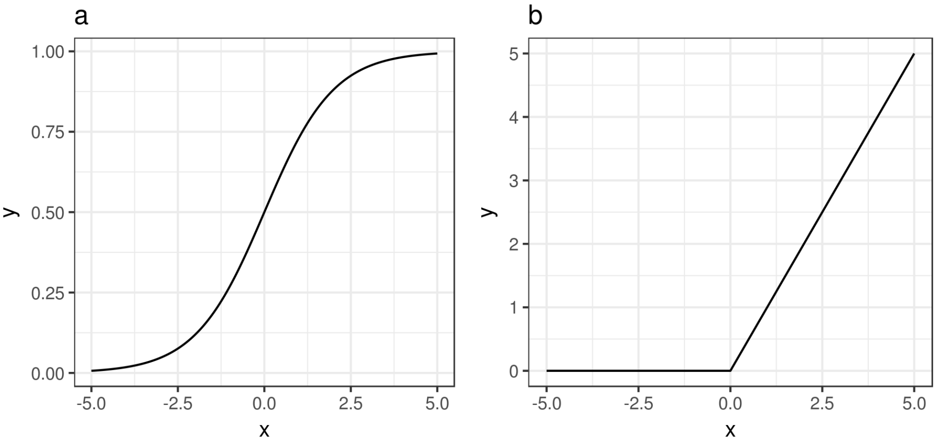


Figure 2.3: a) Logistic activation function given by Equation 2.2 and b) Rectifier Linear Unit (ReLU) given by Equation 2.3.

In practice, employing ReLU as activation function will cause a significant portion of neurons to produce a null output for each input sample. This property, called sparsity, has some desirable features, as discussed by Glorot et al. (2011). One of these features is the information disentanglement between neurons. In other words, ReLUs sparsity promotes each neuron to be an unique feature extractor, reducing redundancy of information. Another desirable property is the “efficient variable-size representation”, underlined by

Glorot et al. (2011). The fact that some neurons can have a null output imply an adaptable effective size/complexity of the neural network. In practice, the DNN can activate different numbers of neurons for different input vectors, adapting itself to the complexity of each input. As a consequence of these properties, ReLUs have become the prevalent activation function in deep learning (LeCun et al., 2015).

“Dropout” is another widely adopted technique in most modern DNNs implementations. It was first proposed by Hinton et al. (2012) as a regularization method that reduces the model complexity to prevent deep architectures from overfitting the training data. Overfitting occurs when the model is complex enough to learn even undesirable noise in the training set, thus worsening the predictions of new observations. Similarly to ReLUs, dropout also compels the nodes to learn independent features by “turning off” a random set of neurons in the forward pass[1] for each training sample. When predicting new data, the model uses an approximation of the average of all possible dropout configurations. Section 5.1. discusses the dropout algorithm in greater detail.

---

[1] Forward pass is the process in which a sample of data is injected at the input layer of an ANN and returns an output after passing through each hidden layer

Chapter 3

Data description

This chapter presents an overview of the datasets used in this study[1]. Two main categories of data were used: 1) precipitation observations and 2) numeric model output and reanalysis. The first corresponds to the target values for the statistical models and the latter corresponds to the explanatory variables. Two types of explanatory datasets were assembled: one representing an operational situation (model output) and the other representing an idealized situation (reanalysis).

The choice of a reliable precipitation dataset is particularly sensitive, considering that the nature of precipitation is highly non-linear both in space and time. CHIRPS (Funk et al., 2014) was chosen as target dataset to train the DNNs because of its high spatial resolution (0.05°) and assimilation of different observational sources. Its assimilation procedure combines: a) monthly precipitation climatology from CHPClim (Climate Hazard Precipitation Climatology); b) infrared (IR) precipitation estimates from geostationary satellites, c) TRMM (Tropical Rainfall Measurement Mission) 3B42 product; d) precipitation fields from CFSv2 (Climate Forecast System version 2) and e) in-situ rain gauges. The next chapter provides a detailed discussion and evaluation of CHIRPS.

The selection of an explanatory dataset is less sensitive than the choice of a target dataset for two main reasons: 1) most atmospheric fields (e.g., temperature, pressure, humidity etc) have relatively smooth spatial configurations and less abrupt changes in time when compared to precipitation; 2) the hidden layers of the DNNs can compensate offsets and systematic errors in the input dataset. Nevertheless, it is important to investigate how

---

[1] Further information is contained in the following chapters.

the quality of the input variables influences the quality of predictions. For this reason, data from the ERA-Interim reanalysis was employed as an idealized forecast and the output from GEFS Reforcast Version 2 was employed as an realistic (operational) forecast. Further details about these datasets are found in the following chapters.

In addition to the GEFS Reforecast and ERA-Interim, three climatic indices were employed as explanatory variables: AMO, SOI (Southern Oscillation Index) and PDO (Pacific Decadal Oscillation). As discussed in Section 1.1., these indices have been reported to influence rainfall in SEB. Also, they can provide information to the model about changes in the global climatic situation.

Chapter 4

CHIRPS evaluation against in-situ observations

As discussed in the previous chapter, the choice of the precipitation dataset is a sensitive issue and, for that reason, this chapter is dedicated to evaluate CHIRPS against purely in-situ observations. CHIRPS itself incorporates rain gauge observations, but it combines satellite and model data as well. At the same time this diversity of sources is a strong feature, it may hinder the interpretation of errors. Furthermore, CHIRPS assimilation procedure is not homogeneous as the availability of its sources changes in time. Figure 4.1 shows the number of rain gauges assimilated in Brazil (4.1a) and its spatial distribution in 1985 (4.1b) and 2015 (4.1c). Other issues can affect the homogeneity of the time series, such as the availability of TRMM, that ceased its operation in April 2015.

4.1 *Methodology*

Xavier et al. (2016) developed an observational gridded (0.25°x 0.25°) dataset with the interpolation of 3625 rain gauges and 735 meteorological stations throughout Brazil (hereafter METBR). All stations used in METBR were subjected to a quality control that verifies homogeneity and removes outliers. This database was chosen to evaluate CHIRPS since it only assimilates rain gauges, facilitating the interpretation of results.

The coefficient of determination (ρ^2), the mean error (ME) and the Root Mean Square Error (RMSE) were employed to compare CHIRPS precipitation rate (P^{chirps}) with METBR precipitation rate (P^{ref}):

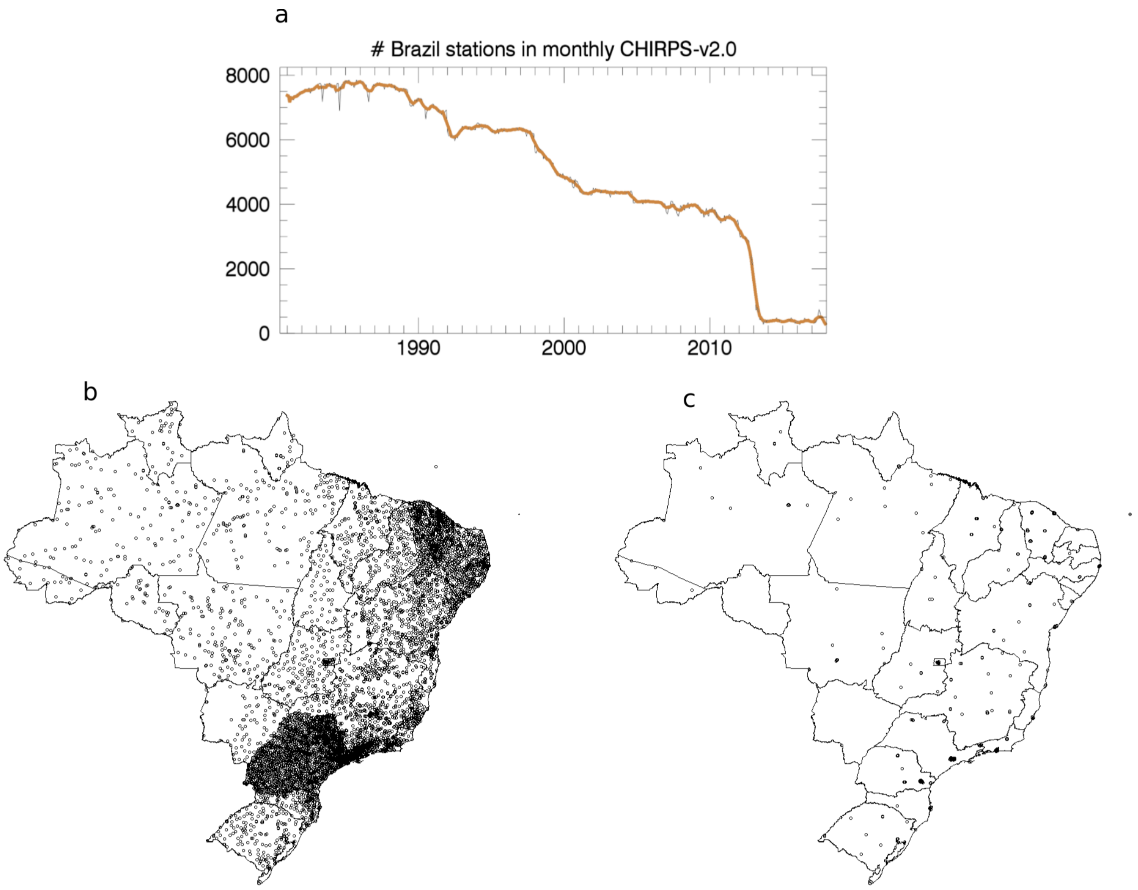


Figure 4.1: Number of Brazil rain gauges used in CHIRPS (a; source: http://chg.geog.ucsb.edu/data/chirps/stations/) and their location in 1985 (b) and 2015 (c)

$$\rho^2 = \left(\frac{cov(P^{chirps}, P^{ref})}{\sigma^{chirps}\sigma^{ref}} \right)^2 \tag{4.1}$$

$$ME = \sum_{i=1}^{n} \frac{P_i^{chirps} - P_i^{ref}}{n} \tag{4.2}$$

$$RMSE = \sum_{i=1}^{n} \sqrt{\frac{(P_i^{chirps} - P_i^{ref})^2}{n}} \tag{4.3}$$

Where σ is the standard deviation.

4.2 Results and discussion: CHIRPS against METBR

Figure 4.2 compares CHIRPS (disaggregated to a 0.25°grid) and METBR daily precipitation rates: 4.2a shows the coefficient of determination ρ^2 (i.e., explained variance in %), 4.2b shows the RMSE, 4.2c shows the ME and 4.2d shows the annual average RMSE and ME. An evident feature of Figures 4.2a and 4.2b is that higher (lower) RMSE (ρ^2) occurs preferentially in regions of higher accumulated rainfall (see Figure 1.1), such as the coastline and the western region of the States of Paraná and Santa Catarina. This generally indicates that, when it comes to intense precipitation systems, such as MCSs and coastal storms, there is less agreement between the two products.

The disagreement between CHIRPS and METBR can be discussed in terms of their different data sources and procedures. For example, METBR, relies on interpolations that may not be valid in situations of high spatial variability. CHIRPS, on the other hand, relies on a numerical model (CFSv2) to disaggregate precipitation pentads [1] into daily rainfall rate. Numerical models, however, have limited ability to reproduce realistic precipitation fields, specially in extreme situations (Clark et al., 2009; Hamill, 2014).

Figure 4.2c reveals a few locations where CHIRPS particularly underestimates METBR precipitation: the valley of São Francisco river in Minas Gerais, the coasts of Rio de Janeiro, Paraná and the southern coast of São Paulo. These underestimations could be related to the geostationary satellite IR sensor inability to quantify precipitation from orographic warm clouds that are favored by the topographic features of these regions. Despite these limitations, Figure 4.2d shows an increasing agreement between CHIRPS and METBR in terms of RMSE and ME.

4.3 Results and discussion: CHIRPS against IAG meteorological station

The meteorological station of the Institute of Astronomy, Geophysics and Atmospheric Sciences (IAG) of the University of São Paulo is a conventional meteorological station located in the city of São Paulo. Its uninterrupted record of observations date back to 1933

---

[1] CHIRPS procedure produces accumulated precipitation pentads (5 days) from observations. The pentads are subsequently disaggregated into daily precipitation based on CFSv2 fields.

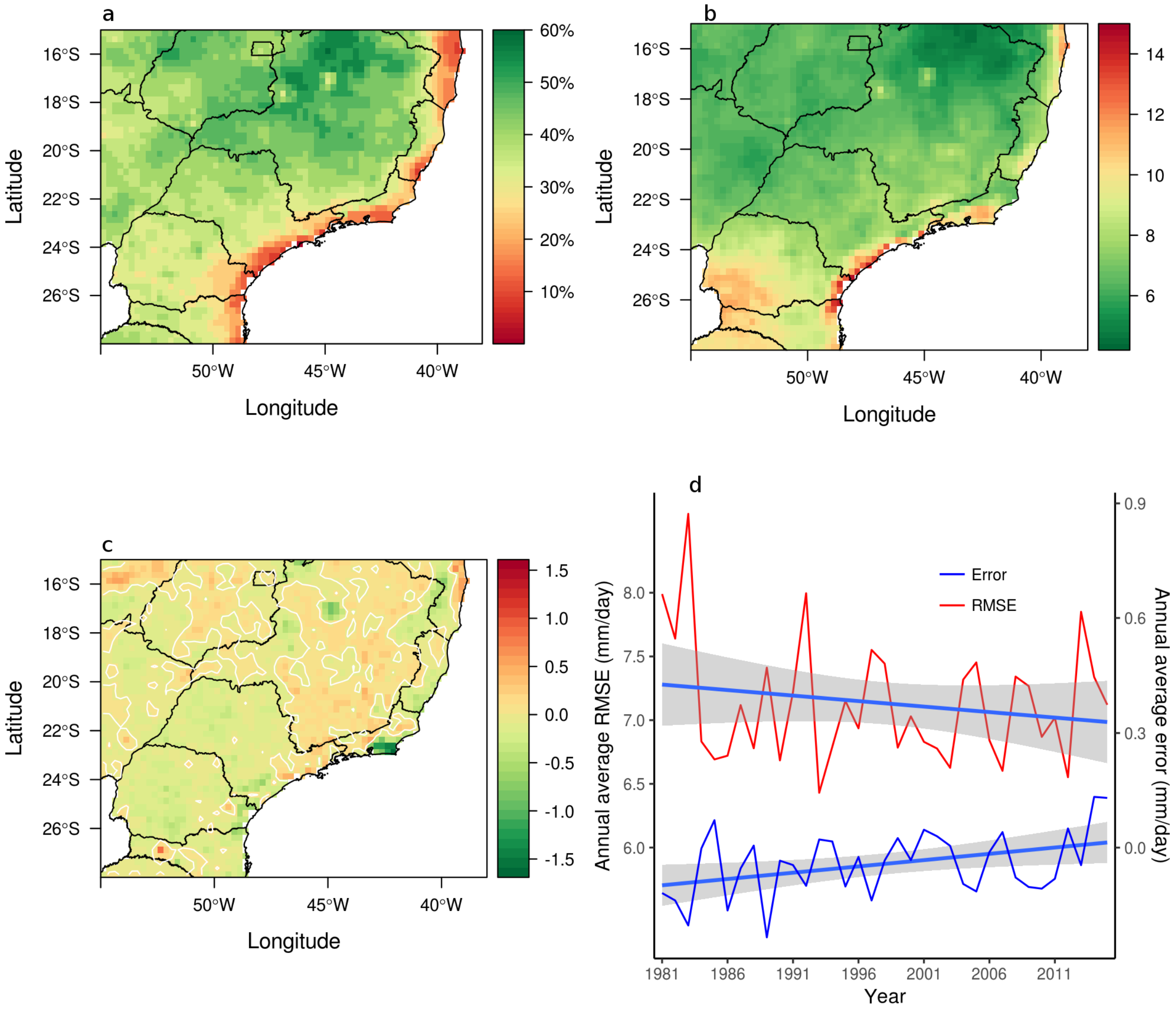


Figure 4.2: Evaluation of CHIRPS daily precipitation rate against METBR between 1981 and 2016: (a) Coefficient of determination (%); (b) RMSE (mm/day); (c) ME (mm/day), where the white contour line separates positive from negative values; (d) Annual ME and RMSE averaged in the same area displayed in “a”, “b” and “c” with the respective linear regressions and 95% confidence bands (shaded).

and its data have been extensively used and evaluated (Sugahara et al., 2012; Silva Dias et al., 2013; Perez and Silva Dias, 2017). On that account, CHIRPS was evaluated against daily precipitation from the IAG station rain gauge.

Figure 4.3 compares CHIRPS and IAG daily precipitation percentiles, considering CHIRPS nearest grid point to IAG station. Each point in the plot (known as quantile-quantile plot or Q-Q plot) corresponds to a percentile (from 0 to 1) and its respective value in the probability distribution of CHIRPS and IAG station daily precipitation. The black line is the 1:1 relationship. Except for a small overestimation, the distributions are very close together below the 0.9 percentile. For higher percentiles, CHIRPS considerably underestimates IAG precipitation rate and the probability distributions drift apart.

Figure 4.4 shows the confusion matrix comparing daily precipitation from CHIRPS and IAG station. The values are binned in 5 and 10 mm intervals. Each position in the matrix corresponds to the number of days in which the same precipitation interval was observed both by CHIRPS and IAG station. The lowest bin (0-5 mm) is the most frequent and has the highest agreement, while the others bins are substantially mismatched. This generally indicates that, despite CHIRPS high spatial resolution (approx. 5 km grid spacing) relative to other precipitation products, the spatial variability of rainfall is still significant inside a 25 km^2 area.

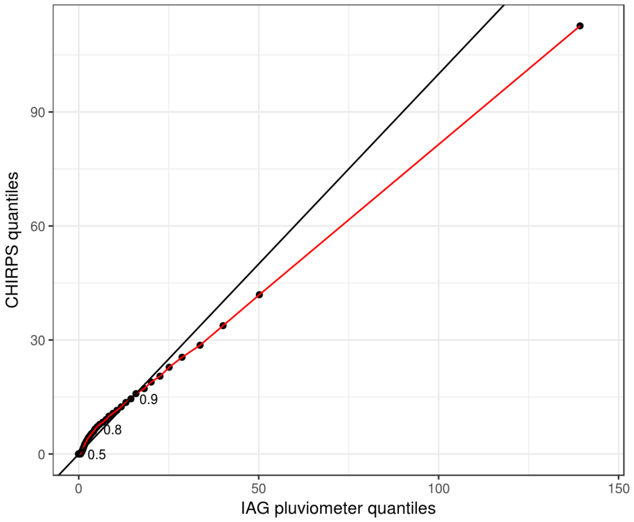


Figure 4.3: Q-Q plot comparing the percentiles of daily precipitation rates (mm/day) from CHIRPS and IAG meteorological station (1981-2016). The 0.5, 0.8 and 0.0 percentiles are identified and the black line illustrates the 1:1.

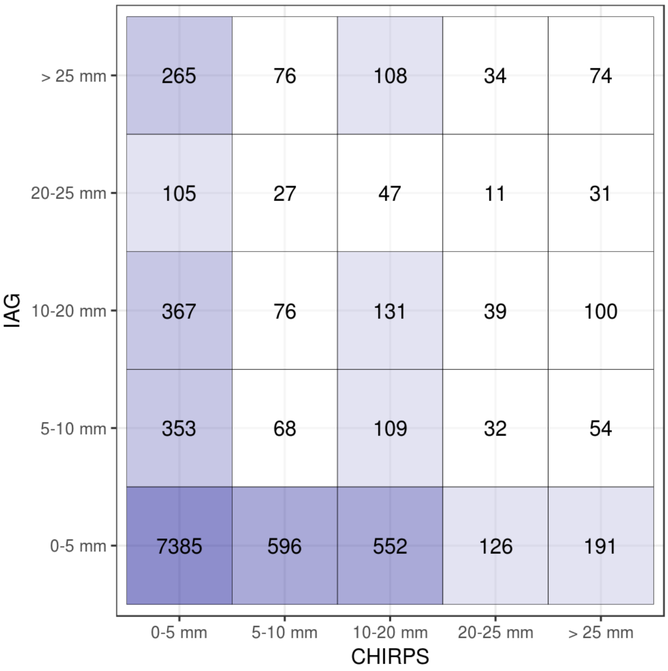


Figure 4.4: Confusion matrix of the daily accumulated precipitation (mm) from CHIRPS and IAG meteorological station station (1981-2016).

Chapter 5

Atmospheric patterns associated with extreme rainfall

The role of the expert in the development, evaluation and improvement of machine learning models is fundamental. The researcher should have a critical eye on the results and the rationale of the learning process to assess the physical coherence and validity of given solutions.

In principle, ANNs can learn solutions (i.e., the weights and biases of the network) that are not bounded by the physical constraints of the phenomena. In purely engineering applications, non-physical solutions are not a problem, as long as they output good results. In opposition to these applications, scientific studies are often concerned in exploiting statistical models to infer physical properties of nature. In that account, deep neural networks can be regarded as potential tools for understanding linear and non-linear phenomena.

It is proposed by Lin et al. (2017) that deep neural networks tend to reconstruct the generative process of physical phenomena. To defend this hypothesis, the authors argue that physical processes can often be described by low-order polynomials and obey some properties such as symmetry and locality. Following that idea, this study aims to employ deep learning not only as a predictive tool, but also as a framework to improve the understanding of rainfall-related processes.

This chapter highlights atmospheric patterns associated with extreme rainfall, comparing the meteorological variables during extreme events and the climatology. Particularly, precipitation events occurred in the city of São Paulo are considered. This analysis aims to underpin the assembly of an explanatory dataset for rainfall forecasting and to support the discussions in Section 5.5.

5.1 Methodology and data

ECMWF's (European Center for Medium-range Weather Forecast) ERA-Interim (Dee et al., 2011) reanalysis combines atmospheric observations with numerical modeling to provide a realistic description of the state of the atmosphere. The atmospheric variables used in this chapter were obtained from ERA-Interim between 1985 and 2015 in a 0.25°grid (Table 5.1).

Table 5.1 - Meteorological variables obtained from ERA-Interim between 1985 and 2015 used to compare the state of the atmosphere during extreme events with the climatology.

Variable name	Time (UTC)	Vertical level (hPa)
Specific humidity	12	1000 and 500
Temperature	12 and 18	1000
Horizontal wind	12 and 18	1000 and 500
Vertical wind	12 and 18	500
Mean sea level pressure	12 and 18	—
Sea surface temperature	12	—

The daily rainfall in São Paulo was calculated by averaging all CHIRPS grid points within the city boundaries. The extreme events were defined when the daily rainfall exceeded the 0.95 percentile. Subsequently, the mean atmospheric fields (from ERA-Interim) during extreme events (X_{ext}) were obtained and compared with the climatology[1] ($\bar{X}$). The Student's t-test[2] (Equation 5.1) was applied to verify whether the climatological average and the extreme events average were significantly different (p-value).

$$t = \frac{X_{ext} - \bar{X}}{\frac{s}{\sqrt{n}}} \tag{5.1}$$

Where n is the total of extreme events s is the extreme events standard deviation.

---

[1] Average of ERA-Interim data from 1985 to 2015, including the extreme set

[2] For the validity of the test, it is assumed that the sample means are normally distributed

The probability of obtaining a t-value $|t'|$ greater than $|t|$ was computed (p-value = $Pr(|t'| \geq |t|)$) and a significance level of 0.01 was established.

5.2 Results and discussion

5.2.1 Specific humidity

Specific humidity is defined by the ratio between the masses of water vapor and moist air. It is a key ingredient to the formation of clouds and rainfall. The moisture available for cloud processes can originate by local recycling or it can be transported from remote regions (Trenberth, 1999; Zhou and Yu, 2005).

Figures 5.1 and 5.2 show the specific humidity from ERA-Interim at the vertical levels of 500 hPa and 1000 hPa, respectively. They reveal that the average concentration of water vapor is higher during extreme precipitation events. Note that, as Figures 5.1 and 5.2 show no gray shading, there is 99% of confidence that the averages are different in all grid points.

The specific humidity, in Figure 5.2b, peaks near the coastline, highlighting the role of the Atlantic Ocean as a source of moisture. The sea breeze is a potential mechanism to link this remote source to extreme precipitation in the city of São Paulo. Perez and Silva Dias (2017) show that the sea breeze frontal passage has a significant contribution to the occurrence of extreme events in São Paulo.

5.2.2 Temperature

Horizontal temperature gradients are important sources of baroclinicity[3] and motion in the atmosphere. Their role in mesoscale circulations can be expressed through the Bjerkenes Circulation Theorem (Equation 5.2). It states that the line integral (in pressure coordinates) of the specific volume (α) along a closed path determines the variation of the absolute circulation (C_a) of the fluid enclosed by that path:

---

[3] Baroclinicity in the atmosphere is defined as a situation where the isosurfaces of density and pressure are not parallel.

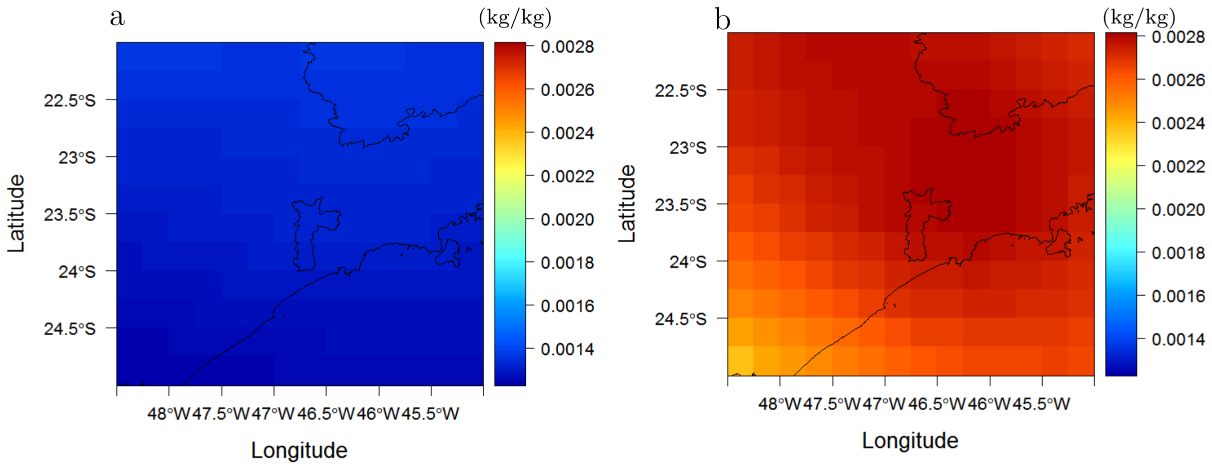


Figure 5.1: a) Specific humidity (kg/kg) climatological average at 12Z and vertical level of 500 hPa (1985-2015, ERA-Interim); b) Same as "a" but averaged on days in which extreme precipitation (> 95%) occurred in the city of São Paulo

$$\frac{dC_a}{dt} = -\oint \alpha dp \text{ , where } \alpha \equiv \frac{1}{\rho} \tag{5.2}$$

Considering the ideal gas law $\alpha = \frac{p}{RT_v}$, the variation of absolute circulation can be re-written in terms of virtual temperature[4] (T_v):

$$\frac{dC_a}{dt} = -\frac{1}{R}\oint \frac{p}{T_v}dp \tag{5.3}$$

Equation 5.3 states that the absolute circulation changes with time provided that temperature is not constant in an isobaric surface. For that reason, horizontal gradients of temperature are essential to the development of local circulations, such as mountain-valley circulation and land/sea breeze.

The importance of the temperature field in the development of clouds can be analyzed in terms of the buoyancy of an air parcel. Buoyancy, as stated by the Archimede's principle, is the vertical acceleration of an element of fluid caused by the density difference between the fluid parcel and the environment. Employing the ideal gas law, the vertical acceleration may be written in terms of the difference between the virtual temperatures of the air parcel

---

[4] Theoretical temperature in which the moist air should have to equal the density and volume of dry air.

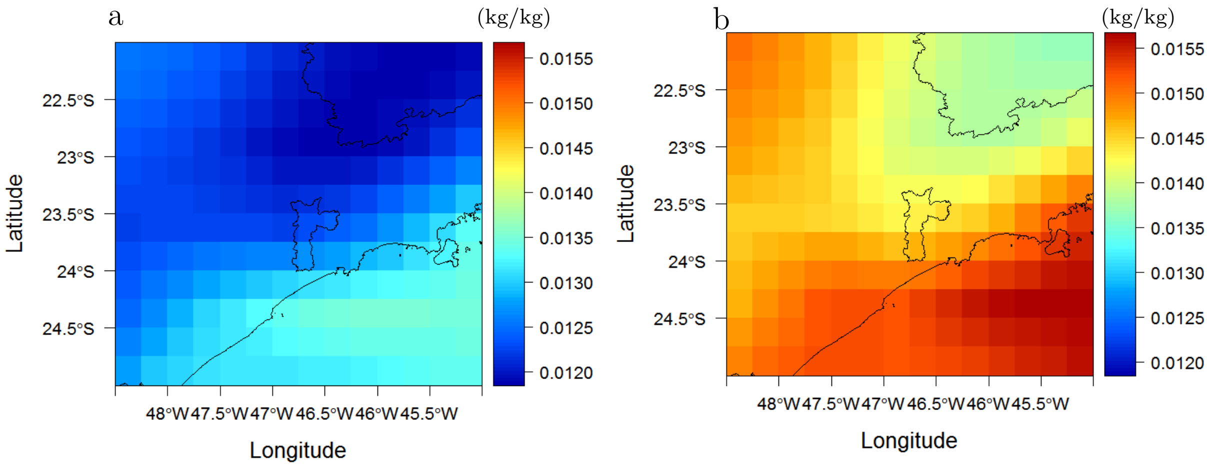


Figure 5.2: a) Specific humidity (kg/kg) climatological average at 12Z and vertical level of 1000 hPa (1985-2015, ERA-Interim); b) Same as "a" but averaged on days in which extreme precipitation (> 95%) occurred in the city of S ao Paulo

(T_v) and the environment (T_v'):

$$\frac{d^2z}{dt^2} = g\frac{T_v' - T_v}{T_v} \tag{5.4}$$

The proximity with the Atlantic Ocean, significant topographic features (Figure 1.3) and contrasting land uses (Figure 1.4) in ESP are ingredients to intense temperature gradients. Figures 5.3 (averaged at 09:00 local time) and 5.4 (averaged at 15:00 local time) compare the climatological temperature with the mean temperature during extreme precipitation days. In the morning (Figure 5.3), warmer temperatures favor the occurrence of extreme events. Oppositely, in the afternoon (Figure 5.4), the average temperature in the continental area is reduced due to the cooling effect of precipitation clouds. In the oceanic area, the air temperature is higher indicating higher sea surface temperature. Note that, in Figures 5.4a and 5.4b, a number of the grid points are shaded, indicating they are not statistically different (t-test).

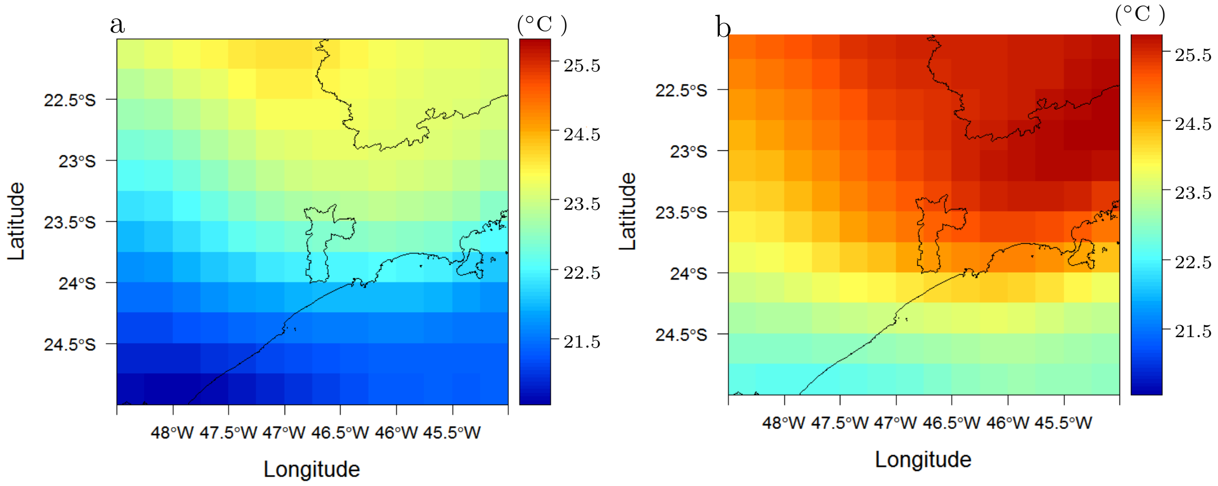


Figure 5.3: a) Temperature (°C) climatological average at 12Z and vertical level of 1000 hPa (1985-2015, ERA-Interim); b) Same as "a" but averaged on days in which extreme precipitation (> 95%) occurred in the city of São Paulo

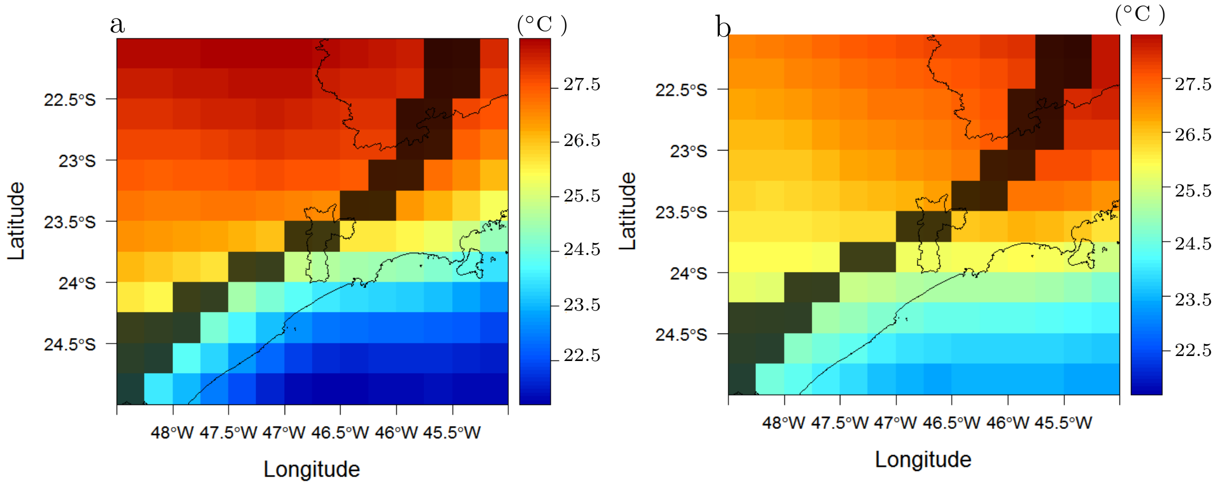


Figure 5.4: a) Temperature (°C) climatological average at 18Z and vertical level of 1000 hPa (1985-2015, ERA-Interim); b) Same as "a" but averaged on days in which extreme precipitation (> 95%) occurred in the city of São Paulo

5.2.3 Wind

The wind field influence on the development of clouds and precipitation ranges from turbulent to synoptic scales. In the eddy scale, the vertical component of the wind is responsible for transporting the water vapor that supplies the formation cloud droplets. In a larger scale perspective, a subsidence region[5] may suppress the ascending currents and prevent the development of clouds, even if local conditions are favorable. The mesoscale and large scale horizontal winds also play an important role transporting properties such as temperature, water vapor and momentum.

Figures 5.5 and 5.6 show the horizontal wind at 1000 hPa from ERA-Interim. The strong northern wind component (5.5b) combined with the temperature field (5.3b) reveals a positive temperature advection towards São Paulo in the morning of extreme rainfall days. Figure 5.6, on the other hand, shows that neither the temperature advection or the strength of the sea breeze are determinant factors in the afternoon. Instead, a clear pattern of convergence between the sea breeze and northwesterly winds can be seen along the coastline.

Figure 5.7 shows the vertical wind speed in pressure coordinates at the level of 500 hPa. The level of 500 hPa situates roughly at 5000 meters of altitude and it is a proxy of the large-scale influence in the ascension of air parcels. Figure 5.7b shows that the vertical speed have a greater magnitude [6] in extreme precipitation days, facilitating the ascension of air parcels from the planetary boundary layer.

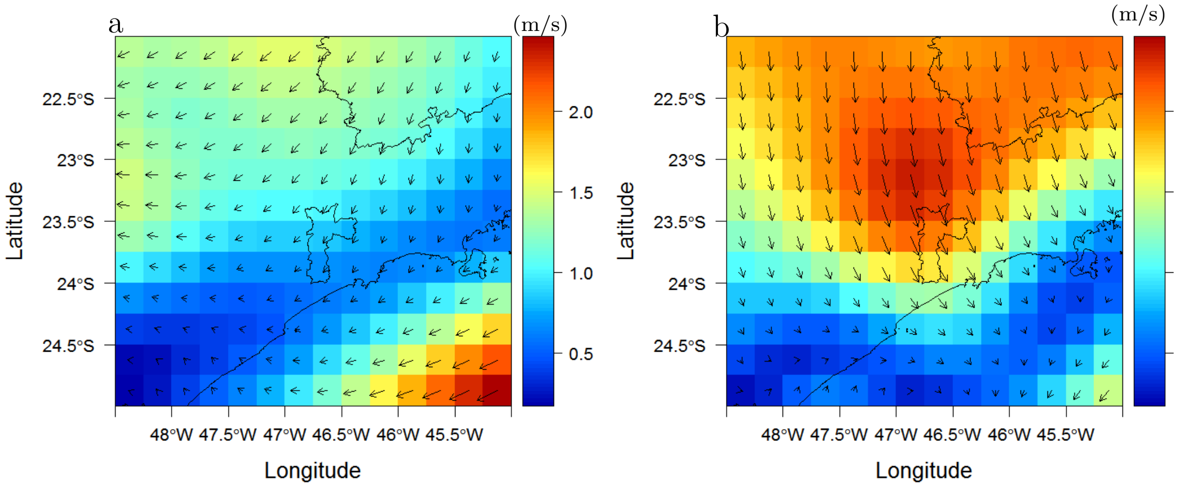


Figure 5.5: Horizontal wind vector field with shaded wind speed (m/s), where (a) is the climatological average (1985-2015, Era-Interim) at 12Z and vertical level of 1000 hPa and (b) is the same as "a" but averaged on days in which extreme precipitation was observed in the city of São Paulo.

---

[5] Region with negative (positive) vertical wind speed in Cartesian (pressure) coordinates

[6] Note that pressure coordinates increases towards the ground

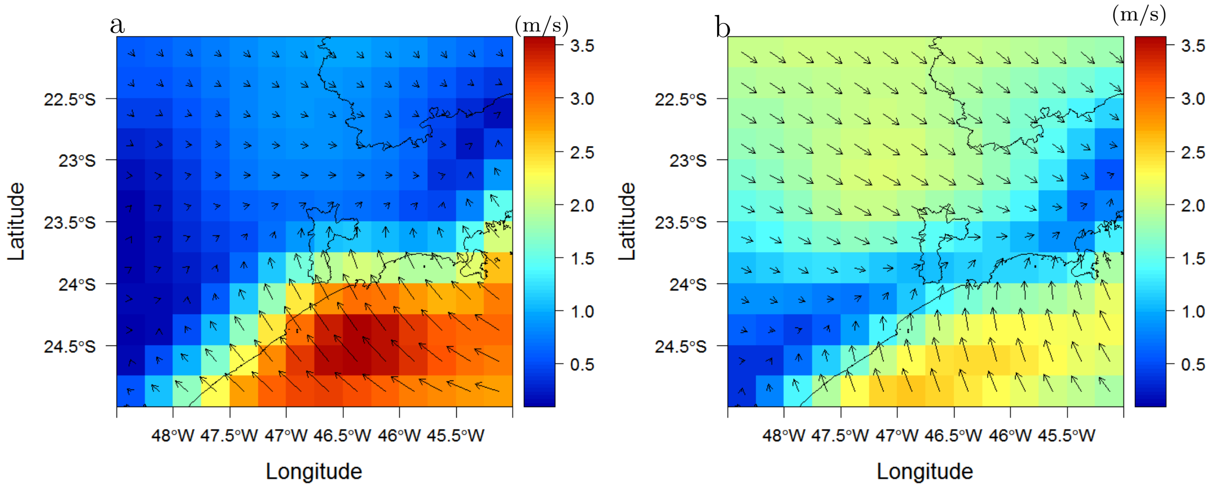


Figure 5.6: Horizontal wind vector field with shaded wind speed (m/s), where (a) is the climatological average (1985-2015, Era-Interim) at 18Z and vertical level of 1000 hPa and (b) is the same as "a" but averaged on days in which extreme precipitation was observed in the city of São Paulo.

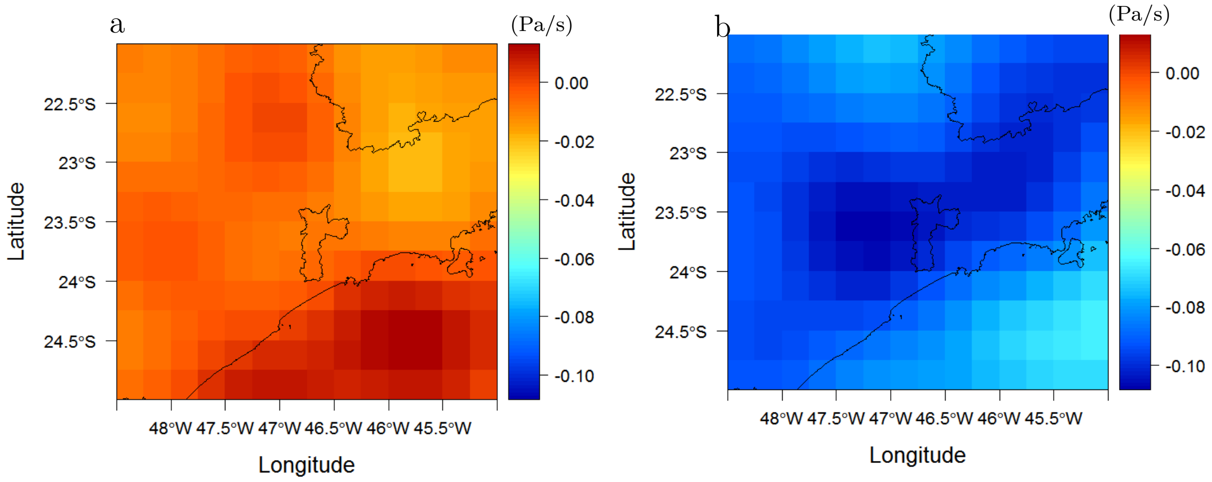


Figure 5.7: Vertical wind speed (Pa/s), where (a) is the climatological average (1985-2015, Era-Interim) at 12Z and vertical level of 500 hPa and (b) is the same as "a" but averaged on days in which extreme precipitation was observed in the city of São Paulo.

Chapter 6

Precipitation forecast in the city of São Paulo

This chapter presents and evaluates a machine learning framework to forecast daily total precipitation in the city of São Paulo. Two situations were explored: an idealized and a realistic. The ERA-Interim reanalysis represents an ideal atmospheric model with minimum uncertainty and the GEFS reforecast represents an operational model available to the weather forecaster. Both of these situations are compared with observations.

6.1 Data

As discussed in Chapter 2, two explanatory datasets were explored: the GEFS reforecast Version 2 (Hamill et al., 2013) and the ERA-Interim reanalysis. The GEFS reforecast is a historical dataset of ensemble weather forecasts produced with a recent version of NOAA's Global Forecast System model. Only the control member of the ensemble was used (hereafter, GFS). Table 6.1 lists the variables chosen as input to the DNNs and Figure 6.1 shows GFS (red) and ERA-Interim (blue) horizontal meshes in the area where the variables where extracted.

ERA-Interim data was not inputed directly in the DNNs. As shown in Figure 6.1, there are 384 grid points per vertical level and time for each variable. Using all ERA-Interim grid points would lead to a very large number of explanatory variables. Dealing with such large number of variables would hinder the comprehensibility of results and significantly increases the computational efforts. The next chapter discusses the approach undertaken to tackle this dimensionality issue.

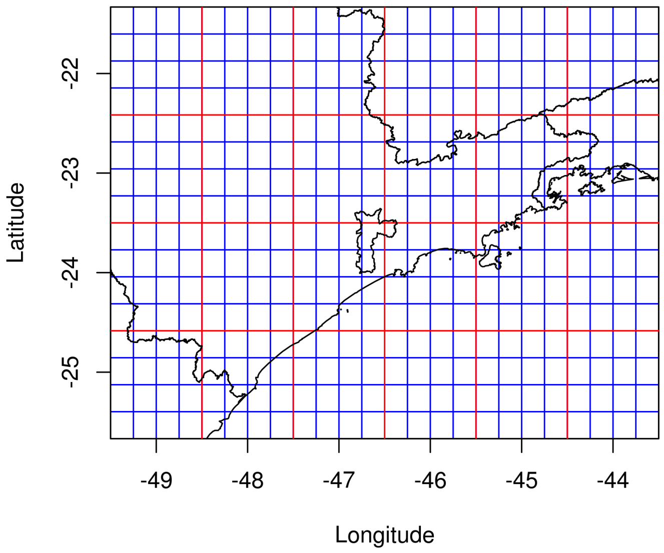


Figure 6.1: Horizontal meshes of ERA-Interim (blue) and GFS (red)

6.2 Linear dimensionality reduction of the ERA-Interim dataset

Considering non-turbulent scales, most atmospheric variables present a relatively smooth and continuous spatial distribution (with some exceptions such as precipitation). The diurnal cycle of such variables is often smooth as well (again, above the time scale of turbulent vortices). Some exceptions may occur in the case of a subtle humid air mass intrusion in a normally dry area by the means of sea breeze, for example. In general, it is reasonable to expect that the explanatory variables listed in Table 6.1 can be, without major losses of information, projected into a lower-dimensional space.

Principal Component Analysis (PCA) is a linear transformation that projects a n-dimensional dataset X into a n-dimensional orthogonal basis in which the variables of the rotated dataset Z are linearly uncorrelated (Figure 6.2). This transformation can be achieved by eigendecomposition of the correlation matrix (Equation 6.1). The eigenvectors of the decomposition correspond to the columns of matrix W (principal components). The eigenvalues $\alpha_{1..n}$ determine the fraction of the linear variance explained by each principal component (PC): $var_i = \frac{\alpha_i}{\sum_{i=1}^{n} \alpha_i}$.

Table 6.1 - Meteorological variables obtained from ERA-Interim, GFS and climatic indices gathered between 1985 and 2015 to assemble the explanatory datasets.

Variable name and abbreviation	Time (UTC)	Vertical level (hPa)	Source
Specific humidity (q)	12 and 18	1000, 500 and 925	GFS and ERA-Interim
Temperature (T)	12 and 18	1000, 500 and 200	GFS and ERA-Interim
Horizontal wind (u and v)	12 and 18	925, 500 and 200	GFS and ERA-Interim
Vertical wind (w)	12 and 18	500	GFS and ERA-Interim
Geopotential height (z)	12 and 18	500 and 200	GFS and ERA-Interim
Mean sea level pressure (P)	12 and 18	—	GFS and ERA-Interim
Total column water (tcw)	12 and 18	—	GFS and ERA-Interim
SOI	(Monthly)	—	NOAA Climate-CPC
AMO	(Monthly)	—	NOAA-PSD
PDO	(Monthly)	—	JISAO

$$Z_{n,m} = X_{n,m} W_{m,m} \tag{6.1}$$

Figure 6.3 shows the fraction of the variance explained by the 20 highest ranked [1] principal components in the eigendecomposition of the original ERA-Interim dataset. The dimensionality reduction is reasonably efficient at upper atmospheric levels (200 hPa). At medium and lower atmospheric levels (500, 925 and 1000 hPa and mean sea level), the

---

[1] Sorted by explained variance

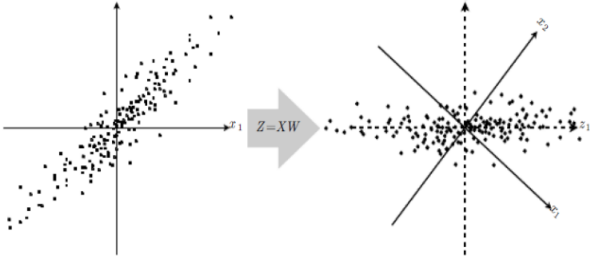


Figure 6.2: 2-dimensional example of PCA rotation from X to Z. Extracted from Goodfellow et al. (2016)

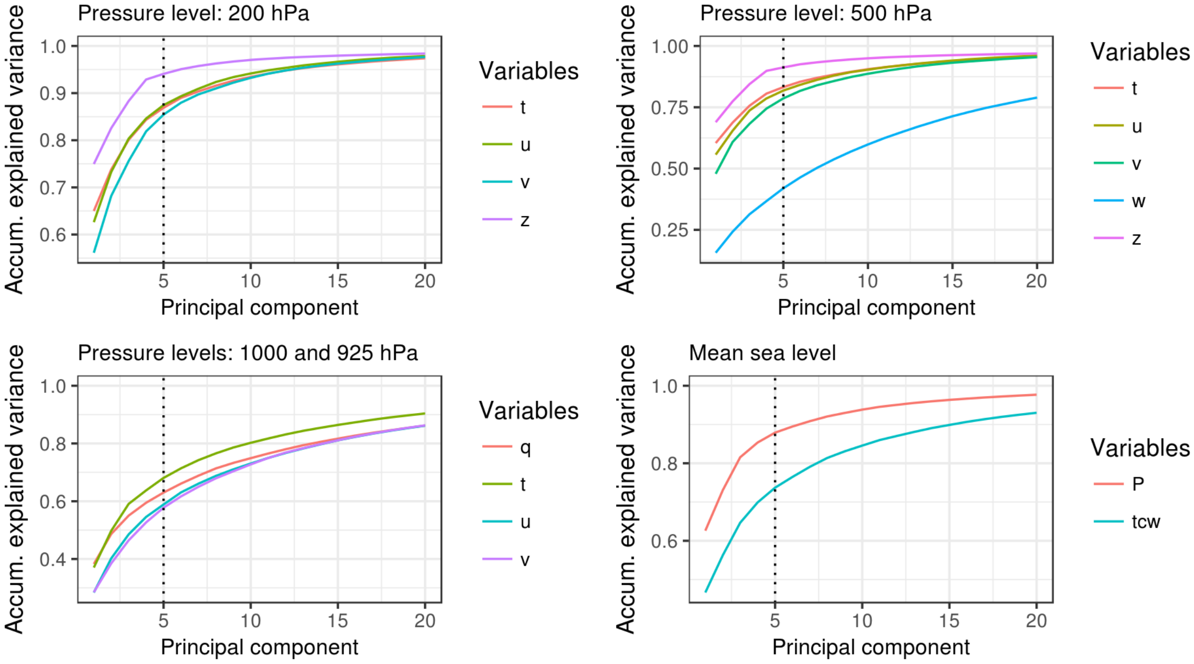


Figure 6.3: Accumulated variance explained by the 20 highest ranked principal components by the eigendecomposition of the original ERA-Interim dataset. The dotted vertical line marks the fifth principal component.

spatial distribution of the explanatory variables becomes increasingly non-linear, causing the PCA dimensionality reduction to be less efficient. Nevertheless, except for the vertical wind component (w), five principal components suffice to preserve more than 50% of the original dataset linear variance. For this reason, the lower dimension dataset was constituted with the projection of the original data in the five highest ranked PCs.

6.3 Deep neural networks architecture and training

6.3.1 Comment on recurrent and convolutional neural networks

There are several neural network architectures available to build discriminator or predictive models depending on the nature of objective function. If the objective function is a time series where the current time depends upon past states of the model, recurrent neural networks (RNNs) are the prevailing technique (Graves et al., 2013). RNNs are frequently employed for speech recognition, text translation, handwriting recognition, among others. In the field of image recognition, the predominant approach is to employ convolutional hidden-layers in the neural network. These convolutional neural networks (CNNs) explore some common properties found in images, such as locality: each pixel tend to be more related to its vicinity. The convolutional layers apply small sized (e.g. 4x4) kernels through the image channels (e.g., red, green and blue) aiming to recognize features such as edges and particular shapes. This convolution "trick" dramatically reduces the number of parameters in the network, allowing processing larger images that would be impracticable in fully connected multi-layer perceptrons.

Even though both RNNs and CNNs conceptually have desirable properties for a predictive model, there is a literature vacuum concerning how these classes of neural networks behave in the context of meteorological processes. RNNs might have been an appropriate choice to forecast the hourly precipitation rate because, in this time scale, the model would take advantage of temporal sequences such as the life cycle of a storm or the propagation of sea breeze fronts. In the case of 24-hour precipitation forecast, however, one cannot assume such temporal dependence exists, except under particular circumstances where a persistent precipitation system, such as the SACZ, is active.

Nevertheless, the ability of CNNs to identify spatial patterns and handle large 2-dimensional grids should not be ignored. Recently, Cao et al. (2017) developed a network combining convolutional and recurrent layers to predict the hourly wind field in the United States. The network was trained to use only the past wind field as input, treating its evolution as a sequence that depends solely upon itself. This approach is valid for a limited time span, since the wind field evolves in function of dynamical forces such as the pressure gradient. Further research must be made in the direction of understanding how convolutional and recursive layers should be employed to predict a meteorological variable (such as precipitation) considering the complete atmospheric state.

6.3.2 Deep multi-layer perceptron architecture

A fully-connected deep multi-layer perceptron (Figure 2.2) was chosen as predictive model for the 24-hour rainfall forecast. It has been found to be an appropriate method because it does not assume any spatial or temporal relationship between the inputs. Moreover, since the dimensionality of the input dataset was reduced with PCA, the size of the weight matrices did not impose any limitations.

The input layer consists of all the variables listed in Table 6.1 (projected in five PCs in the case of ERA-Interim), followed by four hidden-layers of 200 nodes each and a single node on the output layer. Because of the nature of ReLU activation function (Equation 2.3), the precise number of hidden nodes is not a sensible parameter in the model – i.e., similar results could be achieved varying the number of nodes per layer from approximately 150 to 250. Glorot et al. (2011) argue that ReLU gives to the network flexibility to control the effective number of nodes, allowing it to increase or decrease the dimensionality of the representation depending on the amount of information/complexity for each set of explanatory variables. To contextualize this argument with a meteorological example, picture the task of predicting two different types of rainfall systems: local convective storms and large-scale organized precipitation (e.g, cold front). During the training phase, the neural network might learn that, to forecast local convective storms, a more complex representation is required (i.e., more active neurons per layer) than to forecast large-scale organized precipitation.

6.3.3 Training of the deep multi-layer perceptron

This subsection will introduce the algorithms employed during the training phase of the multi-layer perceptron. Initially, the explanatory datasets (Era-Interim + climate indices and GFS + climate indices) and the target dataset (24-hour precipitation from CHIRPS) were divided in three random sets:

1. Training set (50% of the total days) : the data effectively used to train the neural networks;
2. Validation set (20% of the total days): the data used during the training phase to assure that no overfitting occurred;
3. Test set (30% of the total days): the data employed during the test phase, completely ignored during the training phase.

The difference between the validation set and the test set may seem unclear, but the distinction is important. During the training phase, the performance of the neural networks is evaluated against the validation set. This procedure is useful to check if the model is overfitting the training set. The test set, on the other hand, will not be considered during the training phase at all, working as new unseen data.

The first step of the training phase is to feed the training set in the input layer of the neural network and iterate through the net until the output layer. This procedure is called "forward-pass". Algorithm 1 describes the forward-pass algorithm combined with dropout (discussed in Section 2).

Algorithm 1 MLP forward pass with dropout

```
    Given a randomly initialized MLP with L ($0 < l \leq L$) integer layers and I ($0 < i \leq I$)
    nodes per hidden-layer, let:
    $z^l$ be the input vector for layer l;
    $y^l$ be the output vector of layer l;
    $w^l$ be the weight matrix at layer l;
    $b^l$ be the bias at layer l;
    f be an activation function and

    let $r^l$ be a random binary vector such that each element has probability p of being 1
    or zero otherwise.

for $l$ from 1 to L-1
do
    for $i$ from 1 to I
    do
        $z_i^{l+1} = w_i^{l+1} \cdot (y^l \cdot r^l) + b_i^{l+1}$
        $y_i^{l+1} = f(z_i^{l+1})$
```

After the computation of the the network output y^L, a loss function should be defined to evaluate the prediction against the corresponding observed value y_{ref}. Considering that the quantitative precipitation forecast is a regression task, the mean squared error (MSE, Equation 6.2) was chosen as an appropriate loss function. It is defined by:

$$MSE = \sum_{i}^{n} \frac{(y^L - y_{ref})^2}{n} \tag{6.2}$$

The learning process consists in minimizing the loss function MSE. Therefore, the gradient of MSE with respect to each weight $w_{i,l}$ must be computed. The procedure of computing the output layer gradient and propagating it through the previous layers is called backpropagation. A brief intuition of the backpropagation algorithm is given below (a detailed discussion can be found in Bishop (1995)).

To support the intuition of backpropagation, Figure 6.4 shows the final two layers ($L-1$ and L) of a multi-layer perceptron. The gradient of MSE with respect to the output y^L of layer L is given by equation 6.3.

$$\frac{dMSE}{dy^L} = 2(y^L - y_{ref}) \tag{6.3}$$

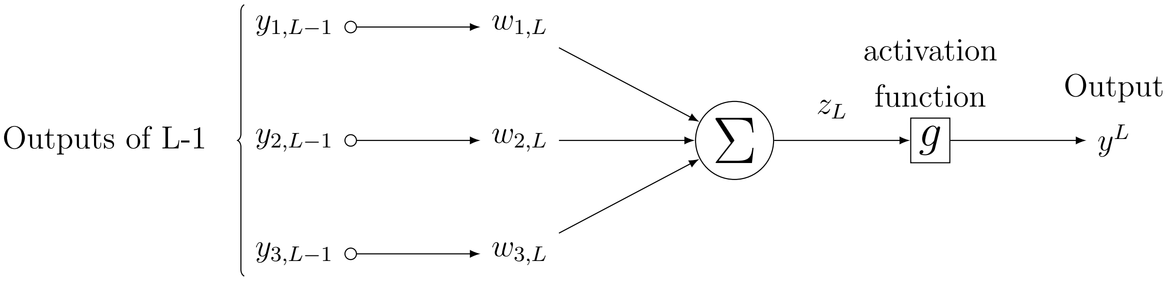


Figure 6.4: Diagram of the last 2 layers of a MLP to support the derivation of the backpropagation algorithm.

The chain-rule of derivatives can be used to obtain the gradient of the loss function with respect to the input z^L of layer L:

$$\frac{dMSE}{dz^L} = \frac{dMSE}{dy^L}\frac{dy^L}{dz^L} \tag{6.4}$$

Similarly, the chain-rule can be applied to compute the gradient of MSE with respect to the weights $w_{1,L}, w_{2,L}$ and $w_{3,L}$:

$$\frac{dMSE}{dw_{i,L}} = \frac{dMSE}{dz^L}\frac{dz^L}{dw_{i,L}} \tag{6.5}$$

Recalling that $z^L = \sum_{i=1}^{3} w_{i,L} y_i$, the term $\dfrac{dz^L}{dw_{iLj}}$ from Equation 6.5 simply equals to $y_{i,L-1}$. Using Equations 6.3 and 6.4, Equation 6.5 becomes:

$$\frac{dMSE}{dw_{i,L}} = 2(y^L - y_{ref})\frac{dy^L}{dz^L} y_{i,L-1} \tag{6.6}$$

The last remaining term to expand is the derivative of the output y^L of layer L with respect to the input z^L. Recalling that $y^L = g(z^L)$, where g is the activation function, it is only required to derive g with respect to z^L.

Iterating this method up to the input layer, it is possible to compute the gradient $g_{i,l}$ of the loss function with respect to the weights. Subsequently, the weights must be updated in order to minimize the MSE. There are several first-order methods available to perform this update. The most popular methods are based on gradient descent, i.e., updating the weights in the opposite direction of gradient g (Equation 6.7).

$$w_{i,l}^{t+1} = w_{i,l}^{t} - \eta g_{i,l} \tag{6.7}$$

The hyper-parameter η is called "learning-rate". It determines the size of the step taken in the opposite direction of the gradient g. The choice of a suitable learning-rate is determinant for the convergence and computational speed of the training phase. A large η may cause the error to oscillate around a local minimum while a small η may lead to a prohibitively slow convergence. The algorithm ADADELTA (Zeiler, 2012) was employed to alleviate the task of choosing an appropriate η. It calculates an adaptive learning-rate based on the running average of the gradient in the previous iterations. Meaning that the convergence accelerates if the gradient points in the same direction for several iterations and the convergence slows down near the local minimum.

6.4 *Precipitation forecast in the city of São Paulo*

To evaluate the 24-hour precipitation forecast in the city of São Paulo, the Hit Rate (HR) and False Alarm Rate (FR) (Equation 6.8) were computed on the test set. HR and FR are evaluation metrics for binary predictions and are calculated in function of the terms a, b, c and d (Equation 6.8). The term a corresponds to the number of times that a precipitation event above a certain threshold was both observed and forecasted. The term b corresponds to the number of forecasted events above a certain threshold that were not observed. The term c corresponds to the number of unpredicted events above a certain threshold. The term d is the number of times that a precipitation event above a certain threshold was neither predicted or observed.

The thresholds defined to compute HR and FR are based on the quantiles of the statistical distribution of the daily precipitation rate in the city of São Paulo. The equivalence between these quantiles and the precipitation rate in milliliters per day is given in Table 6.2.

Table 6.2 - Quantiles of the daily precipitation rate (mm/day) from CHIRPS in São Paulo.

Quantiles	Prec. rate (mm/day)
0.6	0
0.7	2.2
0.8	7.5
0.9	16.2
0.95	24.6

$$HR = \frac{a}{a+c} \qquad \text{and} \qquad FR = \frac{b}{b+d} \tag{6.8}$$

Figure 6.5 shows the HR and FR of the 24-hour precipitation forecast in São Paulo with the ERA-Interim reanalysis as explanatory set. The HR and FR were computed on 50 random subsets of the validation set with 500 observations each. The boxplot shows the first (25%) and third (75%) quartiles as well as the median and outliers. The median HR is above 50% for quantiles below 0.9 and the median FR is below 25% for all quantiles. The FR variance is noticeably smaller than the HR, meaning that it is very unlikely that a high FR is observed in any random set of predictions. As a complementary evaluation, Figure 6.6 shows the confusion matrix of the DNNs predictions compared with CHIRPS observed precipitation.

Similarly to Figure 6.5, Figure 6.7 shows the HR and FR of the DNN forecast in São

Paulo, but trained with the GFS explanatory dataset. The results are very similar to Figure 6.5, except for quantiles above 0.9. This difference highlights the importance of a quality input dataset to forecast extreme precipitation events. Nevertheless, Figure 6.7 shows very promising results for the employment of the methodology in a realistic weather forecasting situation.

To highlight the improvements provided by the deep learning methodology, the same evaluation procedure was employed with the GFS precipitation output. Figure 6.8 shows the HR and FR of the GFS 24-hour precipitation forecast. Comparing with Figure 6.7, it is unequivocal that the neural networks predictions are much more realistic than the GFS precipitation output. It is relevant to recall that neither the precipitation outputs from GFS or ERA-Interim were used as input to the DNNs (see the explanatory variables in Table 6.1).

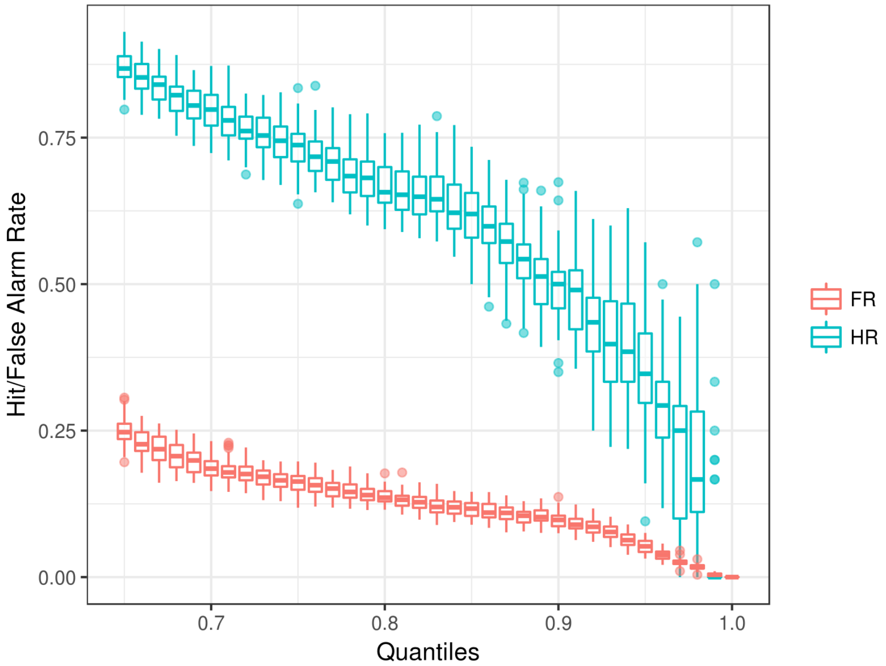
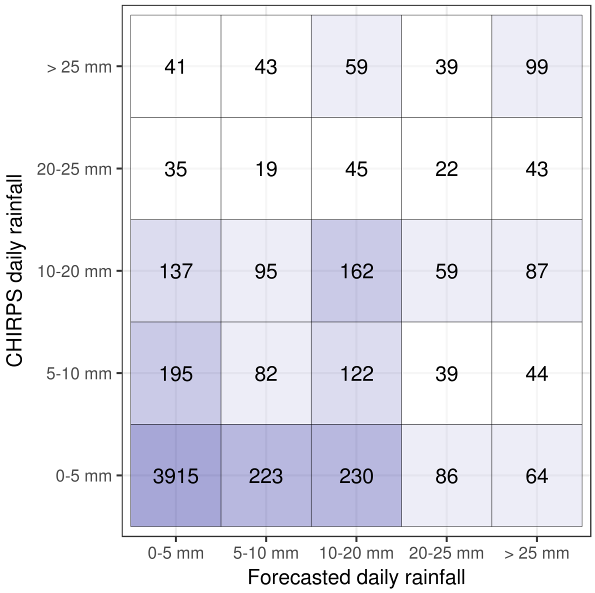


Figure 6.5: Hit Rate (HR) and False Alarm Rate (FR) of the DNN 24-hour precipitation forecast in the city of São Paulo. In this case, the DNN was trained using ERA-Interim explanatory data.



CHIRPS daily rainfall \ Forecasted daily rainfall	0-5 mm	5-10 mm	10-20 mm	20-25 mm	> 25 mm
> 25 mm	41	43	59	39	99
20-25 mm	35	19	45	22	43
10-20 mm	137	95	162	59	87
5-10 mm	195	82	122	39	44
0-5 mm	3915	223	230	86	64

Figure 6.6: Confusion matrix of the DNN 24-hour precipitation forecast in the city of São Paulo. In this case, the DNN was trained using ERA-Interim explanatory data.

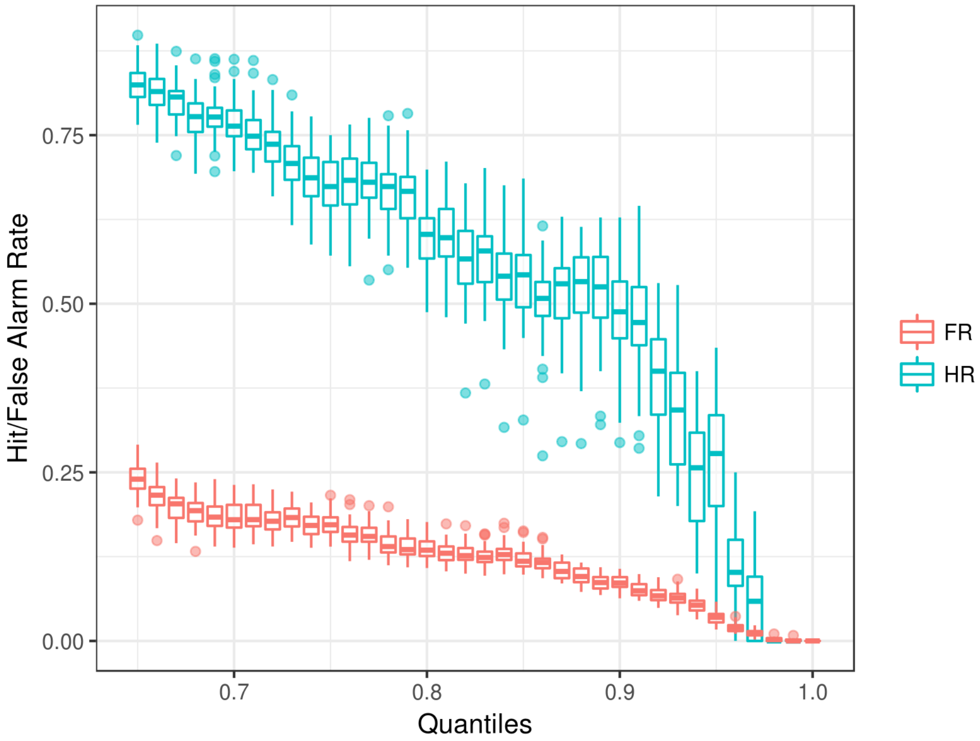


Figure 6.7: Hit Rate (HR) and False Alarm Rate (FR) of the DNN 24-hour precipitation forecast in the city of São Paulo. In this case, the DNN was trained using GFS explanatory data.

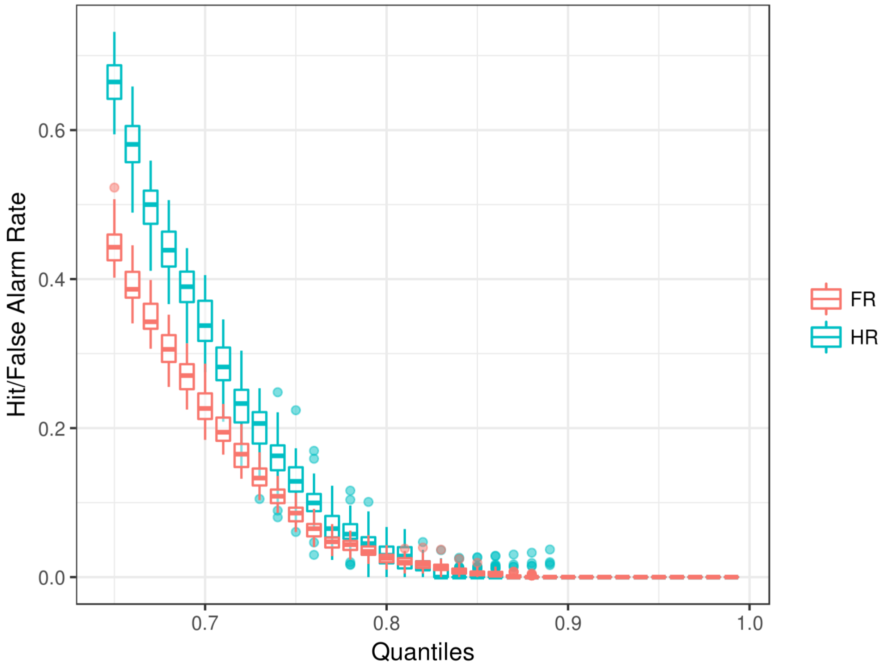


Figure 6.8: Hit Rate (HR) and False Alarm Rate (FR) of the GFS 24-hour precipitation forecast in the city of São Paulo.

6.5 Evaluating the predictors importance

During the learning stage, the stochastic gradient descent updates the weights matrix of the network to minimize the loss function. The weights of the first layer store the strength of the connections (also known as synapses) between the input neurons and the next layer. In a shallow network, i.e., a single-layer ANN, the magnitude of the weights are the direct measure of how much each input influences the output. In a fully-connected multi-layer perceptron, however, such evaluation is not as straightforward. Complex combinations between the inputs are allowed, meaning that, even if an input alone is not a relevant predictor, it may become relevant when combined with a certain configuration of inputs.

An empirical procedure was employed to verify the influence of each input variable alone in the network output. The input layer of the DNN (trained with the ERA-Interim dataset) was stimulated with a one-hot vector [2] for each input variable. The scaled outputs of the one-hot stimulations were defined as the relative importance of each input variable. Table 6.3 shows the relative importances and the meteorological interpretation of the 5

---

[2] One-hot vector is a vector containing the value one in a certain position and zero in the others.

most relevant input variables and Figures 6.9 to 6.13 display the rotation matrix of the principal components listed in Table 6.3.

It may be surprising to note the absence of some variables such as specific humidity and temperature in Table 6.3. Nonetheless, this procedure ignores the interactions between the input variables. Instead, it classifies the variables that can – alone – influence the precipitation forecast. It means, for instance, that a humid atmosphere is not sufficient to produce rainfall without relying on other conditions. Large-scale dynamical mechanisms, on the other hand, seem to have more ability of controlling rainfall alone.

Table 6.3 shows that the first principal component of the vertical wind at 500 hPa is the most relevant predictor. It express the role of the large-scale vertical flow in controlling local convection. This pattern was identified in Figure 5.7b. The second most relevant predictor is the second principal component of the meridional wind at 925 hPa. The pattern revealed in Figure 6.10 can be associated with cold front passage in São Paulo. The third and fourth most relevant predictors are also related with large-scale dynamics: Figures 6.11 and 6.12 reveal the geopotential trough and the high-level jet. The fifth most relevant predictor is the third principal component of the zonal wind at 925 hPa. Figure 6.13 displays a pattern similar to Figure 5.6. This pattern can be associated to a mesoscale low-level convergence between the sea-breeze and northwesterly winds.

The interactions between two input variables were briefly analyzed with two-hot vectors [3] stimulation. The most relevant combination is the first principal component of the 500 hPa vertical wind with the third principal component of specific humidity, highlighting the combination of ascending currents and water vapor availability. The importance of the near-surface temperature is highlighted by the second most relevant combination of predictors: the second principal component of the geopotential height at 500 hPa with the third principal component of temperature at 1000 hPa. This combination identifies the cold front passage associating the geopotential trough with horizontal temperature gradients.

---

[3] The two-hot vector contains the value one in two positions and zero in the others.

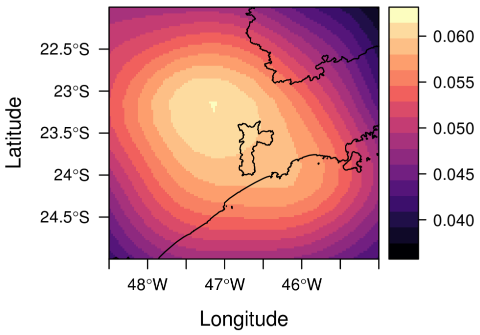


Figure 6.9: First principal component of the vertical wind at 500 hPa.

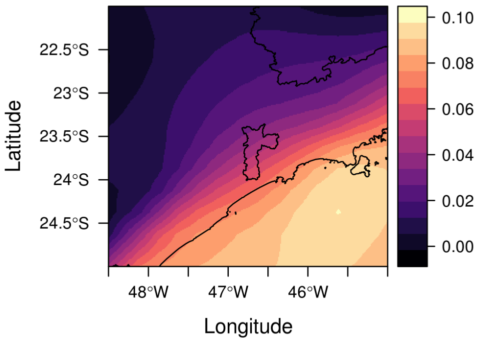


Figure 6.10: Second principal component of the meridional wind at 925 hPa.

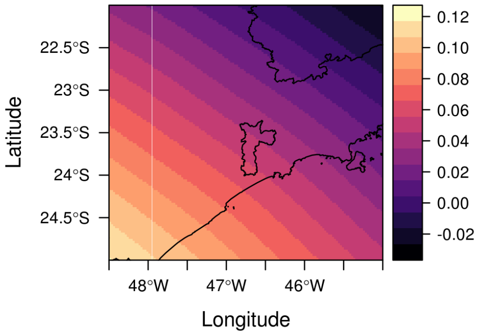


Figure 6.11: Second principal component of the geopotential height at 500 hPa.

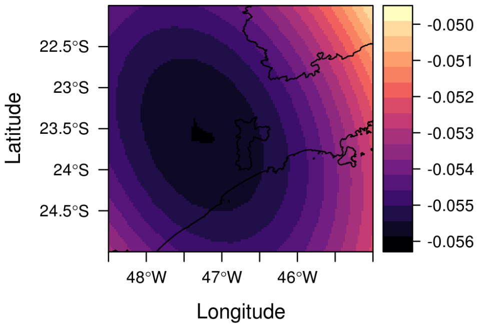


Figure 6.12: First principal component of the meridional wind at 200 hPa

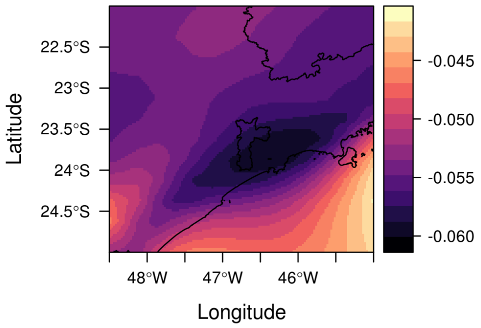


Figure 6.13: Third principal component of the zonal wind at 925 hPa.

Figures 5.9 to 5.13: Spatial patterns of the most relevant explanatory variables ranked in Table 6.3.

Table 6.3 - Relative importance and meteorological interpretation of the five most relevant predictors.

Stimulated variable	Level (hPa)	Princ. Comp.	Relative importance	Meteorological interpretation
w	500	1	1	Medium-level air ascension
v	925	2	0.86	Cold-front passage
z	500	2	0.78	Geopotential trough
v	200	1	0.71	High-level jet
u	925	3	0.67	Sea-breeze and low-level convergence

Chapter 7

Precipitation forecast in the East region of the State of São Paulo

This chapter comprises the methodology and results of the spatial precipitation forecast in ESP. The neural networks architecture and training procedure is similar to Chapter 5. The main methodological difference lies on how to achieve a forecast for hundreds of grid points. Section 6.1. discusses the undertaken approach to reduce the dimensionality of the precipitation field in order to facilitate the spatial precipitation forecast.

Regarding the explanatory dataset, the procedure is similar to Chapter 5. It consists of the ERA-Interim variables listed in Table 6.1 projected on the five first principal components. Figure 7.1a shows the spatial grid where the explanatory variables were extracted and Figure 7.1b shows the grid where CHIRPS precipitation data was extracted.

7.1 Non-linear representation of the precipitation dataset

The precipitation mesh in the study area (Figure 7.1b) consists of 260 grid points. Training a single neural network for each point is feasible, yet computationally expensive. In the case of a larger study area, a single network for each point would become impracticable. For this reason, a more general approach was chosen. Similarly to Section 5.2., a dimensionality reduction technique was employed.

The task of reducing the dimensionality of the precipitation field, however, may not be as simple as other atmospheric variables. Figure 6.3 shows that five principal components

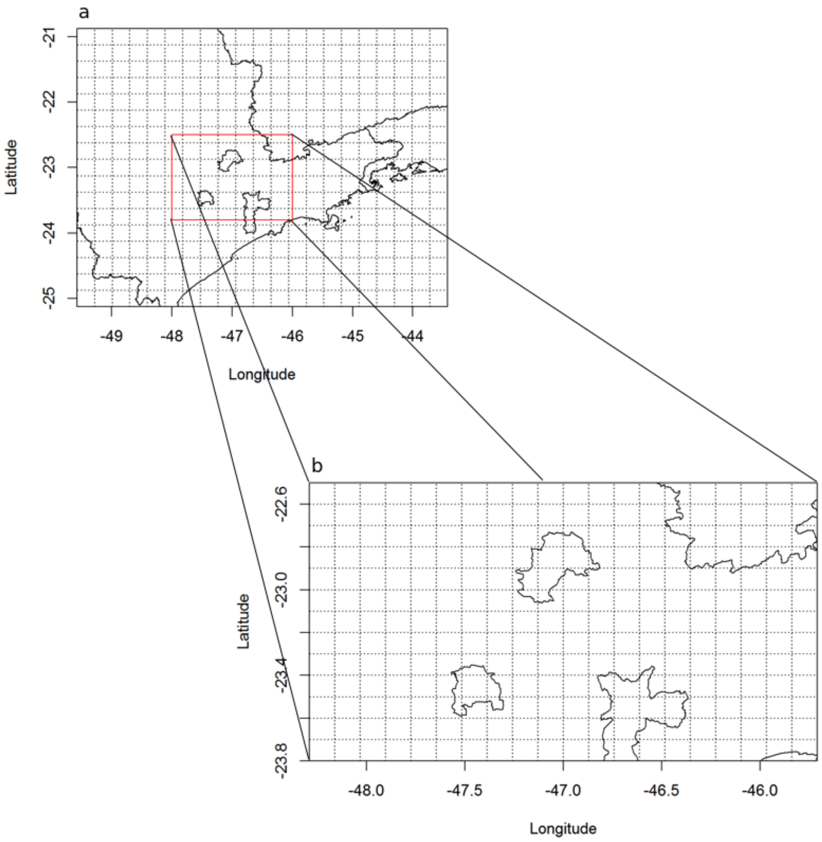


Figure 7.1: a) Spatial grid (0.25°x 0.25°) where the explanatory ERA-Interim dataset was extracted and b) Spatial grid where CHIRPS (0.10°x 0.10°) precipitation data was extracted.

suffice for representing variables with a more linear behavior while they fail when it comes to less linear fields, such as vertical wind speed. Considering the highly non-linear nature of precipitation, a linear technique such as PCA would not be appropriate for an efficient dimensionality reduction.

Hinton and Salakhutdinov (2006) proposed a dimensionality reduction technique based on neural networks called deep autoencoders. Autoencoders are a class of neural networks in which the output layer aim to approximate the input vectors. Figure 7.2 shows a generic diagram of a deep autoencoder based on a multi-layer perceptron. The first hidden-layer encodes the information of a higher dimensional input layer to a lower dimensional "bottleneck layer". The hidden-layer next to the bottleneck layer decodes the lower dimensional information into an output layer of the same dimension as the input. After the network parameters are trained (with stochastic gradient-descent, for example), the outputs of the bottleneck layer correspond to the "non-linear principal components" of the input dataset.

In other words, in a non-linear perspective, it is the best possible lower-dimensional representation of the input dataset. Hinton and Salakhutdinov (2006) shows that well-trained deep autoencoders provide a much better dimensionality reduction than PCA.

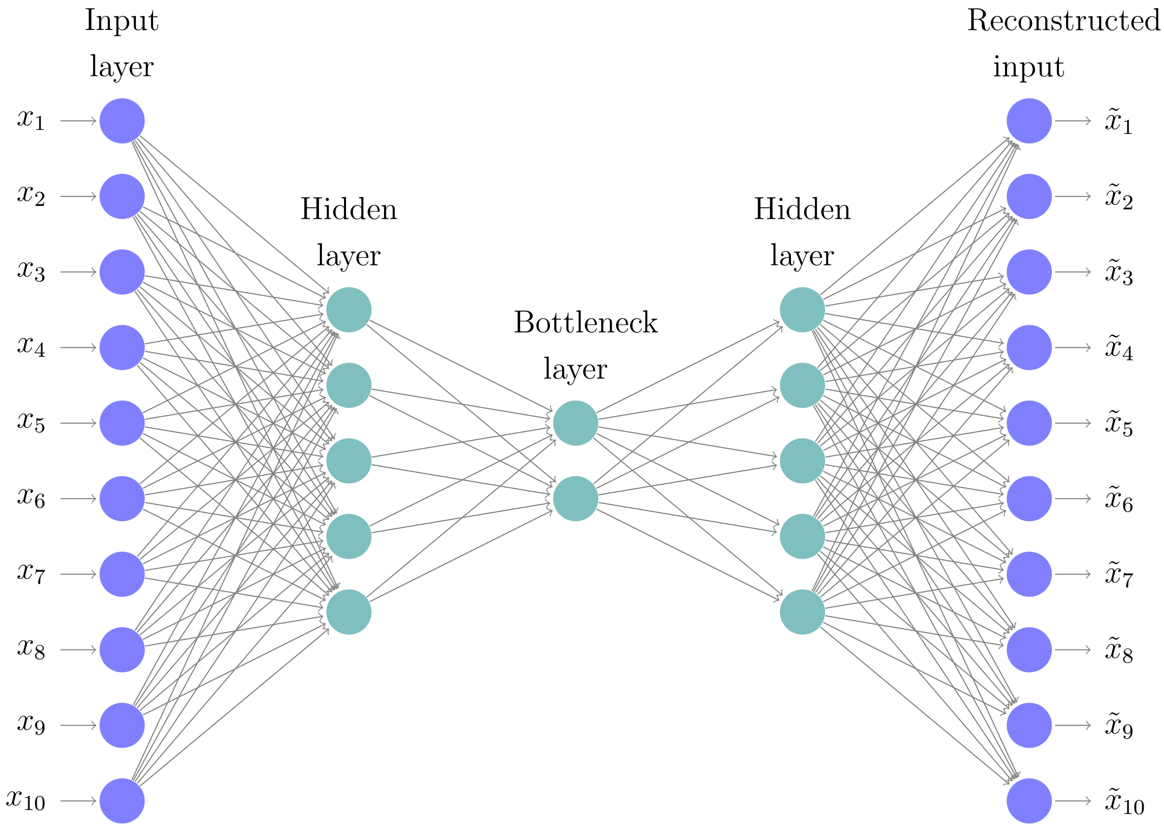


Figure 7.2: Diagram of a generic MLP deep autoencoder architecture with ten input variables and two neurons in the bottleneck layer.

A deep autoencoder with 10 neurons on the bottleneck layer was trained to reduce the dimensionality of the precipitation field. Figure 7.4 exemplifies a decompressed precipitation field after encoding and decoding the original precipitation (Figure 7.3). The reconstructed precipitation have smoother features while keeping the main characteristics of the original field. Now, instead of 260 points, the compressed target dataset consists of 10 non-linear principal components for each day of the training set.

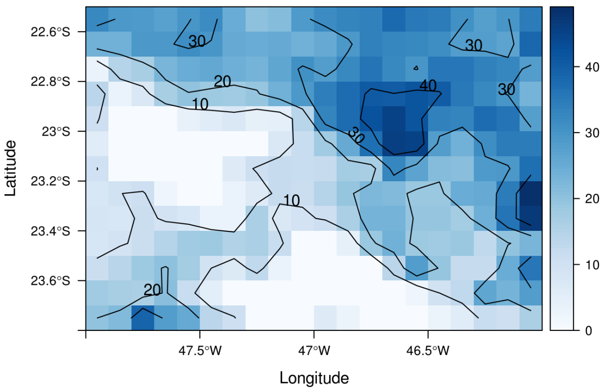


Figure 7.3: 24-hour total precipitation from CHIRPS on 14 April 2009.

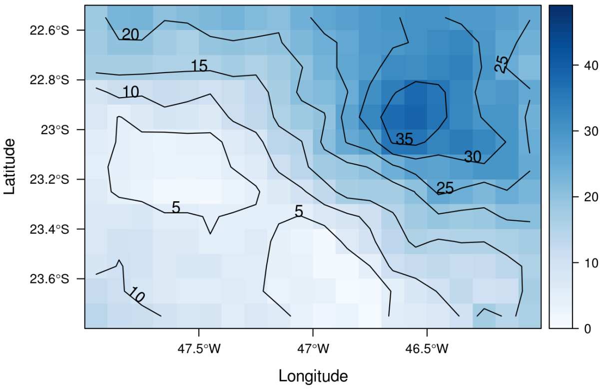


Figure 7.4: Reconstruction of the 24-hour total precipitation on 14 April 2009 by deep autoencoders.

7.2 Results and discussion

The spatial precipitation forecast in the East region of the state of São Paulo was achieved with a combination of multi-layer perceptrons and an autoencoder. As discussed in the previous section, an autoencoder was trained to project the spatial precipitation into 10 non-linear principal components. Following to that, 10 multi-layer perceptrons were trained to predict each non-linear component. The autoencoder is employed one more time to decode the predicted low-dimension projections and construct the predicted precipitation field.

Figures 7.6 and 7.7 shows the RMSE and linear correlation between the predicted precipitation field and CHIRPS data. The evaluation was performed on the test set using ERA-Interim explanatory dataset. Both metrics indicate a good performance on the largest porion of the study area. However, the error rate is higher and the correlation is lower near to the Sierra do Mar mountain range. This loss of performance could be related with two main reasons: (1) the complex topography increases the complexity of physical processes, thus hindering the model's ability to produce realistic rainfall or (2) the CHIRPS rainfall estimate is less realistic near the mountain range. The second argument is related to the satellite IR sensor limitation of estimating warm clouds precipitation.

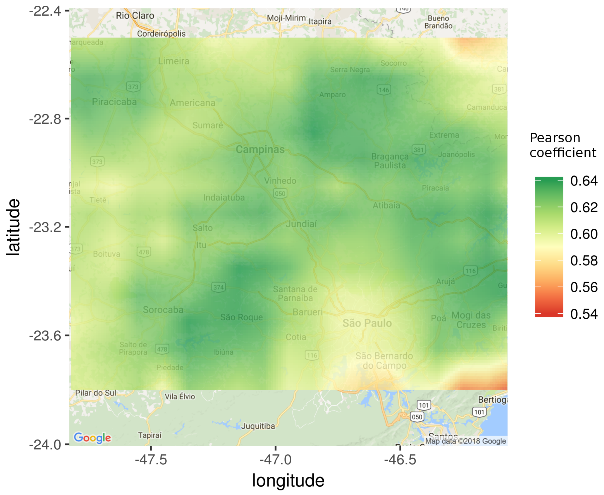


Figure 7.6: Pearson coefficient between the deep learning daily precipitation forecast and CHIRPS data. The coefficient was evaluated on the test set using ERA-Interim explanatory dataset.

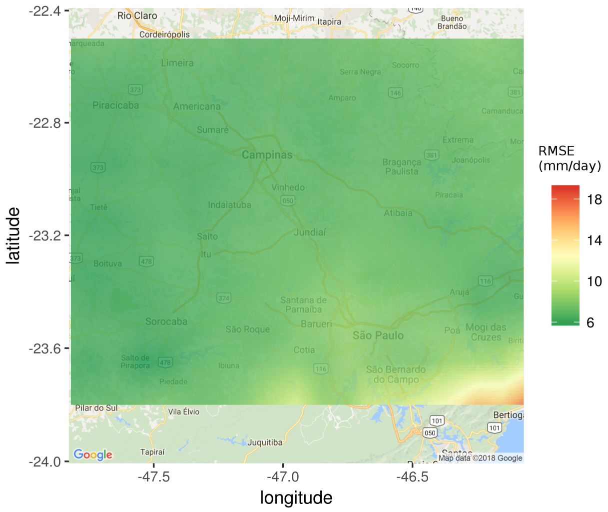


Figure 7.7: RMSE between the deep learning daily precipitation forecast and CHIRPS data. The error was evaluated on the test set using ERA-Interim explanatory dataset.

Chapter 8

Summary, final remarks and future work

The central conclusion of this project is that recent deep learning implementations substantially improve the quantitative precipitation forecast in the area of study. Under a practical point of view, the results presented here highlight the applicability of the methodology in weather forecasting. In a theoretical perspective, the results point towards the understanding of how deep neural networks behave when learning meteorological phenomena. More broadly, the large potential of artificial intelligence applications in atmospheric sciences is emphasized. The improvements obtained here are aligned with trends observed in several other applied areas such as image classification and speech recognition.

The issue of developing a machine learning algorithm for rainfall forecasting begins with the choice of an appropriate precipitation dataset. CHIRPS was the dataset of choice. Chapter 3 presents a short evaluation of CHIRPS daily precipitation estimates against purely in-situ observations. Despite the inhomogeneity of CHIRPS data sources, an overall improvement was observed since 1981. The RMSE, in the order of 7 mm/day, and the determination coefficient, in the order of 40%, were considered satisfactory. The most relevant remark in the usage of this dataset is the limitations that CHIRPS assimilation procedure may have near topographic features such as Sierra do Mar mountain range.

Chapter 4 shows that extreme precipitation events have certain distinguishable patterns: positive temperature advection, higher specific humidity, ascending currents at 500 hPa and low-level convergence along the coastline. This discussion served two purposes. Firstly, to indicate that information about rainfall daily rate is contained in the 0.25°resolution ERA-Interim, and, therefore, it can be used as explanatory dataset. Secondly, to provide reference of which patterns the neural networks were expected to prioritize in Section 5.5.

Chapter 5 is evaluates the DNNs forecasts in the city of São Paulo. In Section 5.2., it is argued that 5 linear principal components are enough to preserve most of the information contained in the ERA-Interim explanatory dataset. The PCA dimensionality reduction was not required in the GFS explanatory dataset, because of its lower spatial resolution. Figures 6.5 to 6.8 conclusively show that DNNs trained with either GFS or ERA-Interim inputs perform far better than the GFS 24-hour rainfall output. Concerning the DNN explanatory datasets, ERA-Interim reanalyisis has a slightly improved performance than GFS reforecast, revealing that better results can be achieved with a more realistic input set.

Except for applied fields where the machine learning developing community [1] is closely involved, such as image recognition, it is seldom discussed in scientific literature how DNNs learn the underlying physical processes of natural phenomena. This fact has created a widespread idea among academicians, particularly in earth sciences, that machine learning algorithms are black-boxes unrelated to physics. In fact, in this moment of the discussion, it is relevant to punctuate that a number of procedures (e.g., convective parameterizations) widely adopted in traditional numerical weather prediction are not necessarily related to physics, instead, they explore statistical properties of the atmosphere. Regardless, recent studies are concerned in unraveling this black-box and providing clearer intepretations of artificial neural networks and machine learning in general. Lin et al. (2017) presented a pioneering discussion about the relationship of DNNs and physical processes, arguing that DNNs take advantage of the hierarchy often found in physical phenomena.

The discussion presented in Section 5.5. intends to narrow the aforementioned gap between machine learning algorithms and physics. When stimulated with one-hot and two-hot vectors, the DNNs behave in agreement with the theoretical understanding of local atmospheric dynamics and with the results in Chapter 4. It indicates that DNNs can be employed to quantitatively evaluate physical relationships in rainfall processes. Ranking the most relevant atmospheric patterns for rainfall prediction could support the weather

---

[1] To name a few: professor Geoffrey Hinton group at University of Toronto, professor Yann LeCun group at New York University and tech companies research labs such as Google, Microsoft and IBM.

forecaster decisions, particularly in regions where the precipitation mechanisms are not well understood.

Chapter 6 shows that it is possible to achieve a satisfactory spatial rainfall forecast with a small number of deep multi-layer perceptrons. To achieve that, a deep autoencoder was employed to create a low-dimensional representation of the spatial precipitation field. The task of predicting the complete precipitation field is then reduced to predicting the non-linear principal components.

Preliminary results not shown in this monograph reveal that similar quality predictions can be achieved for 24 and 48 hour forecast in other areas of Brazil. It indicates that the applicability of this methodology is not limited to ESP and to the 24-hour time window.

The perspective for future studies regarding AI and rainfall forecast is vast. As briefly discussed in Section 5.3.1., modern implementations of recurrent and convolutional neural networks are very successful in a number of prediction and classification tasks. These techniques have interesting properties that might enhance the overall forecast skill. A properly trained CNN, for instance, might dismiss the need of the autoencoder dimensionality reduction, since it is designed to recognize patterns in high-dimensional images. SELU

AI is today one of the most active research field in computer science, applied mathematics and related subjects. It is being progressively recognized by society as an important tool for innovation in science and engineering. The high research activity currently undertaken in AI culminates on a large volume of novel methodologies with potential applications in atmospheric sciences.

Recently, Stanford University has published the 2017 Annual Report (AI Index Team, 2017) to summarize the undergoing activity and progress in AI. The Report highlights that the number of AI papers produced by year is nine times higher today than it was in 1996. In contrast, the average number of papers produced by years in all fields increased only two times. This growth of interest also manifests in the business sector: the number of AI startups in the US increased 14 times and, since 2013, the number of AI related job openings in the US increased 4.5 times.

In conclusion, AI might not provide a definitive solution for meteorological issues such as rainfall forecast. However, the progressively large number of novel techniques may temporarily relief some of meteorology's old struggles.

Bibliography

Achille A., Soatto S., On the emergence of invariance and disentangling in deep representations, arXiv preprint arXiv:1706.01350, 2017

AI Index Team, 2017 2017 Annual Report `http://aiindex.org/2017-report.pdf`

Anochi J., Velho H. C., Mesoscale Precipitation Climate Prediction for Brazilian South Region by Artificial Neural Networks, American Journal of Environmental Engineering, 2016, vol. 6, p. 94

Bishop C. M., Neural networks for pattern recognition. Oxford university press, 1995

Cao J., Farnham D. J., Lall U., Spatial-temporal wind field prediction by Artificial Neural Networks, arXiv preprint arXiv:1712.05293, 2017

Carvalho L. M., Jones C., Liebmann B., Extreme precipitation events in southeastern South America and large-scale convective patterns in the South Atlantic convergence zone, Journal of Climate, 2002, vol. 15, p. 2377

Cavalcanti I. F., Kousky V. E., , 2009 Tempo e clima no Brasil. Oficina de textos

Chaves R. R., Nobre P., Interactions between sea surface temperature over the South Atlantic Ocean and the South Atlantic Convergence Zone, Geophysical Research Letters, 2004, vol. 31

Clark A. J., Gallus Jr W. A., Xue M., Kong F., A comparison of precipitation forecast skill between small convection-allowing and large convection-parameterizing ensembles, Weather and forecasting, 2009, vol. 24, p. 1121

Cybenko G., Approximation by superpositions of a sigmoidal function, Mathematics of control, signals and systems, 1989, vol. 2, p. 303

da Silva A. E., de Carvalho L. M. V., Large-scale index for South America Monsoon (LISAM), Atmospheric Science Letters, 2007, vol. 8, p. 51

de MBS Xavier T., Silva Dias M. A., Xavier A. F., Impact of ENSO episodes on the autumn rainfall patterns near SÃ£o Paulo, Brazil, International Journal of Climatology, 1995, vol. 15, p. 571

Dee D. P., Uppala S., Simmons A., Berrisford P., Poli P., Kobayashi S., Andrae U., Balmaseda M., Balsamo G., Bauer d. P., et al., The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Quarterly Journal of the royal meteorological society, 2011, vol. 137, p. 553

Durkee J. D., Mote T. L., A climatology of warm-season mesoscale convective complexes in subtropical South America, International Journal of Climatology, 2010, vol. 30, p. 418

Freitas E. D., Rozoff C. M., Cotton W. R., Dias P. L. S., Interactions of an urban heat island and sea-breeze circulations during winter over the metropolitan area of São Paulo, Brazil, Boundary-Layer Meteorology, 2007, vol. 122, p. 43

Funk C. C., Peterson P. J., Landsfeld M. F., Pedreros D. H., Verdin J. P., Rowland J. D., Romero B. E., Husak G. J., Michaelsen J. C., Verdin A. P., , 2014 Technical report A quasi-global precipitation time series for drought monitoring. US Geological Survey

Glorot X., Bordes A., Bengio Y., Deep sparse rectifier neural networks. In Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics , 2011, p. 315

Goodfellow I., Bengio Y., Courville A., Bengio Y., Deep learning. vol. 1, MIT press Cambridge, 2016

Graves A., Mohamed A.-r., Hinton G., Speech recognition with deep recurrent neural networks. In Acoustics, speech and signal processing (icassp), 2013 ieee international conference on , 2013, p. 6645

Haddad E. A., Teixeira E., Economic impacts of natural disasters in megacities: The case of floods in São Paulo, Brazil, Habitat International, 2015, vol. 45, p. 106

Hamill T. M., Performance of operational model precipitation forecast guidance during the 2013 Colorado Front-Range floods, Monthly Weather Review, 2014, vol. 142, p. 2609

Hamill T. M., Bates G. T., Whitaker J. S., Murray D. R., Fiorino M., Galarneau Jr T. J., Zhu Y., Lapenta W., NOAA's second-generation global medium-range ensemble reforecast dataset, Bulletin of the American Meteorological Society, 2013, vol. 94, p. 1553

Hamill T. M., Whitaker J. S., Wei X., Ensemble reforecasting: Improving medium-range forecast skill using retrospective forecasts, Monthly Weather Review, 2004, vol. 132, p. 1434

Hinton G. E., Osindero S., Teh Y.-W., A fast learning algorithm for deep belief nets, Neural computation, 2006, vol. 18, p. 1527

Hinton G. E., Salakhutdinov R. R., Reducing the dimensionality of data with neural networks, science, 2006, vol. 313, p. 504

Hinton G. E., Srivastava N., Krizhevsky A., Sutskever I., Salakhutdinov R. R., Improving neural networks by preventing co-adaptation of feature detectors, arXiv preprint arXiv:1207.0580, 2012

Horel J. D., Hahmann A. N., Geisler J. E., An investigation of the annual cycle of convective activity over the tropical Americas, Journal of Climate, 1989, vol. 2, p. 1388

Hornik K., Stinchcombe M., White H., Multilayer feedforward networks are universal approximators, Neural networks, 1989, vol. 2, p. 359

Jorgetti T., da Silva Dias P. L., de Freitas E. D., The relationship between South Atlantic SST and SACZ intensity and positioning, Climate dynamics, 2014, vol. 42, p. 3077

Kang B., Moon S., Regional hydroclimatic projection using an coupled composite downscaling model with statistical bias corrector, KSCE Journal of Civil Engineering, 2017, vol. 21, p. 2991

Kiefer J., Wolfowitz J., Stochastic estimation of the maximum of a regression function, The Annals of Mathematical Statistics, 1952, pp 462–466

LeCun Y., Une procedure d'apprentissage ponr reseau a seuil asymetrique, proceedings of Cognitiva 85, 1985, pp 599–604

LeCun Y., Bengio Y., Hinton G., Deep learning, nature, 2015, vol. 521, p. 436

Lin H. W., Tegmark M., Rolnick D., Why does deep and cheap learning work so well?, Journal of Statistical Physics, 2017, vol. 168, p. 1223

Nasseri M., Asghari K., Abedini M., Optimized scenario for rainfall forecasting using genetic algorithm coupled with artificial neural network, Expert Systems with Applications, 2008, vol. 35, p. 1415

Olson D. A., Junker N. W., Korty B., Evaluation of 33 years of quantitative precipitation forecasting at the NMC, Weather and Forecasting, 1995, vol. 10, p. 498

Perez G. M., Silva Dias M. A., Long-term study of the occurrence and time of passage of sea breeze in São Paulo, 1960–2009, International Journal of Climatology, 2017, vol. 37, p. 1210

Prado F., OscilaÃ§Ã£o interdecadal do PacÃfico e seus impactos no regime de precipitaÃ§Ã£o do Estado de SÃ£o Paulo, University of SÃ£o Paulo, 2010, Master Dissertation

Ramirez M. C. V., de Campos Velho H. F., Ferreira N. J., Artificial neural network technique for rainfall forecasting applied to the Sao Paulo region, Journal of hydrology, 2005, vol. 301, p. 146

Robbins H., Monro S., A stochastic approximation method, The annals of mathematical statistics, 1951, pp 400–407

Rosenblatt F., The perceptron: a probabilistic model for information storage and organization in the brain., Psychological review, 1958, vol. 65, p. 386

Rumelhart D. E., Hinton G. E., Williams R. J., Learning representations by back-propagating errors, nature, 1986, vol. 323, p. 533

Silva Dias M. A., Dias J., Carvalho L. M., Freitas E. D., Dias P. L. S., Changes in extreme daily rainfall for São Paulo, Brazil, Climatic Change, 2013, vol. 116, p. 705

Silva Dias M. A., Vidale P. L., Blanco C. M., Case study and numerical simulation of the summer regional circulation in São Paulo, Brazil, Boundary-Layer Meteorology, 1995, vol. 74, p. 371

Sloughter J. M. L., Raftery A. E., Gneiting T., Fraley C., Probabilistic quantitative precipitation forecasting using Bayesian model averaging, Monthly Weather Review, 2007, vol. 135, p. 3209

Sugahara S., Da Rocha R. P., Ynoue R. Y., Da Silveira R. B., Homogeneity assessment of a station climate series (1933–2005) in the Metropolitan Area of São Paulo: instruments change and urbanization effects, Theoretical and applied climatology, 2012, vol. 107, p. 361

Trenberth K. E., Atmospheric moisture recycling: Role of advection and local evaporation, Journal of Climate, 1999, vol. 12, p. 1368

Velasco I., Fritsch J. M., Mesoscale convective complexes in the Americas, Journal of Geophysical Research: Atmospheres, 1987, vol. 92, p. 9591

Vidal R., Bruna J., Giryes R., Soatto S., Mathematics of Deep Learning, arXiv preprint arXiv:1712.04741, 2017

Werbos P., Beyond regression: new fools for prediction and analysis in the behavioral sciences, Harvard University, 1974, Ph.D. Thesis

Xavier A. C., King C. W., Scanlon B. R., Daily gridded meteorological variables in Brazil (1980–2013), International Journal of Climatology, 2016, vol. 36, p. 2644

Zhou T.-J., Yu R.-C., Atmospheric water vapor transport associated with typical anomalous summer rainfall patterns in China, Journal of Geophysical Research: Atmospheres, 2005, vol. 110

Appendix

Appendix A

Appendix A

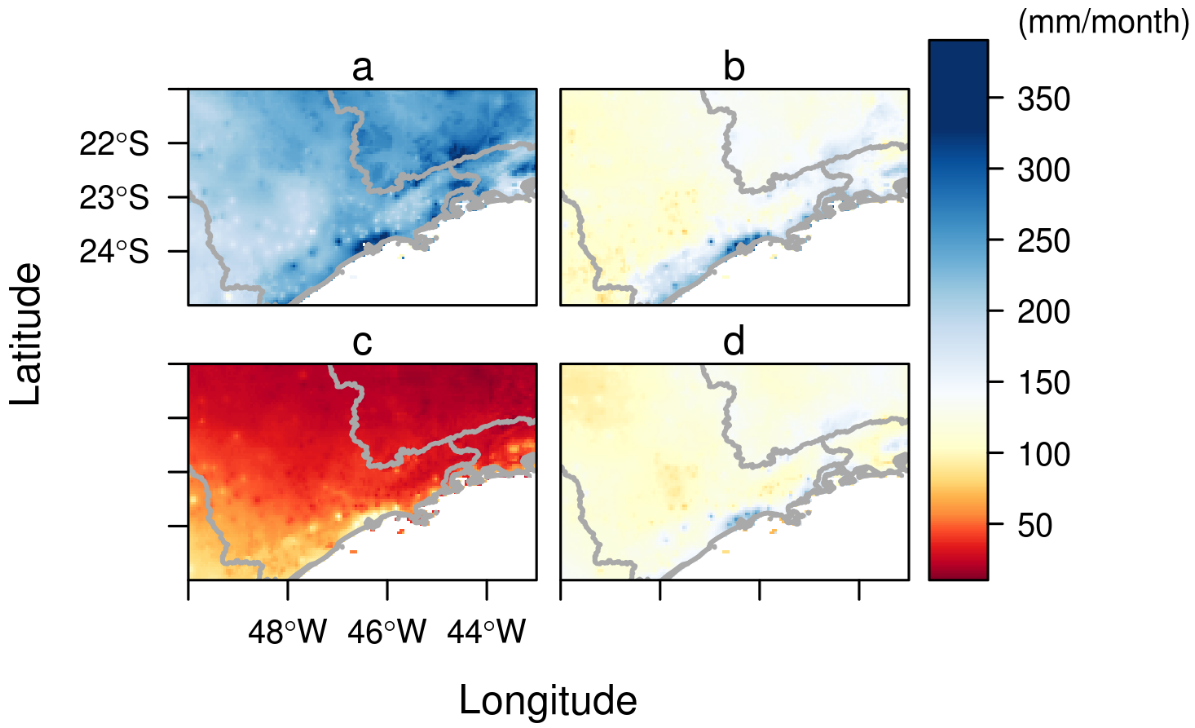


Figure A.1: Monthly averaged precipitation (mm/month, 1981 - 2016) from CHIRPS in ESP during austral summer (a), autumn (b), winter (c) and spring (d).