NEURAL NETWORKS INPUT-BASED MODELS TO CALIBRATE THE MEAN PRECIPITATION OF AN ENSEMBLE PREDICTION SYSTEM

José Roberto Motta Garcia

Doctorate Thesis of the Graduate Course in Applied Computing, guided by Drs. Rafael Duarte Coelho dos Santos, and Christopher Alexander Cunningham Castro, approved in august 24, 2016.

URL of the original document:
<http://urlib.net/8JMKD3MGP3W34P/3M7RCUE>

INPE

São José dos Campos

2016

NEURAL NETWORKS INPUT-BASED MODELS TO CALIBRATE THE MEAN PRECIPITATION OF AN ENSEMBLE PREDICTION SYSTEM

José Roberto Motta Garcia

Doctorate Thesis of the Graduate Course in Applied Computing, guided by Drs. Rafael Duarte Coelho dos Santos, and Christopher Alexander Cunningham Castro, approved in august 24, 2016.

URL of the original document:
<http://urlib.net/8JMKD3MGP3W34P/3M7RCUE>

INPE

São José dos Campos

2016

Aluno (a):   *José Roberto Motta Garcia*

Título: " NEURAL NETWORKS INPUT-BASED MODELS TO CALIBRATE THE MEAN PRECIPITATION OF AN ENSEMBLE PREDICTION SYSTEM".

Aprovado (a) pela Banca Examinadora em cumprimento ao requisito exigido para obtenção do Título de *Doutor(a)* em *Computação Aplicada*

Dr.   Antonio Miguel Vieira Monteiro

Presidente / INPE / SJCampos - SP

Dr.   Rafael Duarte Coelho dos Santos
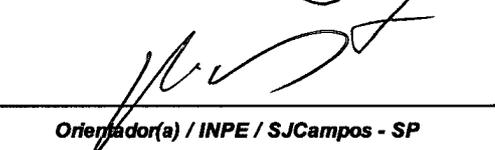
Orientador(a) / INPE / SJCampos - SP

Dr.   Christopher Alexander Cunningham Castro

Orientador(a) / CEMADEN / São José dos Campos - SP

Dr.   Marcos Gonçalves Quiles

Membro da Banca / UNIFESP / São José dos Campos - SP

Dr.   Nandamudi Lankalapalli Vijaykumar

Membro da Banca / INPE / SJCampos - SP

Dr.   Carlos Henrique Quartucci Forster

Convidado(a) / ITA / São José dos Campos - SP

Dr.   Fábio Dall Cortivo

Convidado(a) / ITA / São José dos Campos - SP

Este trabalho foi aprovado por:

( ) *maioria simples*

(X) *unanimidade*

São José dos Campos, 24 de Agosto de 2016

"The best way to predict your future is to create it".

Abraham Lincoln

To my family.

ACKNOWLEDGEMENTS

x

ABSTRACT

Although ensemble forecasting systems provide richer forecasts by adding probabilistic concepts to single deterministic forecasts, they have intrinsic shortcomings caused by the lack of full comprehension of the relationship between meteorological variables. It is especially noticed in medium and large-scale forecasts, whose effects of chaotic behavior of the atmosphere drastically increase as the target forecasting date is lengthened. Improvements on weather forecasting systems can be done either by the meteorology staff concerning physical aspects of weather behavior as well as by implementing computational statistical methods in order to tune the weather forecasting model output. The purpose of this work is to compute, along the forecast horizon, a more accurate precipitation value than the ensemble mean precipitation by post-processing INPE/CPTEC's ensemble prediction output. To achieve the goal, some prognostic fields and derived data are combined and submitted as explanatory variables to an artificial neural network system. Experiments were guided in an exploratory way such that several computational models were generated and thereafter assessed. The study was individually performed at some grid points located within the boundaries of La Plata Basin. Results indicate that the application of this methodology presented values closer to actual values when compared to the ensemble mean precipitation. It also shows that the inclusion of the ensemble mean precipitation itself, as well as data from adjacent grid points, improve the calibration process of the target grid point. In addition, the exploratory approach detects different artificial network models to fit specific location and lead-time. Although this input-driven system computes less than ideal forecasting values, it performs better than the mean output of the ensemble model, which is widely used in various weather forecasting products.

Keywords: Machine learning. Neural nets. Statistical weather forecasting. Rain.

MODELOS DE REDES NEURAIS BASEADOS EM ENTRADA PARA CALIBRAR A PRECIPITAÇÃO MEDIA DE UM SISTEMA DE PREVISÃO POR CONJUNTOS

Embora os sistemas de previsão por conjuntos forneçam previsões mais ricas, acrescentando conceitos probabilísticos às previsões determinísticas simples, eles têm deficiências intrínsecas causadas pela falta de plena compreensão da relação entre variáveis meteorológicas. Isto é especialmente notado nas previsões de média e grande escala, cujos efeitos de comportamento caótico da atmosfera aumentam drasticamente à medida que a data alvo de previsão é estendida. Melhorias nos sistemas de previsão do tempo podem ser feitas tanto pela equipe de meteorologia em relação aos aspectos físicos do comportamento do tempo, bem como através da aplicação de métodos computacionais estatísticos, visando ajustar a saída do modelo de previsão. O objetivo deste trabalho é calcular, ao longo do horizonte de previsão, um valor de precipitação mais preciso do que a precipitação média do sistema de previsão por conjunto através do pós-processamento das saídas do modelo do INPE/CPTEC. Para atingir esta meta, alguns campos de prognóstico e dados derivados são combinados e apresentados como variáveis explanatórias a um sistema de rede neural artificial. Os experimentos foram orientados de forma exploratória onde vários modelos computacionais foram gerados e posteriormente avaliados. O estudo foi realizado individualmente em alguns pontos de grade localizados dentro dos limites da Bacia da Prata. Os resultados indicam que a aplicação desta metodologia apresentou valores mais próximos da realidade do que a média de precipitação do sistema de previsão por conjuntos. Mostra também que ao incluir a própria precipitação média, bem como dados de pontos de grade adjacentes o processo de calibração melhora no ponto de grade alvo. Além disso, a abordagem exploratória traz uma melhora ainda maior pois detecta diferentes modelos de redes neurais para locais e dias de previsão específicos. Embora este sistema baseado em entrada calcule valores de previsão inferiores aos ideais, ele tem um desempenho melhor do que a média do modelo de previsão por conjunto, que é amplamente usado em vários produtos de previsão de tempo.

Palavras-chave: Aprendizado de máquina. Redes neurais artificiais. Previsão de Tempo. Chuva.

LIST OF FIGURES

Page

LIST OF TABLES

LIST OF ACRONYMS AND ABBREVIATIONS

AGCM   Atmospheric General Circulation Model

ANFIS   Adaptive Neuro-Fuzzy Inference Systems

ANN   Artificial Neural Network

BMA   Bayesian Model Averaging

BP   Back-Propagation

CPTEC   Centro de Previsão de Tempo e Estudos Climáticos

CTRL   Control, forecast evolution of the best initial state of the atmosphere

DT   Decision Tree

ECMWF   European Centre for Medium-Range Weather Forecasts

EOF   Empirical Orthogonal Function

EPS   Ensemble Prediction System

EXP   Experiment

FF   Feed-Foward

GDP   Gross Domestic Product

IC   Initial Condition

INPE   Instituto Nacional de Pesquisas Espaciais

LogReg   Logistic Regression

LPB   La Plata Basin

LR   Linear Regression

MEAN   Arithmetic mean of ensemble precipitation values (or ensemble mean)

ML   Machine Learning

MLP   Multi Layer Perceptron

MOS   Model Output Statistics

MSC   Meteorological Service of Canada

NASA   National Aeronautics and Space Administration (USA)

NCEP   National Centers for Environmental Predictions (USA)

NWP   Numeric Weather Prediction

NWPM   Numeric Weather Prediction Model

NWPS   Numeric Weather Prediction System

OBS   Observed precipitation value

PQPF   Probabilistic Quantitative Precipitation Forecast

ReLU   Rectified Linear Unit

RF   Random Forest

RMSE   Root Mean Squared Error

RN   Recurrent Network

RPROP   Resilient Back-Propagation

SS   Skill Score

SVM   Support Vector Machine

WMO   World Meteorological Organization

CONTENTS

1   INTRODUCTION

The current state of the art in weather forecasting is Numerical Weather Prediction Systems (NWPS) and the current global model that simulates the atmospheric processes was idealized in the early 1900s. It is based on single weather estimates built from the evolution of initial measurements of the atmosphere throughout time, which are done by Atmospheric General Circulation Models (AGCM). Over time, scientists have found that single predictions have critical accuracy limitations that are specially verified on predictions longer than 5-6 days, i.e., in the mid-range scale or longer.

The reason for the lack of accuracy is credited to the chaotic behavior of the atmosphere once that imprecise initial states make the forecasting error grow in an exponential way, as time goes by. These imprecision were found to be inherent to weather forecasting since neither the initial state of the atmosphere nor physical equations of prediction systems faithfully represents the actual initial state and weather behavior.

In order to try to capture such initial and modeling uncertainties, multiple deterministic predictions are added to a single control deterministic prediction, composing an ensemble of predictions. These multiple predictions differ from each other by either applying small perturbations to the control initial condition or varying the way that weather evolves through time, i.e., changing the physics of the forecasting model.

The advantages of ensemble prediction over deterministic prediction, which focus only on the information of the magnitude of the event, include the extension of predictability beyond mid-range scale and the gauging the uncertainty of the event in a probabilistic way, which is done ordinary statistical computations from a set of values.

The concept of ensemble spread appeared as a consequence of the ensemble approach and represent how distant the lower ensemble value is from the

1

higher value. A compact data distribution leads to more predictable weather, and vice versa. Another concept that arose is the ensemble mean (MEAN, from now on), which is the arithmetic mean of the ensemble values.

The MEAN is known as the best available estimate of the future state of the atmosphere because it holds smaller error than any of the individual ensemble members. Its use adds an expressive value to ensemble-based forecast products.

Despite endorsed by the meteorological community as a way to enhance the predictability and applicability of weather forecasting, Ensemble Prediction Systems (EPS) propagate the deficiencies caused by systematic errors that are found in the deterministic prediction model. In order to alleviate the effects, the literature has proven that historical ensemble data can be processed to adjust the values of its own output data, ameliorating the effect of the error between forecast and observed value. The literature also strongly signalizes that one of the available techniques to execute such post-processing task are Artificial Neural Networks (ANN), since it is conceptually able to deal with the complex and dynamic relationships that exist among weather variables.

That said, the general goal of this work is to propose a methodology to adjust (calibrate) the ensemble mean precipitation forecast of the ensemble prediction system of the Center for Weather Forecasts and Climate Studies of the Brazilian National Institute for Space Research (CPTEC/INPE).

In order to test and validate the methodology, few locations within La Plata Basin were chosen as spatial domain as well as data from 2009-2012 rainy seasons as temporal domain.

The pioneering spirit of this methodology is based on grouping few meteorological variables, named "experiment", and submit each of them to one hundred ANN structures, aiming to reach the optimum combination of experiment and ANN topology for each location and forecast target date.

To accomplish this task the following specific goals, are:

a) assess how much the chosen meteorological variables are useful in order to approximate OBS;

b) investigate the combination of meteorological variable that best calibrates the ensemble mean precipitation in different lead-times;

c) investigate which MLP topology performs better with respect to the lead-time;

d) analyze if data from adjacent grid points have some influence on data from a specific grid point.

This work is organized as follows: Chapter 2 progressively presents the fundamental knowledge to understand this work, Chapter 3 explains the data and methodology applied to accomplish the goals, and Chapter 4 reports the experiments and results in numbers and plots. Finally, conclusions are exposed in Chapter 5, which is followed by the list of references.

2 FOUNDATION

2.1. Deterministic weather prediction

Since Bjerknes (1904) introduced the deterministic "rational"-numerical method to estimate values of future states of the atmosphere based on the measure of its current state, the weather prediction has been developed and evolved at first to a Numeric Weather Prediction Model (NWPM) (RICHARDSON, 1922) and then to its computational version known as Numerical Weather Prediction Systems (NWPS) (CHARNEY, 1948, 1950, 1951), which is the state of the art of weather prediction.

A NWPM that solves the atmospheric representation in a global domain is called Atmospheric General Circulation Model (AGCM). The future state of the atmosphere (prediction) is estimated from its current state, called Initial Condition (IC). The whole process starts with an IC that is submitted to a loop of numerical solving of complex physics equations that iterates according to time-steps in a way that the last output is used as the input for the next iteration, aiming to represent the evolution of the state of the atmosphere in time. This time stepping occurs up to the final prediction target date in the forecast horizon.

Analysis is a snapshot of the atmosphere at a certain moment of time that is conceived from complex observation gathering and data-assimilation techniques of different data sources, aiming to be the best representation of the current state of the atmosphere as possible. An IC is the starting state of the atmosphere that will be evolved throughout the forecast horizon, it can be identical to the analysis or some variation of it. Time-step is an increment of time that represents the temporal evolution along iterations of the computation of predictions and its size varies according to internal routines of the AGCM, respecting its particular working details.

The parameterization is an important phase in the whole process to provide a way to treat the unresolved effects of the small-scale processes on the resolved scale of the AGCM. It is required because these effects cannot be explicitly predicted in full detail in model forecast equations. Generically, parameterization is used to simulate the effect of shortwave (solar) and long-wave (terrestrial) radiation in the atmosphere, land and sea surface characteristics and their impact and mechanical transfer of heat and moisture (UCAR, 2009).

2.2. **Limitations of deterministic weather prediction**

Having only a single picture of the future has revealed some limitations since forecasters can estimate nothing but the magnitude of the event without any statement of the confidence. Moreover, through time, scientists have proven that atmosphere has a chaotic, hence nonlinear, nature (EADY, 1949; LORENZ, 1963, 1965a, 1965b, 1969), meaning that small differences in the initial state of the atmosphere ultimately result in large differences in the forecast, which generates a time limit specially for longer-range deterministic predictions caused by the rapid growth of errors. Given such evidences, it becomes clear that in addition to an event's prediction in deterministic way, its uncertainty and predictability also needs to be estimated, i.e. how uncertainties evolve during the forecast horizon.

These errors exist at every component within the weather forecasting process because all of them have its own uncertainties, such as (REYNOLDS, 1994):

● Incomplete and imperfect analysis due to hardware limitations; data errors and inconsistencies; interpolation methods over areas with insufficient data; etc. – known as initial uncertainties and responsible for the internal error.

● Physics equations due to lack of a complete understanding of the atmospheric physics process and chemistry it tries to simulate; inherent

limitation of the predictability of highly complex processes forcing the use of approximations and simplifications to represent them – known as model uncertainties and responsible for the external error.

It must also be mentioned that there are trade offs between the appropriate time for publishing the forecasts versus the time to produce them, since forecasts can't take much time to be published under penalty of being useless. Figure 2.1 shows the definitions of the forecast horizons applied in meteorology with respect to the target forecasting date.

Figure 2.1 Forecast horizon range definitions.
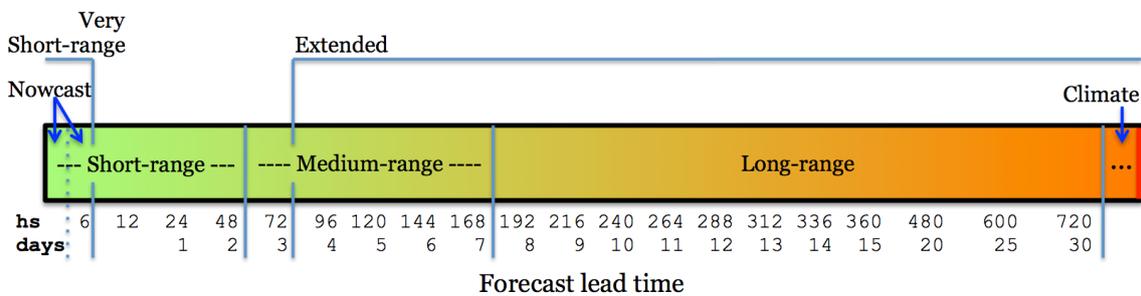
Ranges are expressed in hours (hs) and days. The gradient color from green to orange roughly represents few uncertainties at short-range forecasts and orange represents many uncertainties at longer-range forecasts. Literature is not sharp about limits.

Source: Adapted from Petroliagis and Pinson (2013), AMS (2016), and NOAA (2016).

2.3. **Ensemble prediction**

Aiming to simulate the effects of the chaotic weather behavior, Epstein (1969) idealized the concept of Ensemble Prediction Systems (EPS), which consists on releasing an ensemble of future weather scenarios. Consequently, it enables to gauge the uncertainty of the event in a probabilistic way instead of focusing only on the information of the magnitude of the event, as it happens in the deterministic approach.

An EPS run is composed by multiple AGCM runs (predictions). By having multiple predictions, this approach provide means to go beyond deterministic

prediction and make further estimates like: the range of possible values the prediction can reach, the probability for any individual prediction, and the most likely scenario to occur (UCAR, 2009).

As shown in Figure 2.2, ensemble prediction is composed by a set of deterministic forecasts evolution, called members. Aiming to simulate the IC uncertainties, one of existing manners is to make slight perturbations on the analysis and then run the AGCM with them as new ICs. This process results in a final multiple-prediction scenario that can be used to release a probabilistic forecast emphasizing the confidence of events. In ensemble prediction jargon, the evolution of the analysis is called control member, or CTRL from now on.

Figure 2.2   Evolution of deterministic and ensemble weather prediction
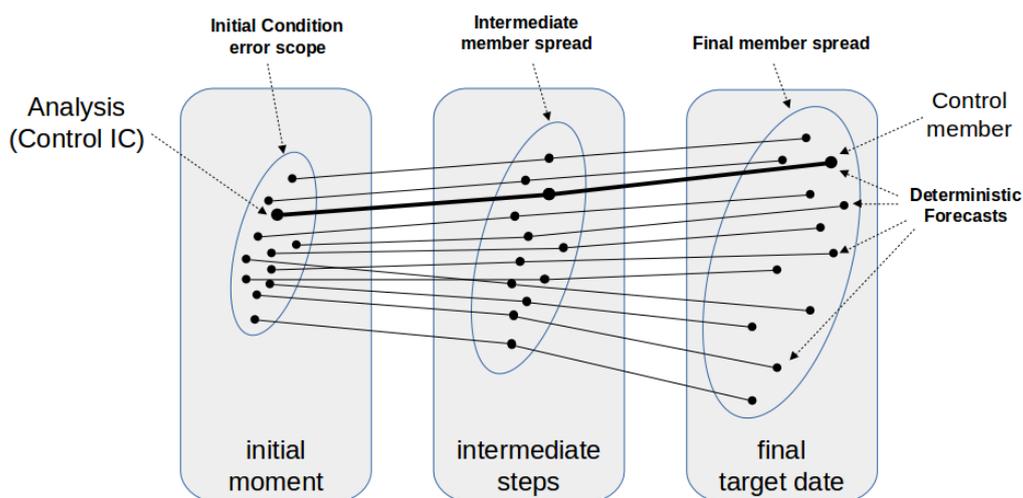
Black-filled circles represents an atmospheric state, thin lines represent the evolution of the forecasts throughout time (members), bold line represents the evolution of CTRL, ellipses represent the range of uncertainties.

Source: Adapted from INPE.CPTEC (2016).

Due to computational time limits, these perturbations must try to best represent the infinite number of values within the range of the initial and model uncertainties, since it is not feasible to process the whole range.

Different strategies can be used for generating members. An ensemble can be composed by either: members, which are built from multiple parameterizations (focusing on model uncertainties); multiple ICs (focusing on the initial uncertainties); or even from a combination of both (BUIZZA et al., 2005).

The evolution of EPSs can be found in Lewis (2005) and Lynch (2007) and some of the next-generation of NWPSs is reported in Hong (2010). In the comparison among the EPS of three big weather centers (ECMWF-Europe, NCEP-USA and MSC-Canada) in Buizza et al. (2005) concluded that, although more valuable than single forecasts, the weaknesses identified in the three systems offer guidelines for the future development of ensemble forecasting techniques. Among them are, "Improving the simulation of model uncertainties" and "Investigating the possibility to generate calibrated products".

2.4. **Weather prediction post-processing related work**

The process of seeking for the best computational model that maps the relationship between the output of a numerical weather prediction (NWP) model and observations, in order to improve forecasts, was concretely conceived in 1965 (GLAHN; LOWRY, 1972) and is called Model Output Statistics (MOS). It is an objective weather forecasting technique that consists in determining a statistical relationship between a variable to be predicted (namely dependent, target, or explained variable) and variables forecast by a numerical model at some projection time (namely independent variables, predictors, or explanatory variables). Since then, it has evolved to MOS-2000 (GLAHN; DALAVALLE, 2000) and has been used in a variety of situations. Examples of statistical post-processing are related in the paragraphs below.

Bocchieri (1977) used MOS for predicting the probability of heavy snow; Zunrndorfer (1980) categorized and forecast the probability of precipitation amount based on MOS; Carter et al. (1989) developed a interpretive forecasting system in the USA National Weather Service (NWS) to serve as a statistical guidance for most weather elements and projections for USA and Alaska;

Taylor and Buizza (2002) used 51 ensemble members as input to an Artificial Neural Network (ANN) to estimate the load demand for the whole England and Wales; Fall et al. (2007) used weather prediction models for hazard mitigation planning based on heavy off-season rains in Senegal; Guarnieri et al. (2007) used the INPE/CPTEC's Regional ETA model outcome as input to an ANN to refine sun radiation prediction; Wei et al. (2008) used Ensemble Transform (ET) method and variations on ensemble forecasts to gauge uncertainties and fed them back to its own analysis fields.

Also, Bravo et al. (2009) used quantitative precipitation forecasts as input to a multilayer feed-forward ANN and a distributed hydrologic model to predict medium-range stream flow on the Furnas Reservoir on the Rio Grande River; Wessel et. al. (2009) improved the short-term wind forecast based on weather forecasts of the German weather service and the online measured power output of forecast wind farms; ECMWF has developed a tool called Extreme Forecast Index to provide forecasters with an indication of potential extreme weather events based on information from the ensemble predictions (PETROLIAGIS; PINSON, 2013); In Kalteh (2013), ANN and SVM *ad hoc* models were used for monthly river flow forecasting.

2.5. Ensemble spread and ensemble mean

EPS output (predictions) can be used to estimate the probability of occurrence of any weather event, since there are multiple predicted values for a single target date. The spread of the values can be very different and directly affects predictability of the atmosphere. In such way, some distributions can be more compact than others, showing that the predictions are more coherent and hence, the final scenario is more likely to be predicted.

The best available estimate of the future state of the atmosphere is the arithmetic mean of the ensemble members, or ensemble mean. In fact, it has, on average, a smaller error than any of the individual ensemble members (LEITH, 1974; MURPHY, 1988). It happens because ensemble mean smooths

out unpredictable detail and simply presents the more predictable elements of the forecast. It can be a good forecasting guide but forecasters must not rely only on it because it will rarely capture extreme events.

The skill that the ensemble mean (MEAN, from now on) has to get closer to the actual observed values (OBS, from now on) is highly related to the size of the ensemble spread. The less spread it is, the more likely ensemble mean will occur, and vice versa. The additional use of a measure of the ensemble distribution adds an expressive value to ensemble-mean-based forecast products. Examples of predictability associated to large and small ensemble precipitation spread during the evolution of the forecast are shown in Figure 2.3.

Figure 2.3   Evolution of the ensemble precipitation forecasting elements.

a)

CPTEC EPS-12hs Precip Fct Spread
Caxambu_MG - Run Date: 2009-02-07

b)

CPTEC EPS-12hs Precip Fct Spread
Caxambu_MG - Run Date: 2009-02-12

c)

CPTEC EPS-12hs Precip Fct Spread
Caxambu_MG - Run Date: 2009-10-03

d)

CPTEC EPS-12hs Precip Fct Spread
Caxambu_MG - Run Date: 2012-02-01
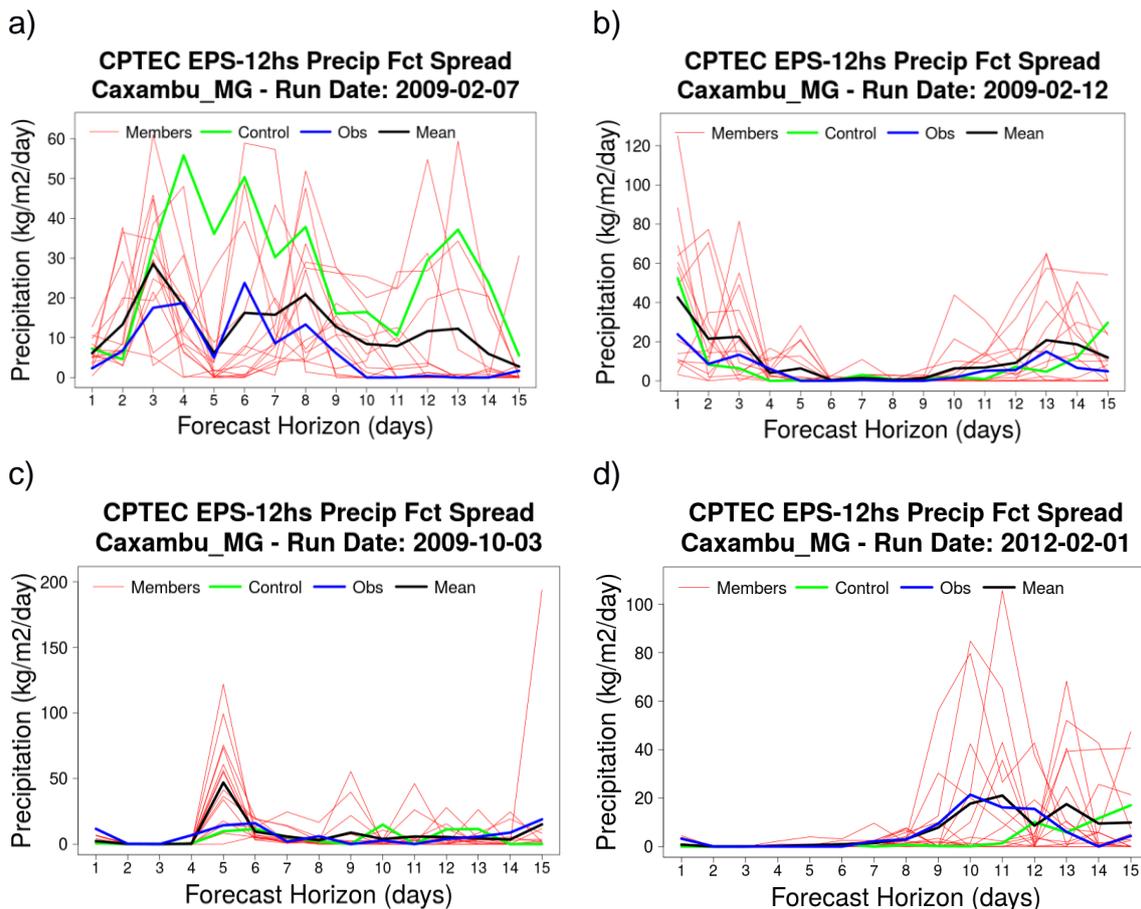
Perturbed ensemble members are in thin red lines, CTRL is the thick green line, MEAN is the thick black line, and OBS is in thick blue line. Lead-time extends along the fifteen

days of the CPTEC-EPS's forecast horizon. Each plot is generated from distinct CPTEC's EPS runs. Data were extracted from a grid point in Caxambu, in the Minas Gerais State.

Source: author's production.

The performance of the MEAN seems to be the best since it roughly follows OBS closer than any other member. As demonstrates Table 2.1, Pearson's correlation coefficient (r) between MEAN and OBS is 0.70 (in bold), which is higher than any other ensemble member.

Table 2.1   Pearson's correlation coefficients of forecasting elements and OBS.

MEMBER	r	MEMBER	r	MEMBER	r	MEMBER	r
MEAN	0.70	PREC_02P	0.29	PREC_04P	0.53	PREC_06P	0.43
PREC_01N	0.36	PREC_03N	0.44	PREC_05N	0.59	PREC_07N	0.39
PREC_01P	0.30	PREC_03P	0.23	PREC_05P	0.35	PREC_07P	0.38
PREC_02N	0.11	PREC_04N	0.36	PREC_06N	0.62	PREC_AVN	0.43

The correlation coefficient (r) was calculated for MEAN and all ensemble members with respect to OBS for the specific dates reported on each of the four plots presented in Figure 2.3. The suffix 01 to 07, N, P and AVN are identification of ensemble members and are explained later.

Source: author's production.

In addition, the four distinct scenarios show peculiar situations, such as but not limited to:

● Large spread causing degradation in MEAN skill, as in days 2-14 in (a), days 1-3 and 13 in (b) and days 9-15 in (d), although it still can be considered the best forecast among all members.

● Chaotic behavior of CTRL leading to a completely wrong forecast, as in (a)

- Small spread that increases MEAN skill, as in days 6-9 in (b), 1-4 in (c) and 1-8 in (d).

- Probabilities that can lead to misleading forecast, as at day #5 in (c) where most members was positioned at higher values causing MEAN to be raised and the actual value was verified among only few members, where the probability were low.

2.6. **Rank histograms**

One of the main tools to diagnose the ensemble spread is the rank histogram (EUMETCAL, 2011). It provides ways to visualize how the spread is related to observed data, serving as a guide to correct (calibrate) the ensemble by evaluating one of its forms (shapes).

It is based on the assumption that ensemble member forecasts are distributed in ranges (or "bins") of the predicted variable such that the probability of occurrence of the observation within each bin is equal, simulating a perfect distribution of members. Thus, in the best case and over a large enough sample, the histogram should take a flat shape, meaning that, on average, the ensemble spread correctly represents the uncertainty in the forecast.

Bins are ranges whose limits are determined by ranking the ensemble members from the lowest to the highest. If there are N ensemble members, then there will be N+1 bins. Rank histograms are built by increasing the relative frequency of each bin by one, where each observation falls into. The outer bins – both lowest and highest – are open-ended, which means that any observed value lower than the ensemble lowest value or higher than the ensemble highest value should be allocated in those bins, respectively.

Hamill and Colucci (1997) performed some experiments on calibrating the spread of members of an ensemble system by post-processing them and had the following results. Figure 2.4 shows rank histograms of three forecast fields

as they were released from model (left column). Distributions are nonuniform, meaning that there is a high tendency for the distributions to be most populated at the extreme ranks. This may indicate systematic errors in the forecast, insufficient variability among ensemble members. Then, a 15-day-based correction was applied and the associated results are show in the right column of the Figure 2.4. When comparing both b) and c), from both columns, there can be noticed that the skewness of the distribution is nearly eliminated.

Figure 2.4   Rank distributions 24-h forecasts from the ensemble.
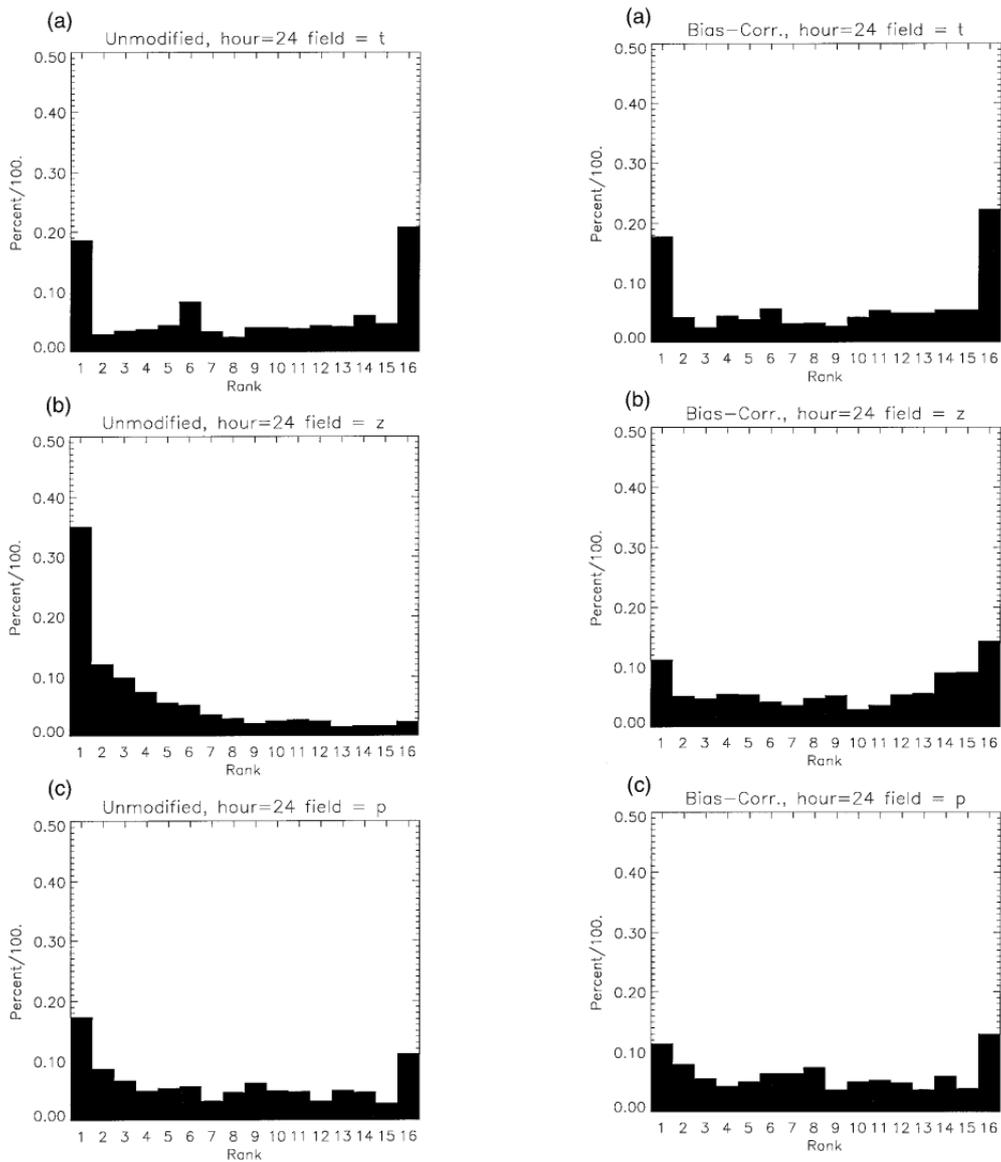
(a) 850-mb temperature, (b) 500-mb geopotential height, and (c) 24-h total precipitation amount. Left columns refers to raw data and right column to calibrated data.

Source: Adapted from Hamill and Colucci (1997).

The authors also stated that there were potentially many causes to not improving case "a": a) the bias corrections may need to be more sophisticated as with MOS; b) the ensemble may truly be insufficiently variable, due to either less than optimal choices for ICs, or model errors. Other commonly rank histogram shapes are show in Figure 2.5.

Figure 2.5   Shapes of commonly found rank histograms.

a)   b)   c)

d)   e)
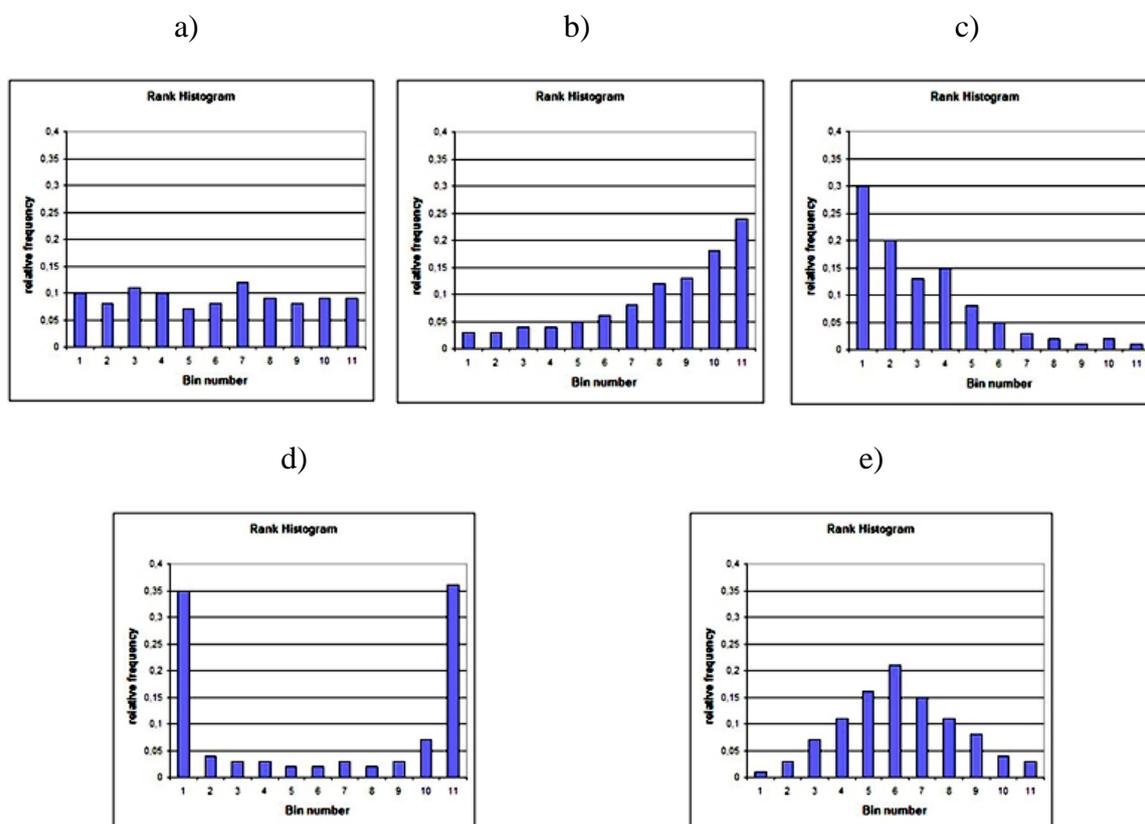
X-axis is the ensemble bins and Y-axis is the relative frequency. Figure 2.5 shows that the shape in plot (a) is "flat", meaning that on average the spread is correct since observations should fall equally into every bin (ideal shape). Plot (b) is left-skewed, meaning that observations are too often higher than the highest value of the ensemble members, indicating that there is a tendency to underforecast, in other words,

members values tend to be lower than observed value. Plot (c), oppositely, is right-skewed, meaning that observations are too often lower than the lowest value of the ensemble members, indicating that there is a tendency to overforecast. The "U" shape presented in (d) shows that the spread tends to be underdispersive, while both lower and higher limits are not enough to capture most observed values. Plot (e) shows a center tendency, meaning that observed values occur too often in the center of the spread distribution, characterizing an overdispersive spread.

Source: Eumetcal (2011).

2.7. **Ensemble spread correction related work**

Despite all improvements EPS has been adding to weather prediction it is not either perfect or an error-free system. It imposes additional difficulties by propagating the uncertainties already exposed in 1.2 (GNEITING; RAFTERY, 2005) leading the ensemble spread to have a systematic error when compared to observations, characterizing a bias. Furthermore, forecast ensembles are typically underdispersive (HAMILL; COLUCCI, 1997; ECKEL; WALTERS, 1998).

In order to handle and correct this issue, researches have been demonstrating that historical ensemble data can be post-processed to produce means to calibrate probabilistic forecast and also, the more cases available, the more sophisticated bias corrections can be applied, improving the probability forecasts (HAMILL; COLUCCI, 1997). Ideally, it is preferable to have as many sample cases as possible and correct for systematic error by location, as is done with Model Output Statistics (MOS; DALLAVALLE et al., 1992).

According to Gneiting et al. (2007) calibration here refers to post-processing techniques that apply statistical adjustments of numerical forecasts in order to increase their sharpness.

Several techniques, which are related in following paragraphs, have been developed to calibrate ensemble members and, consequently, the MEAN. Most

of them are based on the study of the relationship between error, forecast, and observed value related to a specific location.

Eckel and Walters (1998) constructed a calibration method to account for systematic errors that arise from an imperfect forecast model and initial perturbations in order to produce higher quality Probabilistic Quantitative Precipitation Forecasts (PQPFs). Calibration technique noticeably improved the quality of PQPF and extended predictability by about 1 day, but predictability was found to be dependent upon the precipitation category and the calibration was found to be useful only for short-range PQPFs. Applequist et al. (2002) used five different linear and nonlinear statistical methodologies, namely, LR, discriminant analysis, LogReg, and ANN to make comparisons of forecast skill. Results indicated that LogReg performed best among all methodologies but all other also showed significantly greater skill than LR, which was the reference.

Gallus and Seagal (2004), divided the range of forecast precipitation in bins and then applies linear regression to compute the probabilities for other forecast precipitation values. Raftery et al. (2005) proposed a method based on BMA, which is a standard method for combining predictive distributions from different sources, and Hamill and Whitaker (2006) used reforecast analogs to calibrate members based on the most similar situation.

Yuan et al. (2007) applied a feed-forward ANN to calibrate probabilistic quantitative precipitation forecast (PQPF) of the NCEP regional model. The calibration procedure improved the scores for all geographic regions and most precipitation thresholds but degraded the resolution of the PQPFs by systematically producing more forecasts with low nonzero forecast probabilities, mainly where the sample of observed events was relatively small.

Sloughter et al. (2007) applied the Gamma Algorithm where calibrated probabilities are obtained as a linear combination of the individual PDFs associated with each ensemble member. Peña and Van den Dool (2008) analyzed the performance of ridge regression methods for prediction the sea

surface temperature (SST) from multiple seasonal ensemble prediction systems data. Wilks (2009) introduced the concept of extended logistic regression, where a monotonic function of the threshold itself is included as a predictor within the regression equation, and Veenhuis (2013) used a MOS variation called Ensemble Kernel Density Model Output Statistics that uses MOS equations and spread–skill relationships to generate calibrated probabilistic forecasts.

Ruiz and Saulo (2012) developed assessments using several different techniques that were applied to precipitation forecasts since they are one of the most challenging and least accurate products available from numerical weather prediction (EBERT, 2001; STENSRUD; YUSSOUF, 2007). They concluded that methods using only MEAN show very similar results in terms of calibration and skill and that the use of ensemble members does not lead to an improvement of skill or reliability. It should be remarked, however, that forecasts using MEAN are better than those derived from a single deterministic forecast because the errors were found smaller (RUIZ et al., 2009).

The methods here mentioned are sometimes applied alone and sometimes are combinations of them: LR, LogReg, weighted histogram-based rank, simple averaging bias correction, BMA, reforecast analogs; multivariate linear regression, ridge regression, Gaussian regression, and Artificial Neural Network (ANN).

Hamill and Colucci (1998), Eckel and Walters (1998) and Buizza et al. (2005) also reinforced that systematic errors can be reduced by using post-processing techniques in order to produce more calibrated probabilistic forecasts and that calibration of ensemble systems that suffer from underdispersion is highly necessary. It seems clear that post-processing of ensemble forecasts is considered a necessary step to improving forecast quality and benefiting end users.

Moreover, nowadays post-processing is highly recommended instead of improving the model. Hamill et al. (2000) said that mitigation through the improvement of assimilation procedures and model formulations alone poses a significant long-term challenge to the research community, especially for ensemble forecasts because of the increased dimensionality of the prediction system. Moreover, Reynolds et al. (1994) reported that most forecast errors are attributable to predictability error growth rather than model deficiencies, i.e., problems with the IC. Simmons et al. (1995) also showed that improving model itself it's not trivial when they experienced that forecasts at higher resolution amplify initial errors more quickly than previous lower-resolution versions of the model.

2.8. Machine Learning

The post-processing techniques mentioned in the last section are real examples of a branch of artificial intelligence called machine learning (ML), which can be defined as a "Field of study that gives computers the ability to learn without being explicitly programmed" (SAMUEL, 1959).

Roughly, ML techniques execute a learning (or training) phase that builds a model from a training dataset. Then, an optional cross-validation phase can be executed in order to establish the performance of the model from a validation dataset. After that, a generalization phase, in which unknown data (from a test dataset) are applied to the model expecting approximating forecast values and actual values. Commonly, some metric is applied to compute the difference between forecast and actual data.

The term "generalize" means the ability to classify (or categorize) and predict (or regress) unknown data, i.e., measure how good the solution (the model) is when it meets data outside the training set. The term "learning" refers to a process that produces, from a known data (training set), a suitable "ability" to solve a problem according to the importance that each input feature has for the computation of the global solution, in other words, produces a mathematical

forecasting model. Technically, it is a process that searches for the optimized set of parameter values that minimizes the squared error between estimated and actual data (FAUSETT, 1994). This characteristic is particularly very interesting because it avoids the need to specify a set of rules as it happens in the ordinary programming paradigm.

The generalization task can be applied for two reasons: a) classification, which tries to categorize class labels represented by discrete values and includes Decision Tree (DT), Bayes Rule-Based, SVM and Artificial Neural Network (ANN) as examples of classification algorithms; and b) regression, which tries to find the most accurate way to predict an output given some input and includes LR, LogReg, Polynomial Regression and also ANN as examples of regression algorithms.

In addition to the type of generalization`s task there is a need to know the type of the learning process that the system will deal with, which can be inferred from the available data set and from the user particular needs. With respect to the learning type, ML techniques can be classified in: a) supervised learning, i.e., a type of learning that is conducted by a "teacher", in other words, there must be a subset of model features that maps a correct answer (output) to every input pattern; b) unsupervised learning, in which there is no correct answer in the dataset (mainly used for discovering hidden patterns in data), and c) reinforcement learning, in which is given just a partial set of answers and a grade of confidence of it (ABU-MOSTAFA et al., 2012).

ANN is one of the most used supervised learning ML techniques in science for both classification or regression tasks due to its capacity to approximate highly nonlinear functions in order to model complex multivariate dynamical relationships among the associated physical-environmental variables (ABRAHART; SEE, 2007).

It is also particular attractive because requires no prior knowledge of the nature of input/output relationship (GARDNER; DORLING, 1998) neither any human

subjective interference (NICULESCU, 2003), making it flexible enough to accommodate various possible additional constraints. These remarkable qualities have led ANNs to be considered "universal approximators" (CYBENKO, 1989 and references therein).

These characteristics have made ANN to be a suitable tool to create classification and regression models for the atmosphere`s nonlinear dynamical system, since the early 1990's (RIAD et al., 2004; HSIEH, 2009).

2.9. **Artificial neural network (ANN)**

ANNs are nonlinear statistical models whose structure and mechanics are inspired on the biological architecture of human's neurons, the fundamental information-processing unit to the operation of a neural network. The neuron biological elements, which form the basis for designing ANNs and simulate the capabilities of leaning and making associations, are (HAYKIN, 1999):

- Soma: the neuron cell body. In ANN, a processing node (or unit);

- Dendrites: connections in which soma receives signals, i.e., how inputs are accepted by soma. In ANN, the summing function;

- Axons: connections that sends signals out of a neuron. In ANN, the activation function that limits the output signal amplitude of a neuron;

- Synapses: electrochemical contact between neurons. In ANN, weighted links connecting inputs and nodes, i.e., values that represent the "knowledge" acquired from the ANN's learning process.

ANN can be viewed as a structure of nodes interconnected by weighted edges that can be arranged in different manners (architectures) according to the characteristics of the problem to be solved. Two of the most used ANN architecture families for supervised learning are: i) feed-forward neural network (FF-NN), in which nodes have no loops, and ii) recurrent (or feedback) neural

network (R-NN), in which loops occur because of feedback connections. Figure 2.6 shows some examples of ANN architectures.

Figure 2.6   Taxonomy of feed-forward and recurrent network architectures.
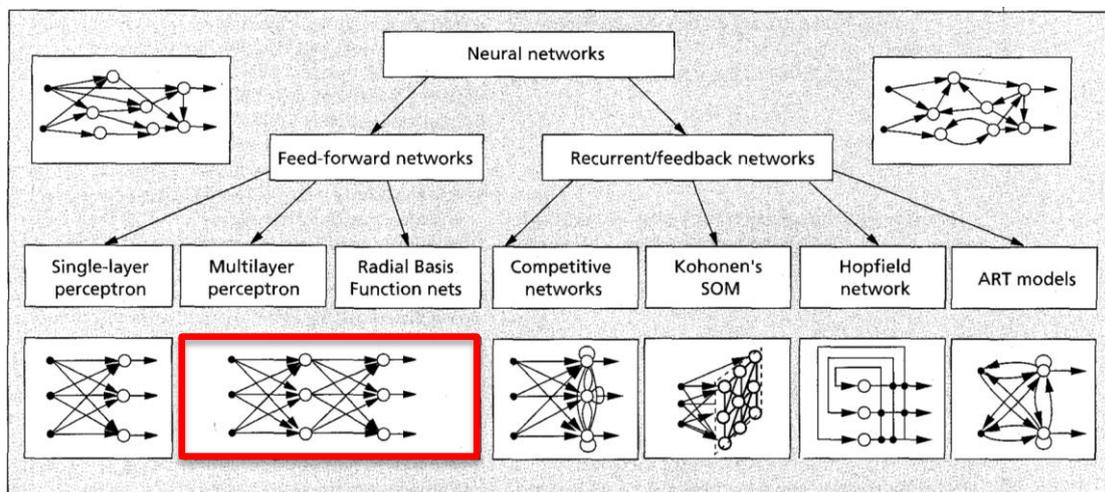
In red, the MLP ANN type.

Source: Jain and Mao (1996).

Multi-Layer Perceptron (MLP) network models are the most popular network architecture used for mathematical modeling in several fields of science and engineering [Yilmaz and Kaynar (2011), Saroha and Aggarwal (2014), Dawson and Wilby (2001), and Chadwick et al. (2011)].

As illustrated in Figure 2.7, MLP's architecture consists in a set of inputs in the input layer followed by sets of neurons organized into one or more intermediate (or hidden) layers, and one output layer. Every neuron is interconnected, in a unidirectional and forward way, to every neuron in the subsequent layer, from the input layer towards the output layer (Bishop, 1995), namely feed-forward (FF) neural networks. MLP's topology is commonly written as a vector of integers collapsed by a dash ('-') where the first element is the number of input units, followed by a vector specifying the number of hidden neurons in each layer, and terminating with the number of output units, as show the legend of Figure 2.7.

Figure 2.7   Example of MLP topology.
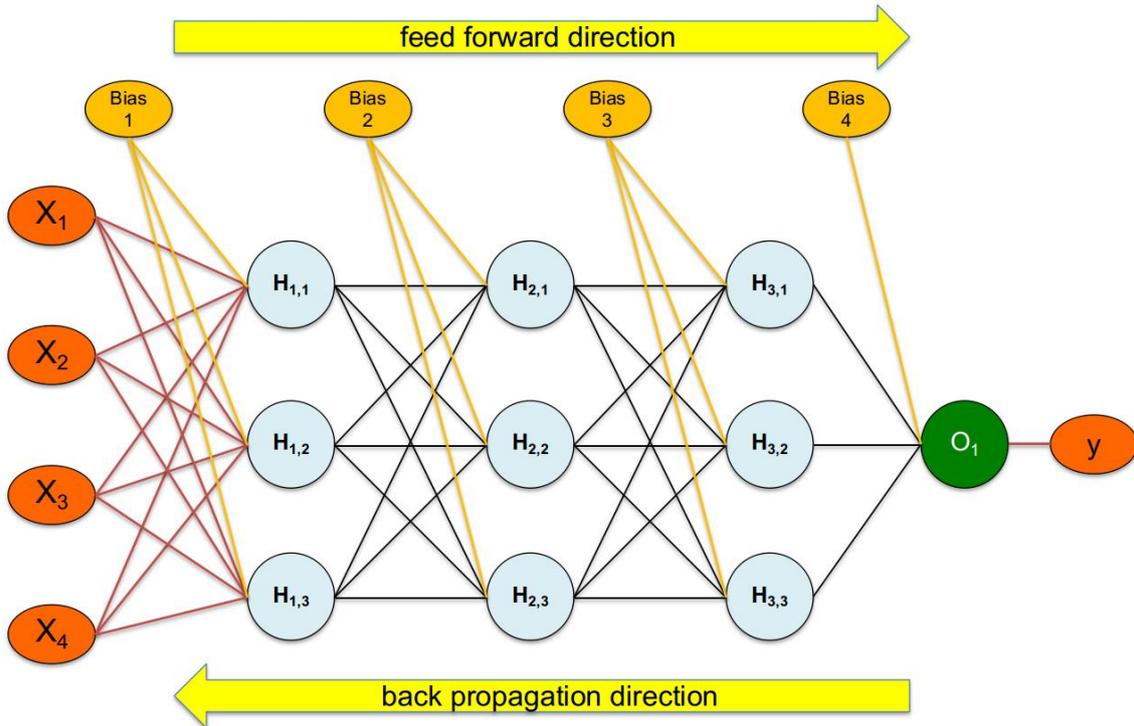
Example of a 4-3-3-3-1 MLP topology composed by four inputs in dark-orange ellipses, three hidden layers with three nodes each in light-blue, and a single output node in green. The biases of each layer are in light-orange ellipses. Colored lines represent the synapses that link each node to every node in the next layer. Subscript of node represents the index of the layer and of the node within it.

Source: author's production.

The model of the neuron describes the output behavior of a neuron, i.e., refers to the manner that the input signal is transferred outside the neuron. McCulloch and Pitts (1943) described the biological neuron as a Threshold Logic Unit (TLU) with L binary (0 or 1) inputs and 1 binary output. It is represented in Figure 2.8 and operates as follows:

    a) The input weights (w_0, w_1, w_2, ..., w_n) are initialized with small random values. Random values help to break the symmetry of the strength of the input signal, if all weights are the same, all units in hidden layer will have the same value of the input and they will compute the same gradients during back-propagation. Small weight values are used in

order to bring the input closer to the neuron output signal, which can lay within [0,1] interval.

b) A weighted sum of the products (namely linear combination) of the input signals composed by the bias term (x_0, which is commonly set to 0 or 1) plus the inputs (x_1, x_2, ..., x_n) and the associated input weights is computed according to the formula:

$$v = \sum_{i=1}^{n} w_i x_i \qquad (1.1)$$

Where *v* is the final summation value, *n* is the total number of input signals, *w* and *x* are the input and associated weight according to the index *i*, respectively. The bias is used to increase or decrease the final summation value according to its sign, and represents a constant influence of some unknown effect in the computation of the neuron's value, something unknown within the relationship.

c) The final value of the Formula 1.1 is submitted to the activation function (φ) that computes the final output according to the threshold (or Heaviside) step function, which assumes only *1 if v >= 0* or *0* otherwise.

Figure 2.8   McCulloch and Pitts-based model of a neuron.
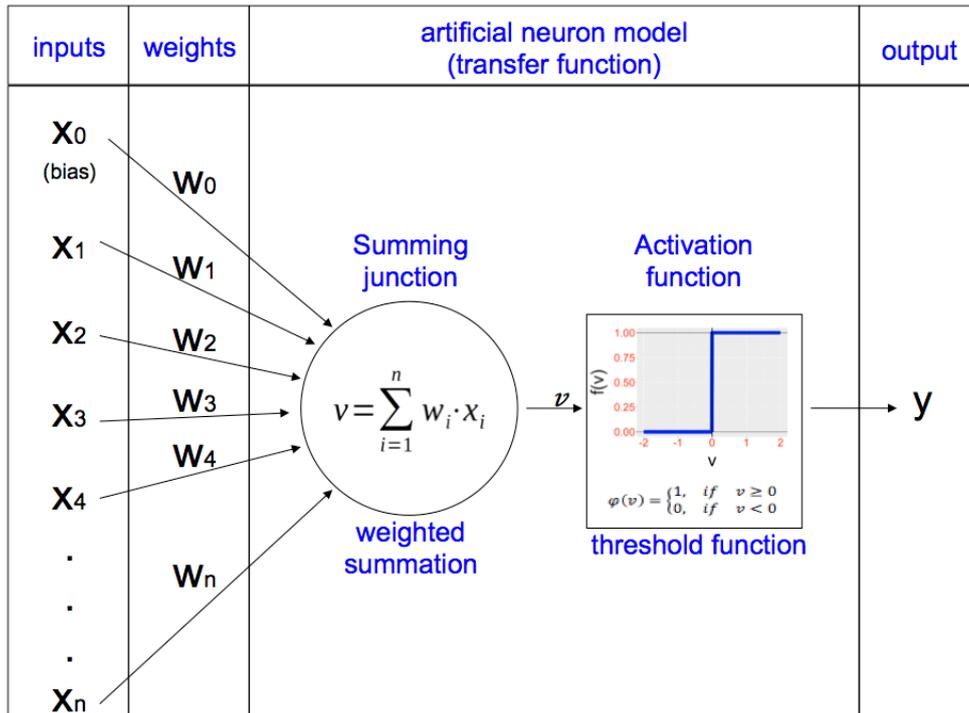
The manner the input is transferred to outside the neuron is ruled by the transfer function. The summing junction represents a way to combine the input data (*x*) and the associated weights (*w*), resulting in the value *v*. It is then passed to the activation function that processes it and generates the output *y*.

Source: author's production

The activation function (or output or squash function) is a nonlinear function used to map the linear input (from the linear combination) to a nonlinear value in a different but limited domain, typically between [0,1] or [-1,1], which makes easier the task of classifying the sample. The TLU, however, has some drawbacks because it is a discontinuous function and limits the output range to only two values, 0 and 1. Concretely, an output equals to zero means that the input should not be considered important to final system (has zero influence) and one it should be. Although enough for binary classification, this behavior brings some issues while modeling nonlinear real world problems, such as the fact that small adjusts on the weights, in general, are not reflected in the output.

Other models were derived from TLU in order to adapt it to particular needs. The main modification was to use smooth continuous activation functions in all units instead of using TLUs. Available options include Piecewise-Linear Function, S-shaped functions like Hyperbolic Tangent and Sigmoid Function (HAYKIN, 1999), and Rectified Linear Unit (ReLU), which Nair and Hilton (2010) reported as a very promising activation function option. Table 2.2 shows more details about activation function.

Table 2.2   Behavior and rules of some activation functions.

Rules and comments	Behavior
Threshold function $$\varphi(v) = \begin{cases} 1, & v \geq 0 \\ 0, & x < 0 \end{cases}$$ Step-like function. Output is only 0 or 1. Mostly used for binary classification.	
Piecewise-Linear function $$\varphi(v) = \begin{cases} 1, & v \geq +\dfrac{1}{2} \\ v + \dfrac{1}{2}, & \dfrac{1}{2} > v \geq -\dfrac{1}{2} \\ 0, & v \leq -\dfrac{1}{2} \end{cases}$$ Linear continuous output when *v* is not large and threshold-like otherwise.	
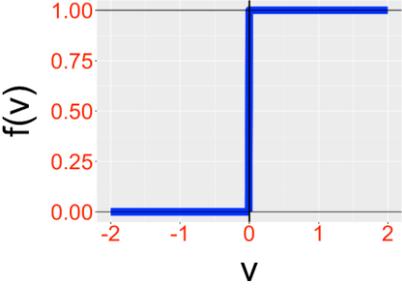
Table 2.2   Behavior and rules of some activation functions (cont.).

Hyperbolic tangent function $$\varphi(v) \equiv tanh(v) \equiv \frac{e^v - e^{-v}}{e^v + e^{-v}}$$ Nonlinear continuous output from -1 to 1.	
Sigmoid function $$\varphi(v) = \frac{1}{1+e^{-av}}$$ Nonlinear continuous output from 0 to 1. *a* is the slope parameter, whose variation changes the slope of the function and impacts in the ANN system concerning training, convergence and generalization performance (SHARMA; CHANDRA, 2010). Higher values generate more step-like functions and lower values generate more linear-like functions.	
ReLU function $$\varphi(v) = \begin{cases} v * a, & v < 0 \\ v, & v \geq 0 \end{cases}$$ Linear output when v ≥ 0. For v < 0 a slope 'a' can be used to avoid 0 values for f(v).	
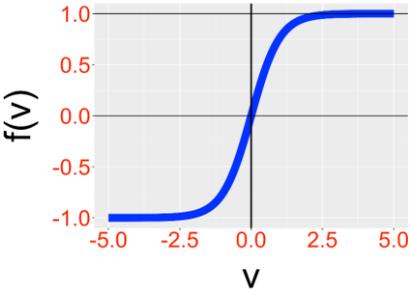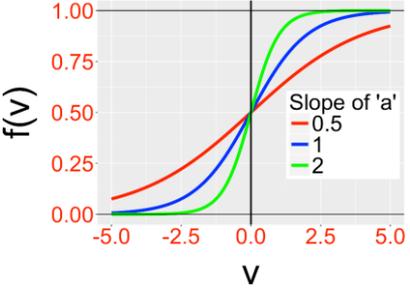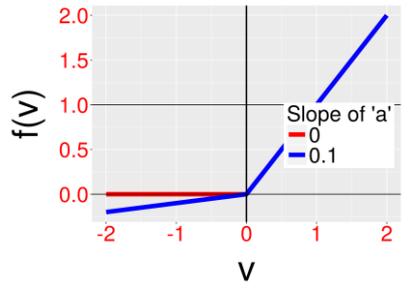
Plots in the right show, in blue line, the behavior of the function given by the rules in the left.

Source: author's production

The first practical form of ANNs was introduced by Rosenblatt (1958) and was named perceptron. It consists in inputs directly linked via weighted connections to a single output layer composed by TLU`s neurons and it was proposed as a model for visual pattern recognition.

However, the limits of the perceptron model were found and it was proven that perceptrons could not be trained to recognize many classes of patterns since the single-layer perceptron could only solve problems that are linearly separable (MINSKY; PAPERT, 1969), and many real-world problems are not linearly separable problems. MLP arose as a way to overcome perceptron`s limitations by introducing extra layers of nodes (hidden layers) between the input and output layers. Together, these neurons become capable of solving complex multipurpose nonlinear systems.

Regarding MLP's training algorithm, the most popular method used to define how the model is adjusted to become closer to the optimal global solution is the back-propagation (BP) (RUMELHART; MCCLELLAND, 1986). BP (or generalized delta rule) (HAYKIN, 1999) is based on the fact that a desired output is given to the system for each input pattern but, commonly, it is different of the actual output. The learning process, i.e., the approximation of forecast and actual values, is done by computing the difference (squared error) between them followed by a backward update of the synaptic weights (namely free variables) on the prior layers, i.e., a backward propagating the error.

The FF and BP processes can be summarized in the following stages:

a) weights initialization: all the weights of the synapses are initialized with small random values;

b) sample feed-forwarding: each hidden node in the first layer computes its activation level (concerning the neuron model) by receiving the input data from each sample in the training set and, iteratively, sends it to the next hidden layer until the last one is reached;

c) network error computation: the difference between each target (ideal) and each final output (actual) is computed according to the squared error function and then summed up.

$$NetError = \sum_{n=1}^{\#outputs} (target_i - output_i)$$

(1.2)

d) Back-propagation: the associated error of each weight is computed in backwards, i.e., how much a change in each individual weight affects the total network error. Technically, each partial derivative of *NetError* with respect to *weight$_i$* (also known as the gradient with respect to *weight$_i$*), is computed with

$$\frac{\partial E_{total}}{\partial w_i}$$

(1.3)

A partial derivative can be viewed as the slope of the tangent line towards the smaller error of the weight. The direction of the step is indicated by the sign and how much the weight need to be adjusted is indicated by the value itself (magnitude). Weights are updated with their respective partial derivatives in backwards and the algorithm loops back and forth until the maximum number of iteration (or epochs) is reached for each sample and the ANN model is determined.

This process repeats until an arbitrary level of generalization, verified by a stop criterion is reached. In ANN, the best-for-the-moment solution is achieved when the accuracy (error) reaches a threshold or the error becomes invariant at some level for some amount iterations, according to arbitrary needs.

It turns out that, for a fairly steep gradient produced by the formula 1.3, the step towards the solution can be too large, overshooting the expected value in almost the same magnitude of the weight value. The weight can, consequently, stay oscillating back and forth the optimum value forever. The use of a learning

rate parameter in the BP algorithm is to determine, along with the magnitude of the gradient, a weight delta small enough to avoid this overshoot. The problem that arises is that the learning rate makes the training phase to be very slow because the steps become smaller and smaller as it approaches the solution (process known as gradient descend). One way to get around this situation is adding a *momentum* term. *Momentum* adds a fraction of the previous weight update to the current one in a manner that increases the size of the steps taken towards the minimum if the gradient keeps pointing in the same direction, otherwise decreases.

BP algorithm requires a good adjustment of these parameters in order to work well. It is also high dependable on the value of the random weights, since a "wrong" initialization of them can lead the solution to a local, and not the best solution. Despite of the large use of the BP algorithm, some other alternatives are worth being considered: the quasi-Newton algorithm (DENNIS; MORÉ ,1977; CORTIVO; CHALHOU; VELHO, 2012), and the resilient back-propagation (RPROP) (RIEDMILLER; BRAUN, 1993).

Although RPROP algorithm is more complex than BP algorithm, it has some advantages, such as: faster training and better generalization (AHMAD, 2008; KISI; UNCUOGLU, 2004), and the independence of free parameters like the learning rate, and the optional momentum term, and of the synapses weights.

RPROP algorithm has two main differences when compared to BP algorithm. It uses only the sign of the gradient to update the weights and maintains individual weight deltas for updating the weights. For each particular weight, if the previous and current partial derivatives have the same sign, it means the weight value hasn't overshot the solution and the function will move even faster in the same direction, since the amount that was added to the weight delta on the last iteration is used to increase the current. Oppositely, if the signs of the previous and current partials have changed, an overshooting has occurred and the algorithm undo the changes by moving the values back to the previous iteration

and decrease the amount added on the last iteration, aiming not to move so far next time.

Once ANN's architecture is chosen, it can still have its structural composition configured in different ways with respect to the amount of neurons at each layer (topology). The architecture, topology and neuron model of ANN can significantly impact on its performance, the estimate of the optimal ANN model must consider:

- the numbers of input and output units;

- the number of training cases;

- the complexity of the function or classification to be learned;

- the type of hidden unit activation function;

- the training algorithm;

- and other ANN's fine tuning.

ANN's training phase can be time-consuming but its application is not. Training is usually performed only once and each application is practically instantaneous since there are only simple computation of floating point additions and multiplications.

This section focuses only in the ANN aspects that were considered relevant to the present work, further details can be found in Fausett (1994), Bishop (1995), and Haykin (1999).

2.10. CPTEC'S ensemble prediction system (CPTEC-EPS)

The ensemble prediction system used in this study is based on the CPTEC-AGCM (CAVALCANTI, 2002), which is operational since November 1994. Coutinho (1999) adapted the methodology described by Zhang and

Krishnamurti (1999) in order to produce an ensemble of forecasts that are distinguishable by analysis' perturbations. The meteorological variables present in the analysis are showed in Table 2.3.

Table 2.3   List of CPTEC-AGCM`s analysis meteorological fields.

Variable	Description	Unit
psfc	SURFACE PRESSURE (p)	HPA
uves	SURFACE ZONAL WIND (u)	M/S
uvel	ZONAL WIND	M/S
vves	SURFACE MERIDIONAL WIND (v)	M/S
vvel	MERIDIONAL WIND	M/S
vort	VORTICITY	1/S
fcor	STREAM FUNCTION	M2/S
potv	VELOCITY POTENTIAL	M2/S
zgeo	GEOPOTENTIAL HEIGHT	GPM
psnm	SEA LEVEL PRESSURE	HPA
tems	SURFACE ABSOLUTE TEMPERATURE	K
temp	ABSOLUTE TEMPERATURE (t)	K
umrs	SURFACE RELATIVE HUMIDITY	NO DIM
umrl	RELATIVE HUMIDITY	NO DIM
umes	SPECIFIC HUMIDITY (q)	KG/KG
agpl	INST. PRECIPITABLE WATER	KG/M2
tsfc	SURFACE TEMPERATURE	K
tp2m	TEMPERATURE AT 2-M FROM SURFACE	K
u10m	U-WIND COMPONENT AT 10-M	M/S
v10m	V-WIND COMPONENT AT 10-M	M/S

Source: author`s production

CPTEC-EPS provides fifteen members, one from the analysis (the unperturbed IC) plus fourteen members from perturbed ICs, which are generated according to the following steps:

a) random small perturbations are added to fields temperature (t), wind components (u,v), surface pressure (psfc) and specific humidity (q) of the control (unperturbed) analysis. These perturbations obey a Gaussian

distribution with mean zero and standard deviation comparable to that of the forecast error verified in 3h forecast, which corresponds to 0.6⁰K for t, 3 m.s^{-1} for u,v, 1 hPa for psfc and the q obeys Table 2.4;

Table 2.4   Standard deviation values proposed by Mendonça and Bonatti (2009).

σ	Std dev	σ	Std dev	σ	Std dev	σ	Std dev
1	0.77	8	0.98	15	0.90	22	0.00
2	0.78	9	1.14	16	0.75	23	0.00
3	0.78	10	1.27	17	0.49	24	0.00
4	0.78	11	1.37	18	0.26	25	0.00
5	0.80	12	1.35	19	0.12	26	0.00
6	0.82	13	1.18	20	0.05	27	0.00
7	0.88	14	1.05	21	0.02	28	0.00

Specific perturbation of the humidity (*q*) for each sigma level (*σ*) of CPTEC-EPS. Values are multiplied by a factor of 10^3.

Source: Mendonça and Bonatti (2009).

b) the resulting randomly perturbed analysis and the control analysis are used to integrate the model up to 36h, having the results saved at every 3h. However, the first 6h of model integration is discarded in order to allow a self-adjustment of the model to the perturbed initial conditions and consequently develop more balanced forecast perturbations;

c) the time series of the difference field forecasts is constructed by subtracting the control forecast from the perturbed forecasts at each time increment of 3h;

d) an EOF analysis is performed for the time series over a tropical domain, in order to determine the eigenvectors whose EOF coefficients increase rapidly with time. These eigenvectors are considered as the EOF perturbations. Figure 2.9 shows examples of the evolution of three coefficients (amplitude) for different eigenmodes along time;

Figure 2.9   Example of EOF coefficients for different eigenmodes.
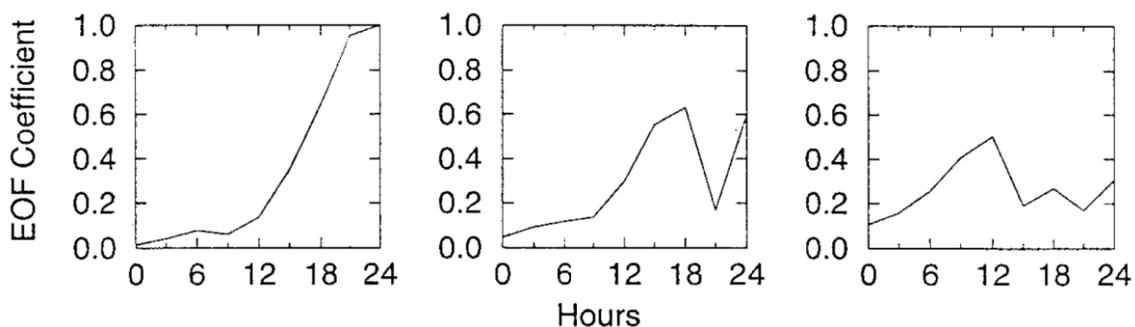
From the highest coefficient eigenmode to the 3rd, from left to right.

Source: Adaptated of from Zhang and Krishnamurti (1999).

e)   adding and subtracting the highest perturbation to and from control analysis produces an ensemble of two initial perturbed states (positive and negative members, respectively) in order to achieve the maximum and the minimum variance caused by the perturbation;

f)   the resulting highest and lowest variance analysis and the control analysis are used to fully integrate the model up to 360h, whose archiving occurs at every 6h.

CPTEC-EPS is in operational mode since October 2001 and runs twice a day at 00h and 12h with a forecast horizon defined from 6h to 360h ahead. Its archiving generates 4.1 GB/run or ~3 TB/year on the Tupã (Cray XE6 supercomputer) archiving system.

The horizontal spectral truncation used in the CPTEC-AGCM is T126 (i.e., triangular truncation at zonal wave number 126). To avoid aliasing in the solution of nonlinear terms of model equations it is necessary to use approximately a number of points in the zonal direction equivalent to 3 times the shortest wave number considered, which corresponds to approximately a 0.948 degrees in the geographic coordinate resolution, i.e., a low resolution of ~100 km x ~100 km, making X-axis to vary from 1 to 385 grid points (longitude 0° to

34

360°) and Y-axis to vary from 1 to 192 grid points (latitude -89.2842° to 89.2842°) (shown in Figure 2.10). In the vertical, the atmosphere is divided into 28 sigma layers (L28). CPTEC-AGCM output variables are listed in Table 2.5.

Figure 2.10   Horizontal domain of CPTEC-AGCM.
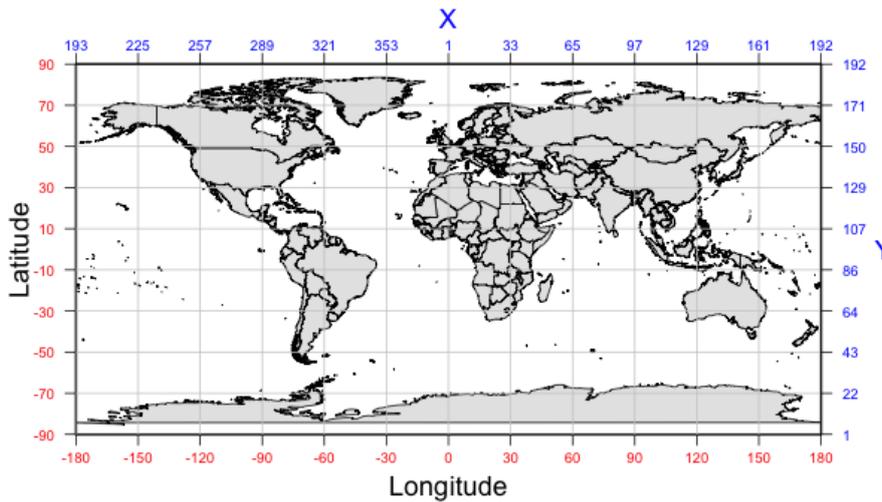
Bottom and left axes (in red) are the geographic coordinates of the domain. Top and right axes (in blue) approximately represents the correspondent X-Y grid coordinate.

Source: author's production

Table 2.5   List of CPTEC-AGCM`s prognostic fields.

Variable	Description	Unit
psfc	SURFACE PRESSURE (p)	HPA
uves	SURFACE ZONAL WIND (u)	M/S
uvel	ZONAL WIND	M/S
vves	SURFACE MERIDIONAL WIND (v)	M/S
vvel	MERIDIONAL WIND	M/S
vort	VORTICITY	1/S
fcor	STREAM FUNCTION	M2/S
potv	VELOCITY POTENTIAL	M2/S
zgeo	GEOPOTENTIAL HEIGHT	GPM
psnm	SEA LEVEL PRESSURE	HPA
tems	SURFACE ABSOLUTE TEMPERATURE	K
temp	ABSOLUTE TEMPERATURE (t)	K
umrs	SURFACE RELATIVE HUMIDITY	NO DIM

Table 2.5   List of CPTEC-AGCM`s prognostic fields (cont.).

umrl	RELATIVE HUMIDITY	NO DIM
umes	SPECIFIC HUMIDITY (q)	KG/KG
agpl	INST. PRECIPITABLE WATER	KG/M2
tsfc	SURFACE TEMPERATURE	K
tp2m	TEMPERATURE AT 2-M FROM SURFACE	K
u10m	U-WIND COMPONENT	M/S
v10m	V-WIND COMPONENT	M/S
uvmt	TIME MEAN ZONAL WIND (U)	M/S
vvmt	TIME MEAN MERIDIONAL WIND (V)	M/S
prec	TOTAL PRECIPITATION	KG/M2/DAY
neve	SNOWFALL	KG/M2/DAY
usst	SURFACE ZONAL WIND STRESS	PA
vsst	SURFACE MERIDIONAL WIND STRESS	PA
cbnv	CLOUD COVER	0-1
tgsc	GROUND/SURFACE COVER TEMPERATURE	K

Source: author`s production

Since 2001 up to current days, no further significant changes occurred in the operational version of the CPTEC-EPS, however, some improvements concerning perturbation method are in the operationalization phase and are described below.

Based on previous studies about the underdispersiveness, i.e., ensemble spread smaller than the Root Mean Squared Error (RMSE) of the ensemble mean, Mendonça and Bonatti (2009) evaluated five experiments using statistical scores like pattern anomaly correlation (PAC), RMSE, and standard deviation spread in order to assess changes on the method that generates the initial perturbation of the temperature (t) and wind components (u, v) and also to assess the inclusion of perturbations to surface pressure (p) and specific humidity (q) fields, in six regions. As the original method, these perturbations also obey a Gaussian distribution with mean zero and standard deviation comparable to that of the 3h forecast error.

Results showed that, on overall, the application of the new method to perturb mid-latitude variables, produced positive impact on the quality of CPTEC-EPS and alleviated the underdispersion of the system. This fact has encouraged CPTEC's staff to make it the new operational version as well as the author of this work to use it as reference.

2.11. CPTEC-EPS`s performance

The new CPTEC-EPS data are already used in this work aiming to make improvements in the foremost system. A preliminary analysis of its spread in Figure 2.11 presents CPTEC-EPS rank histograms from La Plata Basin region for the months JFMA and OND from 2009 to 2012. Each plot represents a lead-time within the forecast horizon, from 24h to 360h, incremented by 72h. The last one represents the rank histogram considering all lead times.

According to this plot, there is a clear lower bias in the ensemble members spread as most observations fall into the last bin. Also, the gap between the last bin and the others decreases as the lead-time approximates to 360h forecast, which corroborates the assumption that an EPS has better performance from mid-range forecasts. Finally, considering all lead times, the spread still shows a lower bias, which means that the overall performance of MEAN has real evidences that it is not good. This picture shows that there is a dispersion problem and that some improvement is needed. Despite there is no clear deficiency of the MEAN, the results of this work show, by comparisons of pure MEAN vs. ANN MEAN, that the performance, whatever it is, can be improved using the proposed methodology.

Figure 2.11   Rank histograms from CPTEC-EPS from La Plata Basin region for the months JFMA and OND from 2009 to 2012.
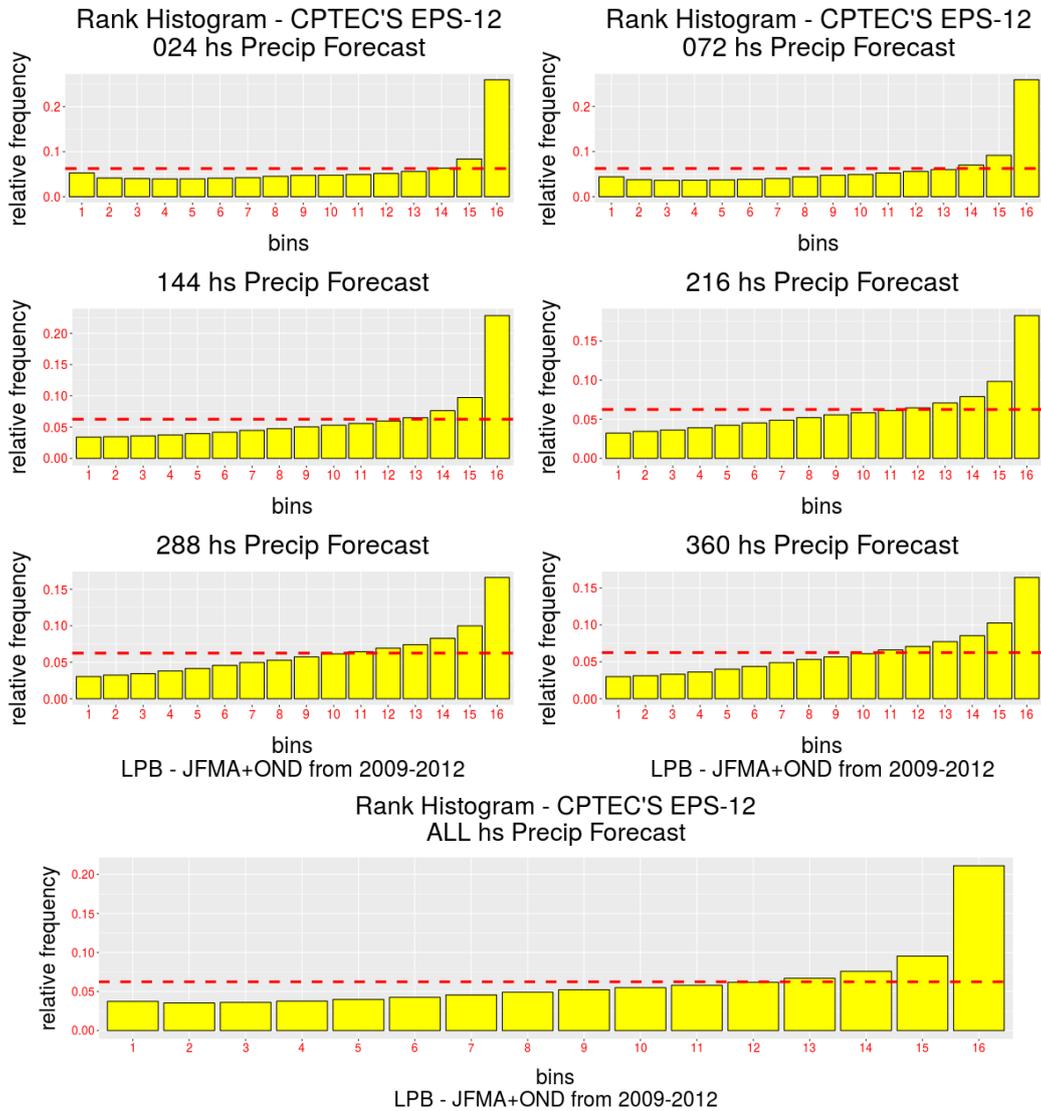
Each plot represents a lead-time within the forecast horizon, from 24h to 360h, incremented by 72h. The last one considers all lead times. The red dotted line is where all bars should be in a flat shape.

Source: author`s production

2.12. **Ensemble mean calibration related work**

MEAN is computed to aggregate reliability to forecasts by minimizing the difficulties caused by ensemble dispersion problems (KALNAY, 2003, p. 26).

However it also carries the underdispersiveness problems from members, which characterizes a systematic error (bias) and consequently degrades the forecast accuracy (HAMILL; COLUCCI, 1997; STENSRUD; YUSSOUF, 2003; BUIZZA et al., 2005).

Despite of some authors have found that MEAN is more better than any individual forecasts members (HOU et al., 2001; GRIMIT; MASS, 2002) it is not as good as the best bias-corrected forecasts (RAFTERY at al., 2005). To obtain it, therefore, it seems necessary to carry out some form of statistical post processing (HAMILL; COLUCCI, 1997, 1998; BUIZZA et al., 2005).

Nowadays post-processing ensemble prediction is considered a necessary step to calibrate the forecast and, as seen, several works that use many different post-processing techniques has been proposed. However, almost all of them are applied over ensemble members, aiming to calibrate the probabilistic forecast and thus the MEAN. Experiments concerning calibrating EM in a direct manner appear to be relatively rare.

Based on such facts and that MEAN is considered a crucial component for the whole forecasting process it was decided to investigate means to try to make some progress in performance skills of the MEAN produced by the CPTEC's EPS by post-processing it directly.

Even though this work uses EPS variables as a way to improve the accuracy of the precipitation forecast, which is one of those variables, this approach looks closely similar to applying soft computing techniques in order to forecast precipitation, since both of them use meteorological variables. The difference is that the later is primarily based on actual meteorological measurements and the former is based on prognostic fields.

2.13. **Precipitation forecasting through ML related work**

The task of predicting the precipitation through the use of ML techniques is extensively explored as demonstrate the following few works.

French et al. (1992) used several ANN configurations for forecasting rainfall in space and at 1h lead-time, using the rainfall field as input and output. They noticed that it is capable of learning the relationship space vs. time mainly when a relatively large number of hidden nodes are utilized (it was from 15 to 100 nodes), as it provides means for storing higher order relationships necessary for adequately abstracting the process. The ANN accuracy, measured in RMSE, showed similar results to two other well-known physics-based methods of short-term forecasting.

Hall et al (1999) used data from a regional model to feed an ANN in order to forecast the probability of precipitation (PoP) and the quantitative precipitation forecast for the Dallas–Fort Worth, Texas, USA area. They got 95% of linear correlation between forecast and OBS.

Ramirez, Velho and Ferreira (2005) used data from CPTEC's ETA model to construct a nonlinear mapping between its outcome and surface rainfall in São Paulo/Brazil. They applied LR and ANN and concluded that the forecasts of the later were superior to the ones obtained by the LR.

Hung et al. (2009) applied ANN for forecasting rainfall of Bangkok, Thailand. Different network types and configurations combined with different types of input information were tested and a Feed-Forward ANN using hyperbolic tangent transfer function achieved the best generalization of rainfall forecast. The input information was a combination of meteorological parameters such as relative humidity, air pressure, wet bulb temperature and cloudiness; the rainfall at the point of forecasting and rainfall at the surrounding stations were also used. Results revealed that ANN forecasts were better compared to simple persistent method.

Dastorani et al. (2010) also explored the ANN capabilities comparing its precipitation prediction against Adaptive Neuro-Fuzzy Inference Systems (ANFIS) in a single meteorological station in Iran. Different architectures of ANN and ANFIS models were applied, as well as various combinations of meteorological parameters as input, including 3-year precipitation moving average, maximum temperatures, mean temperatures, relative humidity, mean wind speed, maximum wind direction and evaporation. ANN types include, Multi Layer Perceptron (MLP), Generalized Feed Forward (GFF), Modular Neural Network, Recurrent Network (RN), and Time Lagged Recurrent Network (TLRN). At the end, RN and TLRN showed better performance for this application.

Moustris et al. (2011) used a 7-5-1 MLP-BP ANN to forecast precipitation of Greece, measured the performance with RMSE and concluded that it had better results than classical statistical methods. The ANN configuration was achieved via test and error approach.

The usage of ANN variations in forecasting rainfall is wide and does not seem to end so soon, other scientists also explored it as well as other techniques to accomplish this task.

Ingsrisawang et al. (2008) used three Machine Learning (ML) techniques available from WEKA (HALL, 2009) for short-term rain forecasting: DT, ANN and SVM. These techniques were applied to classify and predict rainfall forecasts. The results showed that there was no final consensus of which method was the best in overall since the performance varies according to the task (classification o regression), to the number of predictors used (3 to 14), lead-time forecast data (same day to 2 days ahead), and spatial location of the station.

Wu et al. (2010) explored a bunch of components for forecasting daily and monthly rainfall time series seeking a relatively optimal data-driven model based three aspects: model inputs, modeling methods, and data-preprocessing

techniques. Firstly, they applied 7 linear and nonlinear techniques in order to identify the best set of inputs and the ordinary Linear Correlation Analysis (LCA) was capable to perform better. Then, they proposed a model called Modular ANN whose performance was compared to ordinary ANN, K-Nearest Neighbors (K-NN) and LR. The variation of using a preprocessed and non-preprocessed (normal mode) dataset was also explored. Data-preprocessing techniques includes moving average (MA), principal component analysis (PCA), and singular spectrum analysis (SSA). In normal mode, results indicate that MANN performs the best among all prediction models, but this improvement is not significant in monthly rainfall forecasting compared to ANN. The final conclusion is that the proposed optimal rainfall forecasting model can be derived from MANN coupled with SSA.

Hong (2008) used a hybrid model of RNNs (for forecasting rainfall data) and SVM (for solving time series problems), namely RSVR, to forecast rainfall depth values and the results reveal that this model provides a promising alternative. Gagne et al. (2013) applied LogReg and RF to produce calibrated probabilistic forecasts from the raw ensemble data correcting the systematic biases in the ensemble precipitation forecast.

Radzuan et al. (2013) compared three techniques namely LogReg, DT, and RF for precipitation forecast. They concluded that every model has advantage and weakness in making rainfall forecasts. LogReg showed good on ranking variables according to its importance to the process, DT produced a visual model, thus easier to be understood and RF helps to display variable trees in more details.

Soft computing for precipitation forecasting can also be done for predicting seasons rather than single daily events, as have done Navone and Ceccatto (1994). They used hierarchical ANN to predict summer monsoon rainfall over India based on the fact that it can correlate nonlinear relationships. The results showed that the forecasting capabilities was remarkably improved compared to

conventional methods. Also for longer periods, Freiwan and Cigizoglu (2005) estimated monthly precipitation amount also using different configurations of ANNs and stated that the estimations of rainfall resulted in favor of the ANN compared to conventional stochastic models.

El-Shafie et al. (2011) used ANFIS and ANN techniques for forecasting monthly rainfalls in Malaysia. They also used trial and error approach to determine the optimum structure for the ANN. Similarly to other works, different rainfall combinations were used as inputs. They measured the results using 5 different metrics: RMSE, Correlation Coefficient (2R), Nash Sutcliffe coefficient, gamma coefficient (GC) and Spearman correlation coefficient and the final result was that ANFIS had higher accuracy vs. ANN.

2.14. Illustration of foundation and goal

Figure 2.12 illustrates the summary of the foundation and the goal of this Thesis.

Figure 2.12   Foundation and goal.
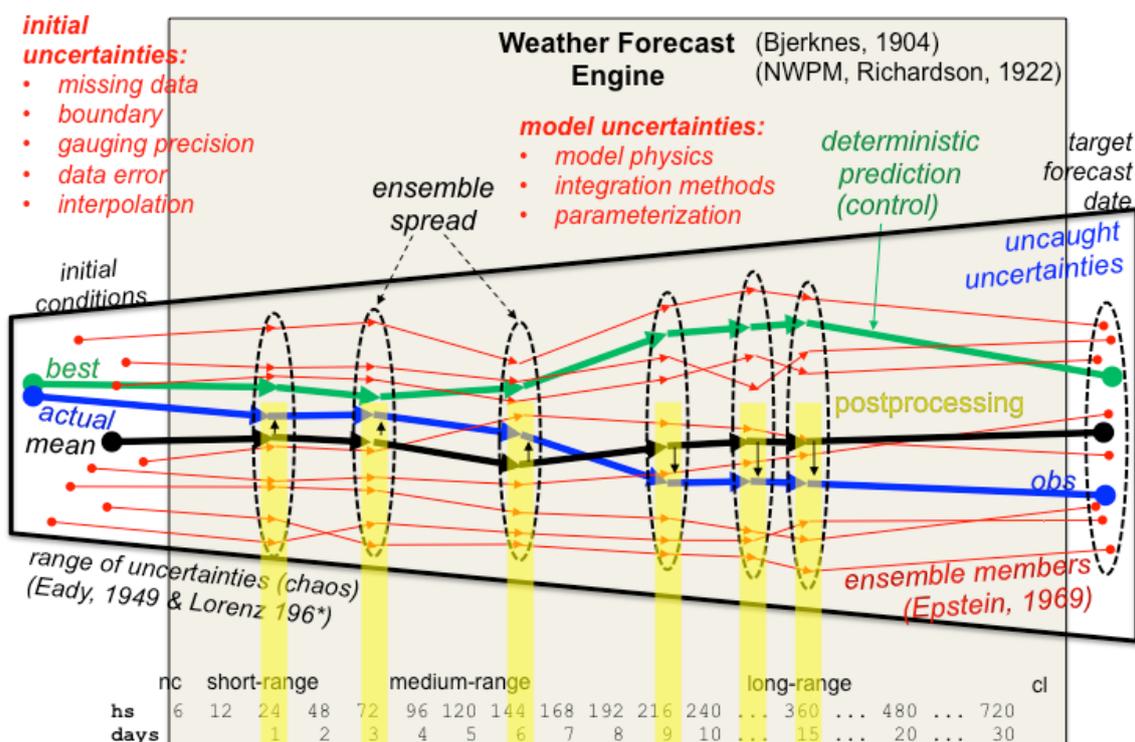
The weather forecasting engine (model), presented by Bjerknes (1904) and improved by Richardson (1922), are represented by the big grey square. It is responsible for the evolution of the "best" state of the atmosphere (filled green circle) by integrating physical equations along time, up to a target forecast date. It turns out that the final values can be too different from actual values especially in longer-range forecasts (blue circles and line), and nothing more than measurements of the magnitude of the error can be computed. Over time, scientists show that it was caused by the uncertainties that are present in the initial state of the atmosphere and even in the way the model computes weather (red texts). This uncertainties represents the difference between what is estimate as states of the atmosphere and its actual values and can lead the forecast to results too far from the truth because of the chaotic behavior of the weather. To alleviate this fact, an ensemble of forecasts (members) was idealized in order to gauge the forecast in a probabilistic way (red circles and lines), since it allows the computation of the event that most likely to occur. An ensemble is composed by multiple evolutions of the weather that are distinguishable either by a different initial state of the atmosphere or by different model operation. In addition to the ensemble, a concept of control member (green line), ensemble mean (black circles and line), and spread (dashed ellipses) arose. The control member is the evolution of the "best" state of the atmosphere. Ensemble mean is de arithmetic mean of all ensemble members and performs better than any individual member. Spread is the measure of how far the lowest member value is from the highest. Besides, being the best, the ensemble mean can, in general, be improved. This work is about post-processing the model output in order to bring the ensemble mean closer to the actual values (lithe black arrows) and it is done for the forecast target dates of 1, 3, 6, 9, 12, and 15 days (yellow transparent rectangles).

Source: author`s production

44

3 DATA AND METHODOLOGY

Given such foundation and motivation, the upcoming sections expose the methodology applied for dataset generation, some exploratory data analysis on the data, and the ANN configuration. The way this work addresses the problem can be classified as an exploratory search of input sets and ANN topologies that, in the end, will be the optimal input-based model that can be applied to calibrate the MEAN forecast of CPTEC-EPS for each lead-time in the forecast horizon.

The programming language used to extract all data from CPTEC-EPS was Fortran (FORmula TRANslator) version 90/95, which is part of the development environment of the CPTEC`s supercomputer, which also stores CPTEC-EPS`s data. The chosen platform used in the development of the procedures for data manipulation, analysis, ANN implementation and assessments was the R Platform (R CORE TEAM, 2015) because it was considered to be a complete platform that could handle all needs.

R is a language and an environment for statistical computing and graphics. It is composed by a core function set, namely base, and optional third-party multi-purpose packages. When applied, R citations are in the form <package>::function().

3.1. Spatial domain

In order to validate the whole process, the La Plata Basin (LPB) region was chosen due to its importance to living society. Coronel (2006) described in some degree of detail the most important aspects. In drainage area terms LPB is world's fifth best ranked and the second with respect to South America, extending over 3.1 million km2, which corresponds to 17% of South America surface (Amazon Basin is the first). As seen in Figure 3.1.a), LPB spreads among five countries: Argentina, Bolivia, Brazil, Paraguay, and Uruguay and includes São Paulo, the largest city in South America, and four of the five

capitals of its countries: Asunción, Brasilia, Buenos Aires, and Montevideo. It is also the most important in population terms accommodating 50% of its five countries population, causing it to be the most populated region within South America.

Moreover, it has important rank in economic activities terms owning 70% of the total Gross Domestic Product (GDP). Its water resources (Figure 3.1.b) sustain the more developed region of South America, where agriculture is one of the most valuable assets. Navigation along the rivers has been increased due to the integration of the regional economies. Also, key ecosystems lay within LPB, such as: The wetland of the Pantanal, Chaco, Pampa's plain, Cerrado, and the Mata Atlantica, revealing abundance and quality of natural resources. In addition, the hydroelectric potential is the main source of the generated power of its countries, which is the core of the economy.

Natural disasters like floods are one of the major concerns in the region due to the damage they cause in almost the whole basin, mainly in the urban areas, as a result of the overflow of the three main rivers of the La Plata basin (Paraná, Uruguay and Paraguay). Improving precipitation forecasts of such area that is highly influenced on water resources is a great step towards mitigating material and economical damages and increasing quality of life for LPB citizens.

Figure 3.1   LPB contour in South America and drainage paths

a)   b)
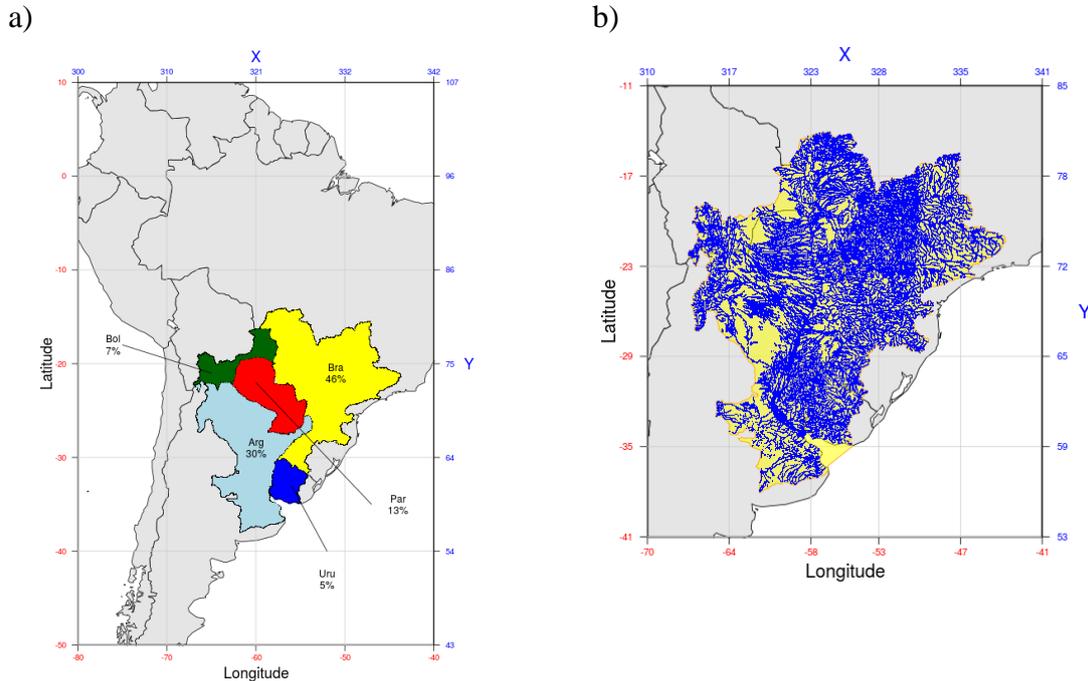
a) LPB contour in South America delimited by colored polygons representing the surface of composing countries and gridded by geographic coordinates (in red) and approximate CPTEC-EPS grid coordinates (in blue). b) Zoomed LPB contour with drainage paths.

Source: author's production based on CIAT (1998) and base::plot().

Further details about LPB can be read in Barros, Clarke, and Dias (2006), which gathers studies aiming to raise awareness in the hydrologic community of the important changes that have occurred in the climate and hydrology of the LPB during recent last decades. Barros et al. (2005) must also be cited because they relate the potential impacts of climate change in the LPB as well as Berbery and Barros (2002) that reports its hydrologic cycle.

Since this work was idealized to operate exactly on each of the CPTEC-EPS grid points, it was necessary to match the LPB geographic coordinates (longitude/latitude pairs) and the CPTEC-EPS grid coordinates (X/Y pairs) in order to extract the data. According to Villela (1984), LPB's extreme geographic coordinates are 14.08 to 37.61 of latitude south and 43.58 to 67.0 of longitude

west, in decimal representation. This domain matches grid coordinates 314 to 339 and 54 to 81, in Y- and X-axis, respectively. However, there must be data from the adjacent grid points for every X-Y pair in LPB domain due to some experiments require it as predictors (details later), therefore, it was necessary to enlarge all four extreme grid coordinates at one grid point. After all, for extraction purposes, the final extreme grid coordinates were set to 313 to 340 (in X-axis) and 53 to 82 (in Y-axis), composing a 28 x 29 grid (812 grid points). Figure 3.2 shows a representation of both working and extraction domains.

Figure 3.2   Extreme extraction and working coordinates of LPB.
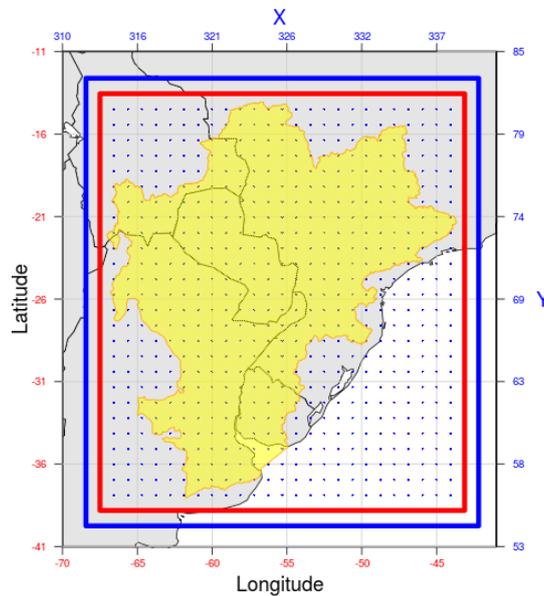
A zoomed view of the LPB extreme grid coordinates and contour (in yellow). Red line bounds the working domain and blue line the extraction domain. Blue dots represent CPTEC-EPS grid points.

Source: author's production based on CIAT (1998) and base::plot().

The term working domain (LPB domain from now on) means the square area limited by the extreme grid coordinates that is effectively used as a source of choice of the grid points for the experiments. This domain is the minimum grid-based square area that encloses the LPB contour and is used to build the final

data sets. The term extraction domain refers to an expansion of the working domain by one grid point on both horizontal and vertical directions. The extra data points of the extraction domain are only used to aggregate information of the adjacent grid points of the working domain grid points.

However, the implementation of all experiments of a pilot project of machine learning system for the whole LPB domain was considered not feasible since it would be difficult to assess the results to each grid point as well as very time-consuming. The initial strategy was set to the following steps: a) pick up some grid point within LPB contour; b) perform all the experiments on it; c) assess the results; d) reapply then on other sparsely located grid points still within LPB contour, and e) compare the results of all chosen grid points.

In order to start the experiments in a location that best represents the rainy season (defined later), the choice of the first grid point was based on the amount of rainy days among all grid points. The MEAN values at every grid point within LPB domain were used to classify the rainner grid points. Figure 3.3 shows a color-ranked map based on the amount of rainy days that was used to choose the first grid point.

Figure 3.3   Color-ranked LPB domain and top 30 best-ranked rainy grid points.
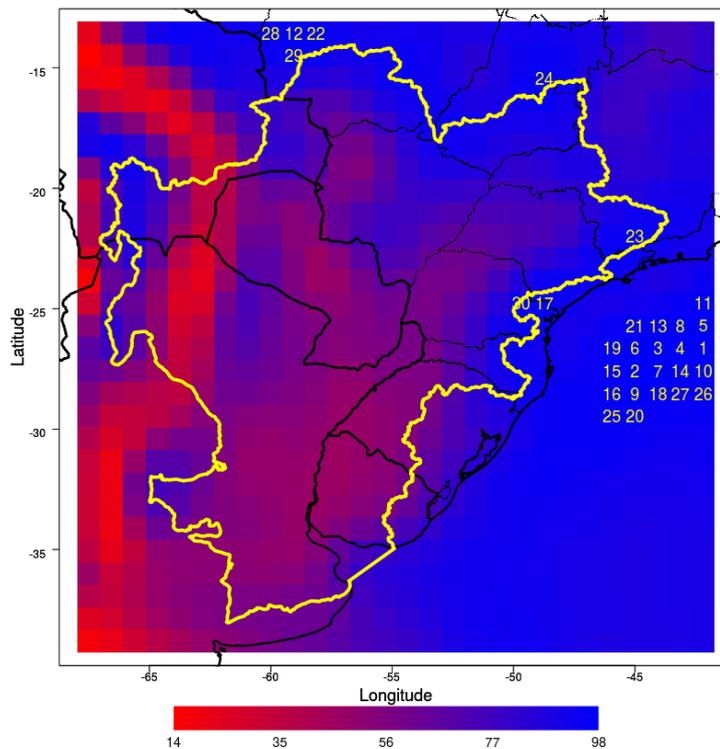
LPB domain filled with a gradient color from red, which represents lower rates of rainy days, to blue, which represents higher rates. Numbers in yellow represent the rank and location of the top 30 best rated grid points, in descending order. The lowest rainy rate within spatial domain was 14% and the highest was 98%. The yellow polygon represents LBP physical contour.

Source: author's production based on CIAT (1998) and base::plot().

As can be seem in Figure 3.3, the first grid point that lies within LPB contour is the 23rd best-ranked grid point. A closer look at its physical map location shows that it lies near to the town of Caxambu, state of Minas Gerais, Brazil. Thereafter, it was chosen as the first grid point for the experiments.

As shows Figure 3.4, the nearest town from the grid point is Caxambu, thus it is be referenced as so. The same naming logic is applied for the other grid points, i.e., the nearest town names it.

Figure 3.4   Terrain view of the 23rd best-ranked rainy grid point and surroundings.

Red filled circle is the location of the 23rd best-ranked grid point within LPB contour in terms of rainy days.

Source: author's production based on ggmap:googlemap().

In addition to Caxambu, the other grid points, which were chosen by spatial distribution and/or importance (such as being the capital of state or country), sum up to twelve and are listed in more details in Table 3.1 and spatially showed on Figure 3.5.

Table 3.1   List of chosen LPB grid points.

X	Y	LON	LAT	SNAME	NAME	PERC %
337	73	-45.0000	-21.9739	CX	Caxambu_MG	96.4
332	69	-49.6875	-25.7142	CT	Curitiba_PR	95.4
335	71	-46.8750	-23.8441	SP	Sao Paulo_SP	94.9
334	80	-47.8125	-15.4285	BR	Brasilia_DF	94.2
315	76	-65.6250	-19.1688	PO	Potosi_BOL	92.8
325	80	-56.2500	-15.4285	CB	Cuiaba_MT	88.5
327	75	-54.3750	-20.1038	CG	CampoGrande_MS	79.4
325	59	-56.2500	-35.0648	MV	Montevideo_URU	74.6
315	70	-65.6250	-24.7791	ST	Salta_ARG	64.6
327	63	-54.3750	-31.3246	BG	Bage_RS	57.8

| 323 | 59 | -58.1250 | -35.0648 | BA | BuenosAires_ARG | 55.1 |
| 324 | 69 | -57.1875 | -25.7142 | AS | Asuncion_PAR | 54.9 |

SNAME stands for short name. Letters after '_' represent a state, while in Brazil, or a country: MG=Minas Gerais, PR=Paraná, SP=São Paulo, DF=Brazilian Federal District, BOL=Bolivia, MT=Mato Grosso, MS=Mato Grosso do Sul, URU=Uruguay, ARG=Argentina, and PAR=Paraguay. In descending order of percentage of rain, indexed by the PERC column.

Source: author's production.

Figure 3.5   Color-ranked rainy LPB domain grid points and the 12 grid points.
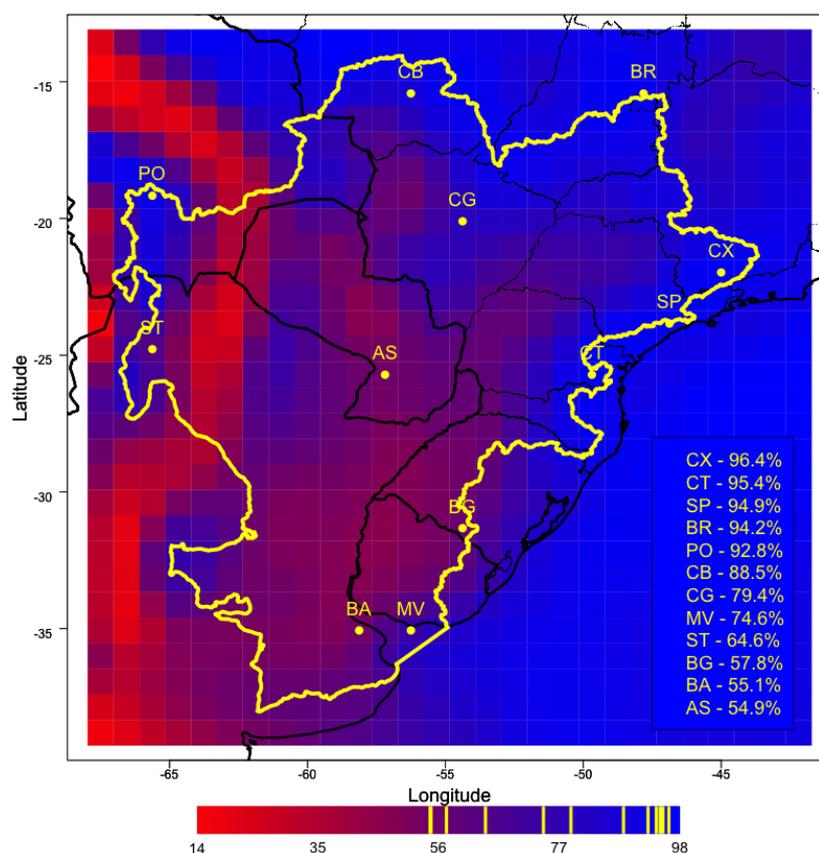
LPB domain filled with a gradient color from red, which represents lower rates of rainy days, to blue, which represents higher rates. Yellow dots are the location of the twelve chosen grid points. The lowest rainy rate within spatial domain was 14% and the highest was 98%. The yellow polygon represents LBP physical contour. The yellow

lines on the bottom legend represent the position of each grid point in the gradient scale.

Source: author's production based on CIAT (1998) and base::plot().

3.2. **Main dataset**

The main dataset is essentially composed by forecast data and associated OBS. It has also been created a few sub-datasets based on combinations or transformations of the raw data.

3.2.1. **Forecasting data**

The forecast data comes from the CPTEC-EPS past forecasts. Datasets like this are usually referred to reforecast (HAMILL et al., 2004, 2006) that, in summary, is a companion dataset generated by retrospective runs of a fixed numerical model in order to be re-analyzed in terms of diagnosing model bias, statistically correcting weather forecasts, studying atmospheric predictability, etc. In general, these sets are produced utilizing other than the model version that is running operationally, and because of that, become computationally expensive as they are great resource consumers. The term fixed here refers to a version of the EPS that is "frozen", i.e., neither programming code nor parameterization values are modified along the runs.

3.2.2. **Data selection**

The process for getting data ready for the ANN algorithm starts with data selection. The following variables listed in Table 3.2 were chosen from Table 2.5 in order to serve as initial predictors for the post-processing training system, expecting that the relationship with OBS can be mapped.

As bottom topography substantially affects the linear baroclinic instability (CHEN; KAMENKOVICH, 2013), the location of data points will be used as a way to filter the data used to build the dataset for the experiments, aiming to

capture how location influences in weather behavior. Thus, in addition to the five meteorological fields (uves, vves, psnm, tsfc, and prec), the data point location (in grid (x/y) and in geographic (lon/lat) coordinates), and the target forecast date, were also extracted.

Table 3.2   List of CPTEC-EPS`s raw variables.

Variable	Description	Unit
x	LOCATION OF THE GRID POINT IN X-AXIS	-
y	LOCATION OF THE GRID POINT IN Y-AXIS	-
fct_date	TARGET FORECAST DATE	DATE
lon	LOCATION OF THE GRID POINT IN LONGITUDE	DEGREES
lat	LOCATION OF THE GRID POINT IN LATITUDE	DEGREES
uves	SURFACE ZONAL WIND (u)	M/S
vves	SURFACE MERIDIONAL WIND (v)	M/S
psnm	SEA LEVEL PRESSURE	HPA
tsfc	SURFACE TEMPERATURE	K
prec	TOTAL PRECIPITATION	KG/M2/DAY

Source: author`s production

3.2.3. Data preprocessing

CPTEC-EPS`S data are stored in gridded data (GRIB format) (WMO, 2016) and its archiving is done by member (15 members) and lead-time (15 days) for each run, which generates 225 forecasting GRIB files. For the purposes of this work the lead-times considered were 1, 3, 6, 9, 12, and 15 days (which corresponds to 24, 72, 144, 216, 288, and 360h, respectively). Moreover, aiming to perform distinct assessments per lead-time, one dataset for each was generated, having the target forecast date as the key that distinguishes the occurrences of the same grid point. After data sampling and formatting, the dataset structure is composed as described in Table 3.3. The final number of columns sums up to 80, as there are 5 key columns and 75 data columns (5 variables x 15 members).

Table 3.3   Members dataset structure.

Variable	Type	Contents
X	int	313 313 313 313 313 313 313 313 313 313 ...
Y	int	55 55 55 55 55 55 55 55 55 55 ...
LON	num	-67.5 -67.5 -67.5 -67.5 -67.5 -67.5 -67.5 -67.5 -67.5 ...
LAT	num	-38.8 -38.8 -38.8 -38.8 -38.8 ...
FCT_DATE	Date	"2009-01-16" "2009-01-17" "2009-01-18" ...
UVES_01N	num	0.81 4.84 2.37 3.65 6.75 2.89 6.76 -1.48 3.26 5.37 ...
VVES_01N	num	1.54 4.94 3.11 5.3 -0.02 4.13 0.2 -3.29 2.09 4.36 ...
PSNM_01N	num	1018 1015 1015 1015 1021 ...
PREC_01N	num	0 0 0 0 0 0 0 0 0 0 ...
TSFC_01N	num	293 296 296 298 293 ...
UVES_01P	num	8.48 -5.45 7.78 -0.47 -4.26 4.24 1.64 -1.36 4.1 4.41 ...
VVES_01P	num	2.23 6.03 0.36 1.44 -0.99 4.54 -2.62 -0.07 5.84 3.86 ...
PSNM_01P	num	1013 1028 1012 1019 1019 ...
PREC_01P	num	0 0 0 0 0 0 0 0 0 0 ...
TSFC_01P	num	293 294 293 294 292 ...
UVES_02N	num	-3.37 -4.15 -0.2 -3.78 0.75 5.73 -0.03 6.24 7.1 2.21 ...
VVES_02N	num	1.95 -2.83 -2.92 -3.63 4.84 2.44 2.16 3.88 0.03 6.45 ...
PSNM_02N	num	1016 1027 1018 1018 1012 ...
PREC_02N	num	0.22 0 0 0 0 0 0 0 0 0 ...
TSFC_02N	num	296 293 295 299 299 ...

.

. continues up to 75 meteorological variables

.

UVES_CTR	num	0.68 1.97 2.66 5.06 5.7 4.32 -0.46 2.03 2.75 -1.73 ...
VVES_CTR	num	-3.32 2.11 -0.69 6.34 5.2 5.85 3.01 -0.91 1.43 6.09 ...
PSNM_CTR	num	1019 1018 1011 1013 1014 ...
PREC_CTR	num	0 0 0 0 0 0 0 0 0 0 ...
TSFC_CTR	num	297 295 295 296 294 ...

In the variable name, "01" and "02" means the first two perturbations (of seven), "N" or "P" means the negative and the positive perturbation, respectively. CTR means the CTRL member. UVES, VVES, PSNM, PREC, and TSFC are the variables extracted from the GRIB datafile and are explained in Table 3.2.

Source: author`s production based on base::str().

3.2.4. Data transformation

3.2.4.1. Averaging by arithmetic mean

Once the raw dataset is ready, some data transformation was considered in order to observe whether and how new features contribute to the whole calibration process. Moreover, as the goal of this work is calibrating MEAN, a natural step would be creating a dataset averaged by the arithmetic mean of the members. Thus, all fifteen members from each of the five meteorological fields were averaged and suffixed with '_MEAN', as shown in Table 3.4.

Table 3.4   Averaged dataset structure.

Variable	Type	Contents
X	int	313 313 313 313 313 313 313 313 313 313 ...
Y	int	55 55 55 55 55 55 55 55 55 55 ...
LON	num	-67.5 -67.5 -67.5 -67.5 -67.5 -67.5 -67.5 -67.5 -67.5 ...
LAT	num	-38.8 -38.8 -38.8 -38.8 -38.8 ...
FCT_DATE	Date	"2009-01-16" "2009-01-17" "2009-01-18" ...
UVES_MEAN	num	0.584 1.601 2.279 4.666 5.346 ...
VVES_MEAN	num	-2.9 2.23 -1.1 6.3 5.32 ...
PSNM_MEAN	num	1019 1018 1011 1012 1014 ...
PREC_MEAN	num	0 0 0 0 0.000667 ...
TSFC_MEAN	num	296 295 296 296 294 ...

Source: author`s production based on base::str().

It is opportune to mention that PREC_MEAN is, actually, the MEAN, which this work aims to calibrate.

3.2.4.2. Neighborhood generation

Aiming to capture the influence that neighbors could have on each data point, it was considered to associate the data of all five variables from the eight adjacent data points to each sample case line (also known as observation) in both members and mean data sets. Figure 3.6 illustrates the idea behind neighborhood's data inclusion.

56

Figure 3.6   Representation of the adjacent neighborhood of a data point.
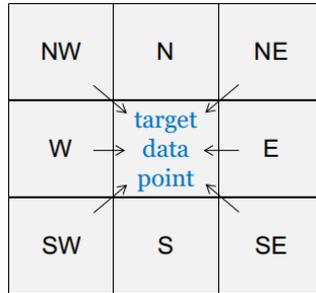
Each target grid point is composed by its own data and the data from the eight adjacent neighbors in cardinal and ordinal directions: northwest (NW), north (N), northeast (NE), west (W), east (E), southwest (SW), south (S), and southeast (SE).

Source: author`s production

This construction makes each sample line in the dataset, and consequently each input set of the ANN, to be composed by 9 values (grid point + 8 neighbors). This combination made the total number of columns of both mean and members data sets to be increased by a factor of eight, for every meteorological feature. As shown in the examples in Table 3.5, the process of naming the new features was done by prefixing each variable with the letters that represents each compass direction.

Table 3.5   Examples of neighbor variables.

Variable	Type	Contents
E_UVES_01N	num	0.67 4.26 0.8 3.44 6.25 2.77 6.17 -2.7 4.76 ...
SW_UVES_01N	num	1.53 4.47 2.51 4.72 9.42 3.73 8.08 0.68 4.84 ...
S_UVES_01N	num	1.51 3.85 1.01 4.19 8.96 3.46 7 -0.46 4.14 ...
NW_VVES_01N	num	-0.4 4.81 3.71 5.27 0.4 3.91 0.19 -3.28 4.55 ...
....		
SW_PSNM_MEAN	num	1018 1019 1018 1015 1014 ...
SE_PSNM_MEAN	num	1018 1019 1017 1015 1013 ...
NW_PREC_MEAN	num	0.367 0.137 1.867 0 0 ...
N_PREC_MEAN	num	0.0333 0 1.366 0 0 ...
NE_PREC_MEAN	num	0.000667 0 0.78 0.001333 0

Source: author`s production, based on base::str().

The number of resulting columns in each dataset is summarized in Table 3.6.

Table 3.6   Number of columns in data sets.

Dataset	Key	Meteorological	Neighbor	Total
Members	5	15 (members) x 5 = 75	75 * 8 = 600	680
Mean	5	5	5 * 8 = 40	50

Both members and mean data sets are composed by five key columns (X, Y, LON, LAT, and FCT_DATE), five meteorological variables (UVES, VVES, PSNM, TSFC, and PREC), and the eight neighbors for each of them.

Source: author`s production.

3.2.5. Sampling and temporal domain

The CPTEC-EPS runs initialized at 12 UTC were was used in order to match the accumulation timetable of the OBS dataset, in which is computed exactly at 12h, which avoids further conversions. Due to computational constraints it was generated data only from January 2009 up to April 2012 and, moreover, to avoid no-rain data, only rainy regime was considered, i.e., data from months October, November, December, January, February, March, and April. Table 3.7 summarizes the amount of the available data per month and year at every grid point of LPB domain.

Table 3.7   Amount of data samples per grid point.

Year\Month	Jan	Feb	Mar	Apr	Oct	Nov	Dec	Total
2009	*30	28	31	30	30	30	31	210
2010	31	28	31	30	30	30	31	211
2011	31	28	31	30	30	30	31	211
2012	**31**	**29**	**30**	**30**	**-**	**-**	**-**	**120**

Numbers are the amount of samples cases for each grid point. Data from 2009 to 2011 was used for training, summing up to 632 samples. Data from 2012 (in bold) was used for testing, summing up to 120 cases. (*) Data generation problem this month.

Source: author`s production, based on base::table().

The ordinary rainy season range includes months October to March but according to Figure 3.7, the precipitation in April is comparable to October, thus, it seems natural to include April in the rainy season, helping to increase sample case number.

Figure 3.7   Average precipitation of the whole La Plata Basin.
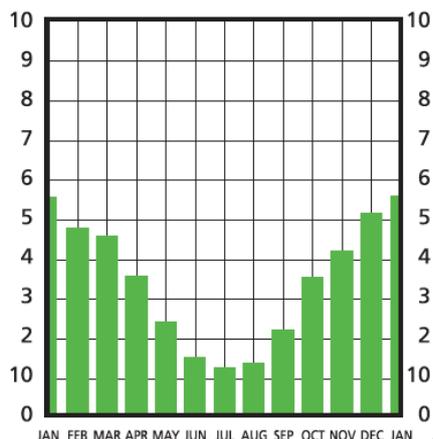
Precipitation in mm/day.

Source: Barros, Clarke, and Dias (2006:p22)

3.2.6. Observed precipitation data

OBS data come from a product that is a combination of satellite precipitation estimates with surface observations over the South American continent – namely MERGE (ROZANTE et al., 2010). The former comes from a joint project – the Tropical Rainfall Measuring Mission (TRMM) - between the National Aeronautics and Space Administration (NASA) and the Japan Aerospace Exploration Agency (JAXA). They provide different products based on different satellite sensors and the 3B42RT product (HUFFMAN et al., 2007) is the one MERGE uses as one of the combining elements. The other component comes from a CPTEC'S database fed up by approximately 1500 surface stations that report data to CPTEC on a daily basis. Its spatial distribution is very irregular, as shown by white dots in Figure 3.8 (a), since there is a major concentration of the surface stations in the eastern portion of South America and becomes

drastically scarce towards both the north and south. The main goal of MERGE is to minimize problems with interpolation in regions of low-density observation networks, as well as the low and high biases of the TRMM product.

To achieve the goal, MERGE technique consists in combining both sources of data according to the weight they have in terms of being the most reliable information, it is assumed that surface observations are considered "the truth", thus, have more weight than satellite data. Once verified which data is the most valuable, it is used to establish the value of the grid box itself and of the two rows of adjacent grid boxes (resulting in 24 grid points). The establishment of values is done by an interpolation method idealized by Barnes (1973). Figure 3.8 (b) shows that the white dots in (a) have its dimension increased due to this technique. And, for the remaining grid boxes in which there are no surface observations, the TRMM data are considered.

Results show the performance is superior over areas with sparse observations and equivalent in terms of simply averaging the stations over areas with a high density of observations. Using MERGE in this work is a way to guarantee that the observational data will be at least equivalent in performance compared to data from surface stations since various grid points in LPB domain lay within areas with scarce data.

Figure 3.8   The TRMM gauging operation.
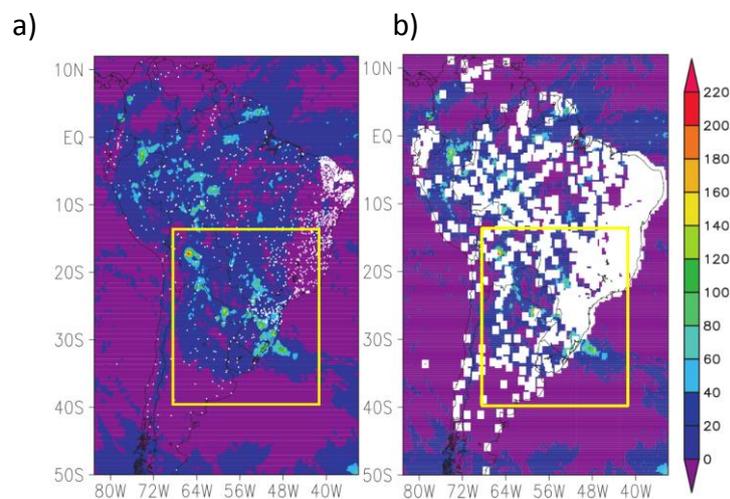
The TRMM 24-h accumulated precipitation estimate (shaded) for 8 Mar 2003. (a) White dots represent the reporting rain gauges on that particular day. (b) White squares represent the 0.258 TRMM grid boxes near the rain gauges shown in (a). The box in yellow represents the LPB geographic boundaries.

Source: Adapted from Rozante et al. (2010).

As shown in Figure 3.9, the final merged product has a resolution of 0.25 degrees and CPTEC-EPS has a resolution of approximately 0.94 degrees, which requires a *regrid* (conversion from one grid resolution to another). To fix it, the simple arithmetic mean of the overlaid MERGE grid boxes was calculated for each grid box in CPTEC-EPS resolution.

Figure 3.9   TRMM to CPTEC-EPS grid resolution conversion.
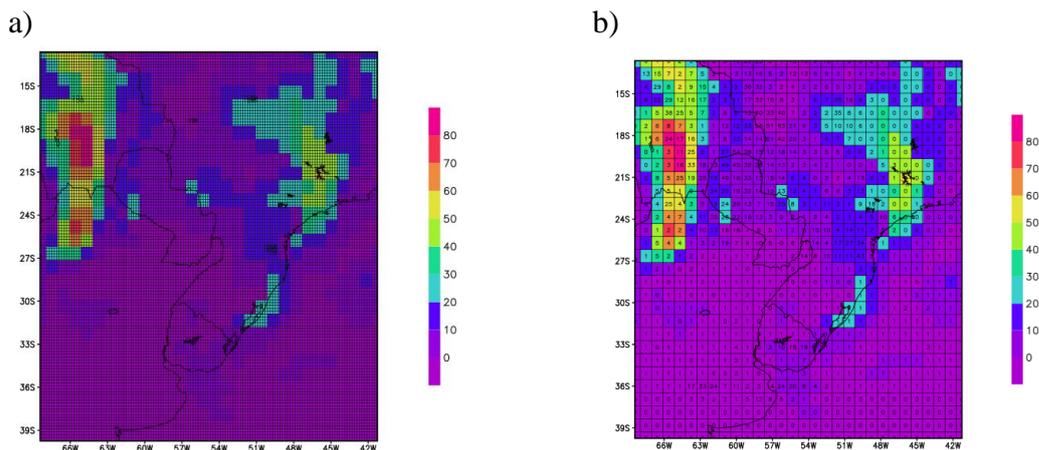
TRMM`s original spatial grid resolution in (a) and *regrided* in (b), matching CPTEC-EPS spatial grid resolution. Colors mean the amount of rainfall for each grid point in some date (irrelevant in this case).

Source: Dr. Christopher Castro's production.

Data from MERGE were also extracted to another particular dataset in order to serve as the source of the target variable for the training process as well as assessments in the testing phase. Table 3.8 shows its structure.

Table 3.8　　　OBS dataset structure.

Variable	Type	Contents
X	int	313 313 313 313 313 313 313 313 313 313 ...
Y	int	55 55 55 55 55 55 55 55 55 55 ...
LON	num	-67.5 -67.5 -67.5 -67.5 -67.5 -67.5 -67.5 -67.5 -67.5 ...
LAT	num	-38.8 -38.8 -38.8 -38.8 -38.8 ...
FCT_DATE	Date	"2009-01-16" "2009-01-17" "2009-01-18" ...
PREC_OBS	num	0.466343 0 0 0 0.000000323

Source: author`s production based on base::str().

3.2.7. Normalization

Another data pre-processing technique is the normalization, in which, here, refers to the process of adjusting data values measured in different scales to some normal form, i.e., to a common scale. It must be applied to all input and output features within the system. As shown in Table 3.9, there are mean values among the five meteorological variables that differ from other in some order of magnitude. The difference of the scaling factor between input or target variables tends to make back-propagation type algorithms to be slow and have inaccurate results (SARLE, 2002). It's also opportune to mention that it is not about modeling because ANNs are flexible enough to bypass any problem caused by data distribution of the input, but it is only about numerical issues.

Table 3.9　　　Descriptive statistics of meteorological variables.

Statistics	PREC_01N	TSFC_01N	PSNM_01N	UVES_01N	VVES_01N
#Samples	589568	589568	589568	589568	589568
Mean	2.43	296.26	1013.67	-0.45	-0.15
Std. Dev.	7.96	5.78	4.44	3.17	3.40
Median	0.00	297.35	1013.18	-0.39	-0.32
Min	0.00	259.53	986.62	-19.49	-17.13
Max	250.91	311.13	1034.94	18.32	19.48
Range	250.91	51.60	48.32	37.81	36.61
Skew	7.42	-1.36	0.41	0.21	0.17

Descriptive statistical measurements on the 24h member dataset. #Samples is 589,568 because there are 784 grid points and 752 days.

Source: author`s production, based on psych::describe()

Normalizing either input or target variables ensure that values are equally important to the training process (DA SILVA, 2001). As the target variable of this work is OBS, whose lower limit is zero, it seems natural to normalize the target within the range [0, 1], and consequently, apply the same normalization process for all input features. There are many methods for data normalization in literature. Some commonly found are min-max, Z score, and decimal scaling normalization.

Based on Jiawei et al. (2012), let A be a numeric attribute with n sample values, $v_1, v_2, \dots, v_n$. Min-max normalization performs a linear transformation on the original data ranging them to the interval within [0,1]. The drawback is that an "out-of-bounds" error is raised if a future input case for normalization falls outside of the original data range. Min-max normalization maps a value, v_j, of A to v_j' in the range [0, 1] by computing

$$v_j' = \frac{v_i - min_A}{max_A - min_A}$$
(2.1)

where min_A and max_A are the minimum and maximum values of an attribute, A.

Z-score normalization (or zero-mean normalization) normalizes the values based on the arithmetic mean and standard deviation. It is a zero-centered-based algorithm in which lower and upper values are not necessarily symmetric. It is useful when the actual minimum and maximum of attribute are unknown, or when there are outliers that dominate the min-max normalization. Z-score normalization maps a value, v_j, of A to v_j' in the range [0, 1] by computing

$$v_j' = \frac{v_i - \bar{A}}{\sigma_A}$$
(2.2)

where $\bar{A}$ and σ are the mean and standard deviation, respectively, of attribute A.

Decimal scaling normalization acts by moving the decimal point of values of the attribute according to the number of decimal points moved on the maximum absolute value of the attribute. It is a quasi-zero-centered-based algorithm in which lower and upper values are not necessarily symmetric. Decimal scaling normalization maps a value, v_j, of A to v_j' in the range [0, 1] by computing

$$v_j' = \frac{v_i}{10^j} \qquad (2.3)$$

Where j is the smallest integer such that $\max(|v_j'|) < 1$.

Table 3.10 shows the application of these normalization methods on the meteorological variables. As can be noticed, the skew factor remains the same as original values, which is shown in Table 3.6, for every normalization method, indicating that the relationships that exist in the original data were preserved.

Table 3.10   Descriptive statistics of meteorological variables, normalized by three different methods.

Statistics	PREC			VVES		
	Min-max	Z-Score	Decimal Scaling	Min-max	Z-Score	Decimal Scaling
Mean	0.01	0.00	0.00	0.46	0.00	0.00
Std. Dev.	0.03	1.00	0.01	0.09	1.00	0.03
Median	0.00	-0.30	0.00	0.46	-0.05	0.00
Min	0.00	-0.30	0.00	0.00	-5.00	-0.17
Max	1.00	31.22	0.25	1.00	5.77	0.19
Range	1.00	31.52	0.25	1.00	10.77	0.37
Skew	7.42	7.42	7.42	0.17	0.17	0.17
Statistics	PSNM			TSFC		
	Min-max	Z-Score	Decimal Scaling	Min-max	Z-Score	Decimal Scaling
Mean	0.56	0.00	0.10	0.71	0.00	0.30
Std. Dev.	0.09	1.00	0.00	0.11	1.00	0.01
Median	0.55	-0.11	0.10	0.73	0.19	0.30
Min	0.00	-6.09	0.10	0.00	-6.36	0.26
Max	1.00	4.79	0.10	1.00	2.57	0.31
Range	1.00	10.88	0.00	1.00	8.93	0.05
Skew	0.41	0.41	0.41	-1.36	-1.36	-1.36

Data extracted from 24h-members dataset. All variables refer to member 01N. UVES is omitted once that VVES has similar behavior.

Source: author`s production, based on Jiawei et al. (2012) and psych::describe().

However, preserving the original relationship between the attributes is an important component for succeeding in machine learning but it is not the only one. Concerning ANNs, it is highly recommended that the output of the activation function match the distribution of the targets, avoiding forcing the data to conform to the activation function output. In other words, if the activation function outputs values within the range of [0,1], then the target values should lie within that range.

As exposed in the section 2.3, the activation function chosen was the sigmoid function which outputs continuous values within range [0,1]. The natural choice should be choosing the Min-max normalization but the shortcoming that raises an exception for new data that lies outside the original range could be a problem if the validation dataset encloses that case. The remaining Z-score and Decimal Scaling are zero-centered and have free lower and upper limits but it does not become a problem in this work because the original range of the target variable is always zero or higher, which makes the normalization process to output values that are also zero or higher, but in a small scale.

This fact should make both methods suitable for this work, but a closer look at Table 3.10 shows that the minimum and maximum values for PREC ranges from -0.30 to -31.22 in Z-score normalization, while in Decimal Scaling it ranges from 0 to 0.25. Since a normalization process aims to convert numbers to a common scale, the range of the Z-score were considered very high, leading us to chose Decimal scaling as the normalization method.

It should also be mentioned that it is necessary to save the normalization parameters (j, for the Decimal Scaling) in order to denormalize the data back.

3.3. Artificial Neural Network (ANN)

The supervised learning approach is the natural choice of this work since approximating to OBS is the main goal. The foundation exposed in Chapter 1 has revealed that the MLP model architecture is suitable to address the task of handling complex dynamic non-linear relationships regardless data distribution and prior knowledge of the input/output relationship, thus it was chosen.

Two R packages that implement ANNs are 'neuralnet' (FRITSCH; GUENTHER, 2012) and 'nnet' (RIPLEY; VENABLES, 2016). The drawback of nnet package is that it supports only one hidden layer, but was considered too limited for the purposes of this work, which include evaluating few numbers of layers. Thus, thus, 'neuralnet' package was chosen.

MPL networks are applied in this work, however, the use of back-propagation algorithm as the learning algorithm has been avoided due to the fact that it has some problems associated with its convergence (HAYKIN, 1999), such as slow training, getting stuck in the local minimum, or even requiring many processes in order to obtain a response within an acceptable error interval (CORTIVO; CHALHOUB; VELHO, 2012). The chosen learning algorithm was RPROP.

'neuralnet' supports sigmoid and hyperbolic tangent (tanh) as activation functions. Since the target value assumes values greater or equal zero, the sigmoid function was chosen, with a fixed default slope parameter set to 1. The value that specifies the threshold for the partial derivatives of the error function as stopping criteria was set to 10^{-3}. The epoch (maximum steps for the training of the neural network) was set to 1×10^{6}. The weights were set to be randomly initialized and no constraints were added to limit them. In order to become reproducible the "seed" 1968 was set before the each unique 'neuralnet' invocation.

Concerning topology, the suitable model for this work is neither proved nor has a consensus in the literature. Statements similar to "In most situations, there is

no way to determine the best number of hidden units without training several networks and estimating the generalization error of each" (SARLE, 2002), encouraged to set the trial and error strategy as the approach to determine the optimal ANN topology.

All training process was run in a HP Z820, a 32-core and 62GB-RAM workstation. Taking advantage of the multi-process capabilities provided by the R packages 'foreach' (CALAWAY; WESTON, 2015), and 'doParallel' (CALAWAY et al., 2015), it was idealized one hundred ANN topology variations. The idea was to loop the amount of hidden units from 1 to 10 within 1 to 10 hidden layers, which generates 100 topologies to be evaluated for each lead-time of the experiment. This approach generates only 'regular' topologies of hidden layers obeying the terminology *UxL*, where *U* is the number of hidden units in each of the *L* layers. The term 'regular' means that there is no variation on the number of hidden units within the hidden layers.

Figure 3.10 show some examples of the topology and the common terminology that describes the whole topology (1st line) as well as the *UxL* terminology (2nd line) obtained with this approach, which is used to describe the optimal ANN topology of the experiments.

Figure 3.10   Examples of ANN topologies applied to the experiments.

Figure 3.10   Examples of ANN topologies applied to the experiments (cont.).

d) 8-6-6-6-6-6-6-6-1 UxL: 7x6	e) 15-10-1 UxL: 10x1	f) 45-10-10-10-10-10-10- 10-10-10-10-1 UxL: 10x10
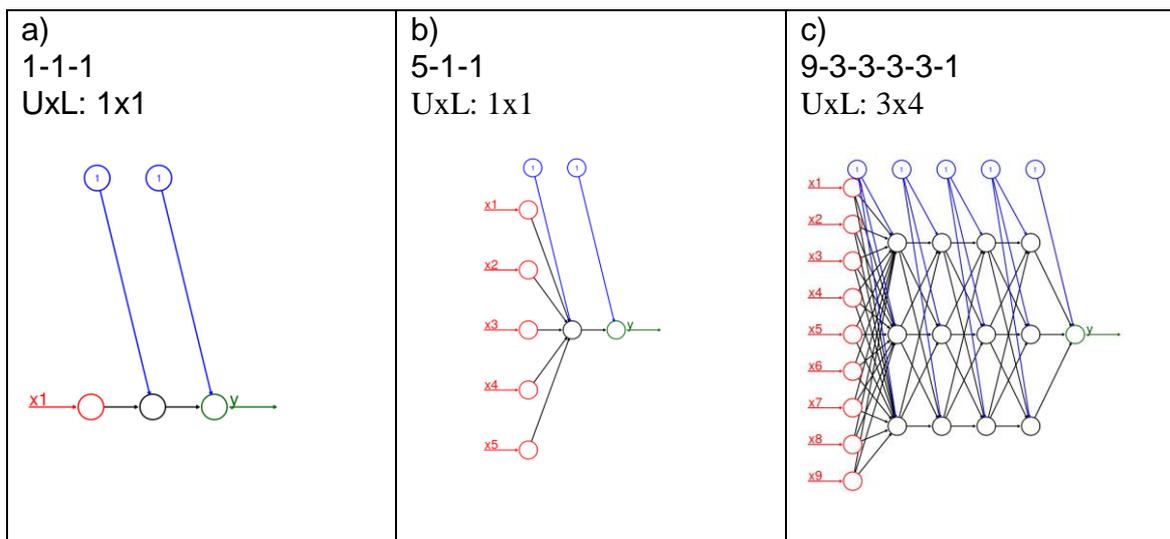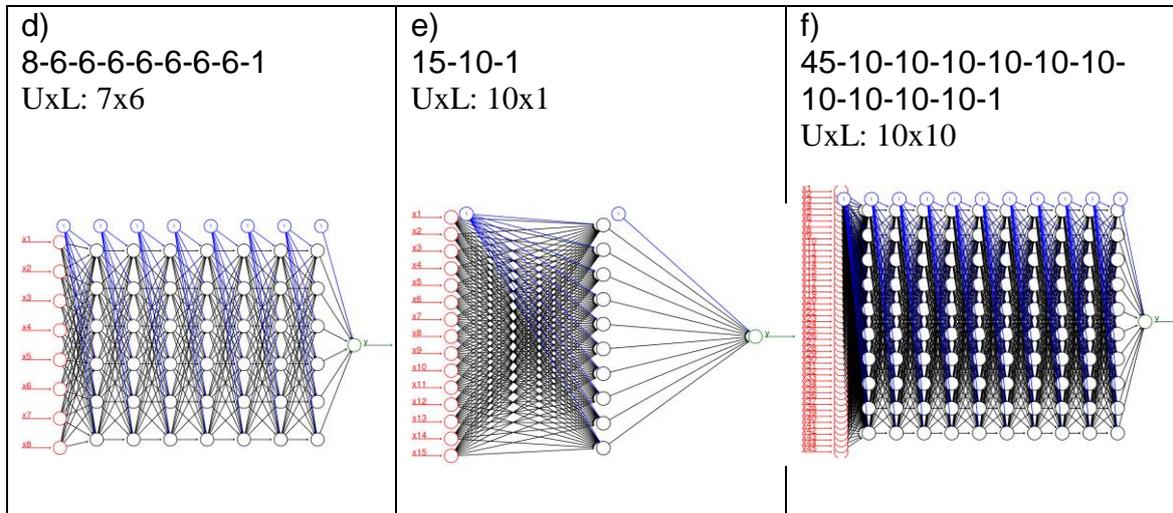
In red, the input sets. In blue, the bias and corresponding synapses linked to every neuron in the next layer. In black, hidden neurons and corresponding synapses. In green, the output set, which has one element in this case. a) Minimum topology used in the experiments and f) the maximum. 1st header line is the terminology commonly found in literature that describes the topology, and the 2nd line is the *UxL* terminology.

Source: author`s production, based on neuralnet::plot.nn().

There are a multitude of possibilities that might be feasible to consider regarding the combination of architecture, topology, and configuration of ANN. Even from few examples in literature there is no consensus about which ANN model is better. As have been highlighted in related work sections, lots of work have been done by training several combinations of architecture, topology and configuration with a subsequent comparison of the generalization error of each one in order to determine the optimal ANN model, just as was done here.

4 EXPERIMENTS AND RESULTS

4.1. **Performance of MEAN calibration**

The inter-comparisons between OBS versus MEAN and ANN output were done through Root Mean Squared Error (RMSE), which is a highly recommended verification score of forecasting precipitation amount (WMO, 2008) and measures the magnitude of the mean error giving greater weight to larger errors. Like every error measurement, values closer to zero is better.

In addition, RMSE skill score (SS) was also applied in order to evidence percentage improvements in ANN models with respect to the MEAN, which is the reference score. They are expressed by Formulas 3.1 and 3.2, in sequence.

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - \hat{y}_j)^2} \tag{3.1}$$

where n = number of samples, y = target output and $\hat{y}$ = ANN estimate.

$$SS_{ANN} = \left(\frac{RMSE_{MEAN} - RMSE_{ANN}}{RMSE_{MEAN}}\right)100\% \tag{3.2}$$

where $RMSE_{ANN}$ is the RMSE of the ANN estimate and $RMSE_{MEAN}$ is the RMSE of the MEAN. The percentage improvement is relative to its magnitude. A positive value indicates a better performance of the ANN compared to the MEAN and a negative value, the opposite.

4.2. **Outlining the experiments**

This study is an attempt to seek an optimal input-driven model for calibrating MEAN given both location and forecast horizon. The term input-driven means that each experiment for approximating MEAN towards OBS is designed based on specific set of predictors set, whose selection were done aiming exploratory purposes. The whole process to build an experiment can be summarized in the

following steps: i) predictors selection; ii) grid point selection: in order to filter data by location; iii) data set creation: creates six distinct data sets, one for each lead-time (1, 3, 6, 9, 12, and 15 days); iv) ANN processing: train and test each data set in each of the one hundred ANN topologies; and v) analysis: gathers and compares the results at each lead-time.

The general goal of each experiment is to measure the ability of the chosen predictors to make MEAN approximate OBS and to check which ANN topology performs better at each lead-time.

In order to explain the predictors used on each experiment in a compact way the following terms, listed in Table 4.1, are used to compose the experiment identifier, which informs the dataset and the features used in the experiments. The experiment identifier is written in upper or lower case interchangeably.

Table 4.1   Meaning of terms used to define the experiments.

TERM	PREDICTOR MEANING
P	Mean precipitation data were used from the mean dataset.
V	Mean data from the variables PREC, TSFC, UVES, VVES, and PSNM were used from the mean dataset.
M\<x\>	Data from member \<x\> were used from members dataset, such that x $\subseteq$ {C, N, P, A}, where C is CTRL, N is member 01N, P is member 01P, and ALL means all 15 members.
GP	Data from target grid point was used.
N	Data from the eight adjacent grid points of the target location were used.
+	Prior set is repeated for N (neighbors).

Source: author`s production.

In the next section, the experiments at Caxambu are described and the results are graphically shown in two charts. In general, more than one experiment is gathered together for clarity purposes but do not necessarily express the order that they were done. Description includes: the experiment identifier, the number

of predictors, the correspondent specific goal, and a discussion about the results based on the charts and numbers.

In line chart (a), the evolution of each experiment within the group is show along the forecast horizon. Moreover, three dark-blue lines, which are located immediately below their associated lead-time, express the RMSE, the ANN topology, and the identifier of the best experiment (EXP). In addition, some legends show the overall RMSE of the experiments and of the MEAN. In bar chart (b), multiple bars illustrate the percentage improvement of each experiment also according to the associated lead-time. The colors of both charts match.

All figures and tables in this section were produced by the author based mainly on base: data.table(), table(), reshape(), dcast(), and ggplot().

4.3. **Experiments at Caxambu**

4.3.1. **Experiment set #1**

Pgp; (1); Check whether ANN is able to model the relationship between MEAN and OBS using only MEAN as predictor.

Pn; (8); Check whether MEAN from neighbors improve ANN forecast at the target location.

Pgp+n; (9); Check whether submitting both MEAN from the target grid point and from its neighborhood improves the calibration.
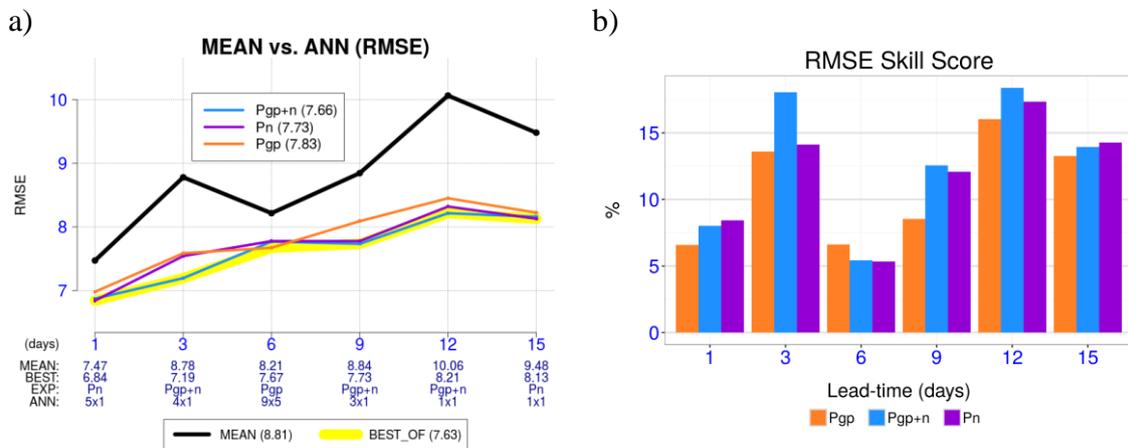
Discussion: ANN calibration from all these experiments surpasses MEAN at every lead-time and, consequently, in the overall performance. The two peaks registered at 3rd and 12th days for MEAN imply that there are larger errors in these target days when compared to ANN. Moreover, a better overall performance was achieved when using both precipitation data from the target location and from the neighbors as predictors (thin blue line) when compared to their individual submission. The "best of" measure (in thick yellow line) indicates the best RMSE at each lead-time, helping to perform calibration using the most suitable ANN model for each of them, in order to boost up the performance of the calibration process.

4.3.2. Experiment set #2

Vgp; (5); Check whether submitting MEAN and other mean variables from the target grid point improves the calibration process.

Vn; (40); Check whether submitting MEAN and other mean variables from the neighbors can improve the calibration.
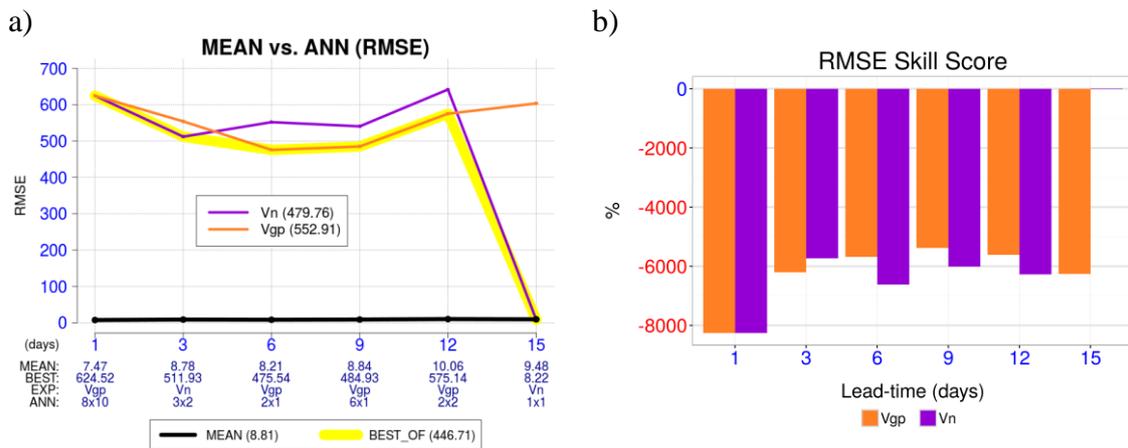
Discussion: The performance of the neural network was worse than MEAN. Large RMSE values possibly indicate that the maximum number of iterations was reached and the ANN did not converge to the solution, most likely because either got stuck in a local minimum or the weights became too high or too low, causing the denominator of the sigmoid function to vary high (or low), characterizing an overflow (namely saturation). Vn presented a small RMSE at the 15th day but it was still worst than the experiments in set #1. The use of the five mean variables together brought none improvements to MEAN calibration, thus, no further experiments were done with them.

4.3.3. Experiment set #3

McPgp; (1); Check how an input built from "the best snapshot of the atmosphere" (the control member) performs along the forecast horizon.

MnPgp; (1); Check how an input built from physical laws (weather model) performs, in case, it is a negative leg of the perturbation.

a)

b)
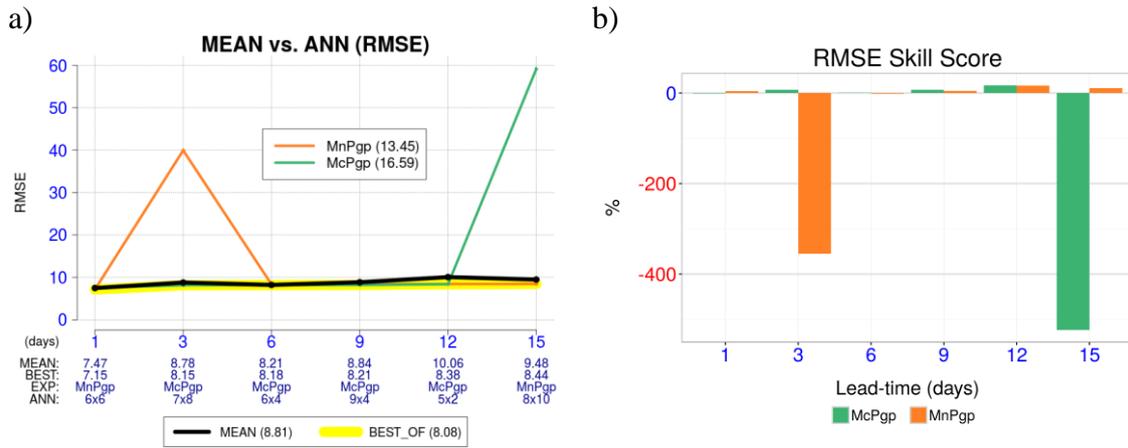
Discussion: Both experiments showed some instability in some forecast day and high overall RMSE. One of possible reasons is that, both CTRL (green) and perturbed members (orange) are deterministic predictions and, because of that, its variability can be too different from OBS, making the relationship between them harder to comprehend, as exposed in Chapter 1.

4.3.4. Experiment set #4

MpPgp; (1); Similar to MnPgp, but using the positive perturbation member.

MaPgp; (15); Measure how the ANN performs using all precipitation data of members from the target location as input.

MaPgp+n; (75); Check whether adding precipitation data of members from neighbors to the precipitation data of members from the target location improves calibration.

c)

d)
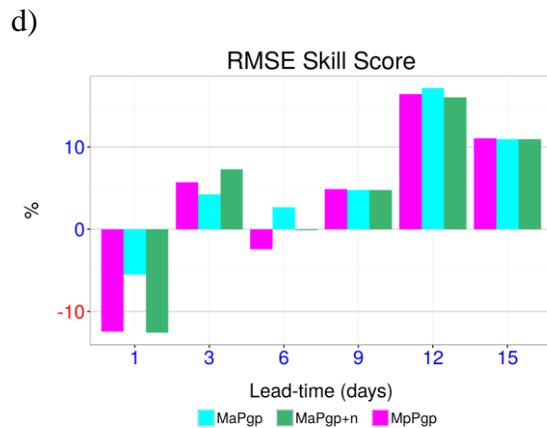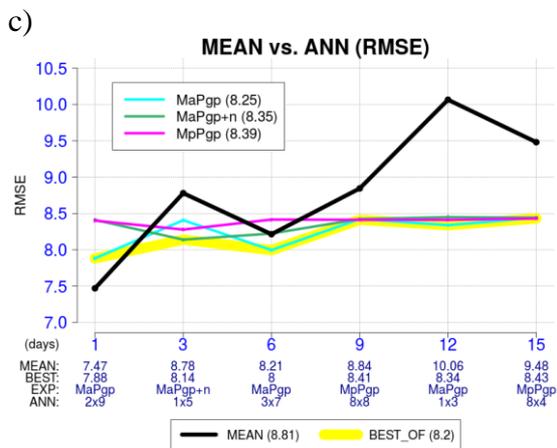
Discussion:

As an exception of the behavior of individual members among the experiments, MpPgp remained stable at certain RMSE level for all lead-times, however, its overall performance was worst that the experiments in set #1. Besides members are, conceptually, individual deterministic predictions, which causes some difficulties on approximating the solution, CPTEC-EPS's members are built from a random mechanism and, therefore, the relationship between a set of values and OBS is not kept among the samples cases. Both experiments concerning all fifteen members presented better RMSE compared to MEAN, but none could surpass the experiments in set #1.

4.3.5. Optimum predictors selection and ANN topology

The optimum set of predictors for each lead-time and concerning Caxambu can be inferred from on Table 4.2, which summarizes the results of the experiment sets.

Table 4.2   RMSE summarization by experiment and lead-time

EXP.ID	OV.ALL	1	3	6	9	12	15
Pgp+n	7.66	6.87	7.19	7.77	7.73	8.21	8.16
Pn	7.73	6.84	7.54	7.78	7.78	8.32	8.13
Pgp	7.83	6.98	7.58	7.67	8.09	8.45	8.22
MaPgp	8.25	7.88	8.41	8.00	8.42	8.34	8.44
MaPgp+n	8.35	8.41	8.14	8.22	8.42	8.45	8.44
MpPgp	8.39	8.40	8.28	8.42	8.41	8.41	8.43
MEAN	8.81	7.47	8.78	8.21	8.84	10.06	9.48
MnPgp	**13.45**	**7.15**	**39.94**	**8.35**	**8.41**	**8.43**	**8.44**
McPgp	**16.59**	**7.52**	**8.15**	**8.18**	**8.21**	**8.38**	**59.11**
Vn	**479.76**	**624.54**	**511.93**	**551.85**	**540.33**	**641.66**	**8.22**
Vgp	**552.91**	**624.52**	**553.97**	**475.54**	**484.93**	**575.14**	**603.37**

Sorted by overall performance (OV.ALL). Shows RMSE at every lead-time (1-15). Bold lines in back have worst overall performance compared to MEAN (RMSE=8.81, in red). Delimited values refer to the best RMSE at each lead-time.

Source: author`s production.

Based on the results it can be seen that only the first three experiments had at least one of its values as the best of some lead-time, which are delimited by black rectangles in Table 4.2. The experiments that used only the MEAN from the target grid point, the one that used only the MEAN from neighbors, and the one that used both data together are comprised by experiment set #1. A detailed view of the numeric results is shown in Table 4.3. Bold lines refer to the optimum experiment (best predictors set) and ANN topology at each lead-time that was computed for Caxambu.

Table 4.3　　　RMSE results of experiment set #1 by lead-time.

EXP.ID	1	3	6	9	12	15
MEAN	7.47	8.78	8.21	8.84	10.06	9.48
BEST	6.84	7.19	7.67	7.73	8.21	8.13
EXP	Pn	Pgp+n	Pgp	Pgp+n	Pgp+n	Pn
ANN	5x1	4x1	9x5	3x1	1x1	1x1

Source: author`s production.

And the final optimum predictor set and ANN topology for Caxambu is summarized in Table 4.4.

Table 4.4　　　Optimum predictor set and ANN topology for Caxambu.

Day: 1 Pn (8) 5x1 or 8-5-1	Day: 3 Pgp+n (9) 4x1 or 9-4-1	Day: 6 Pgp (1) 9x5 or 1-9-9-9-9-9-1
Day: 9 Pgp+n (9) 3x1 or 9-3-1	Day: 12 Pgp+n (9) 1x1 or 9-1-1	Day: 15 Pn (8) 1x1 or 8-1-1
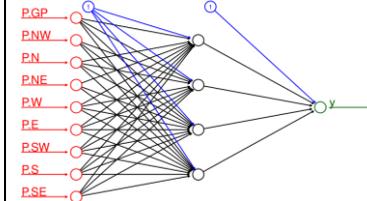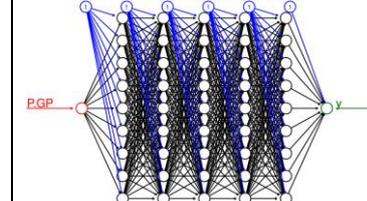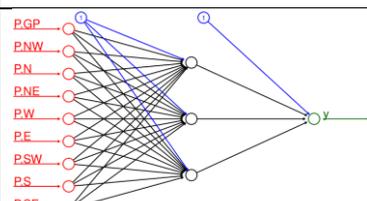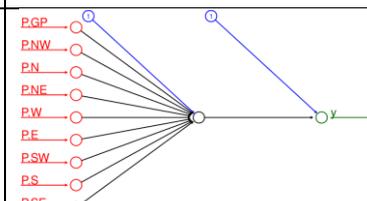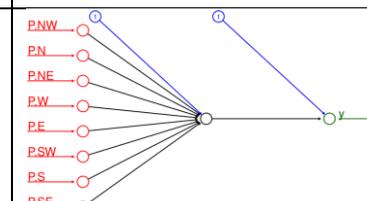
First line: forecast days 1, 3, 6, 9, 12, and 15. Second line: Experiment ID plus number of predictors. Third line: compact ANN topology notation plus common notation. Chart: representation of the resulting ANN topology. Red=inputs, blue=bias and its synapses, black= hidden neurons and its synapses, green=output.

Source: author`s production.

In order to prevent from applying underperforming experiments to other grid points, a cutting criterion was established aiming to ignore every experiment whose overall performance was below MEAN overall performance (RMSE=8.81). Table 4.2 demonstrates the ignored experiments in bold and the remaining experiments are: Pgp+n, Pn, Pgp, MaPgp, MaPgp+n, and MpPgp.

4.4. **Experiments at remaining locations**

4.4.1. **Assuncion (PAR) - AS**

a)

b)

c)

d)
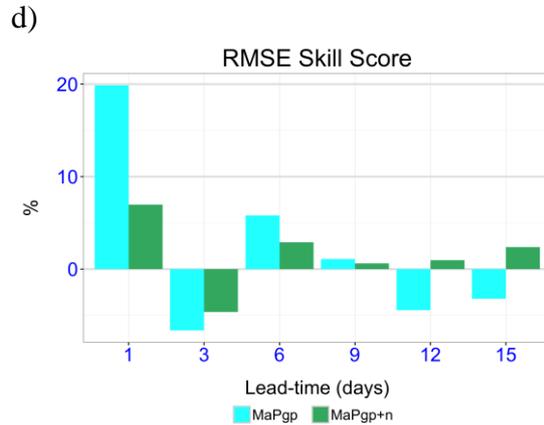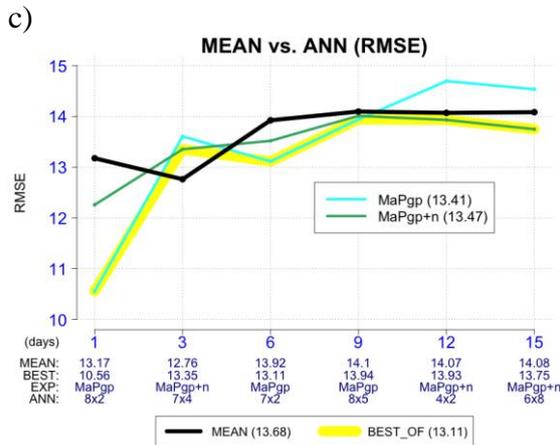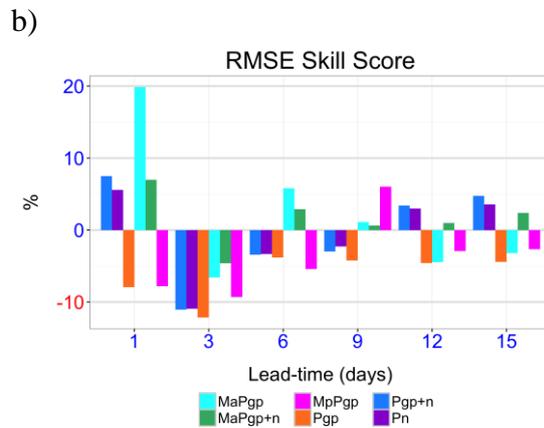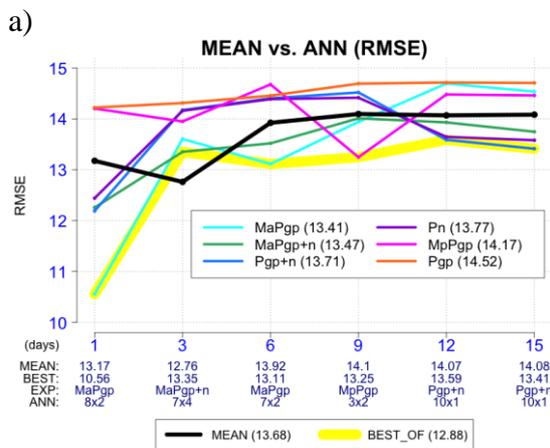
Four of six experiments that performed worst than MEAN were eliminated. The 1st target lead-time achieved the maximum improvement, reaching 20%, but in overall ANN could help only a little since it was around 13%.

4.4.2. Buenos Aires (ARG) - BA

a)

b)

c)

d)
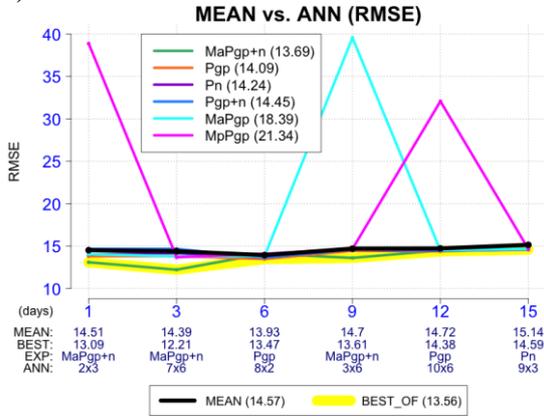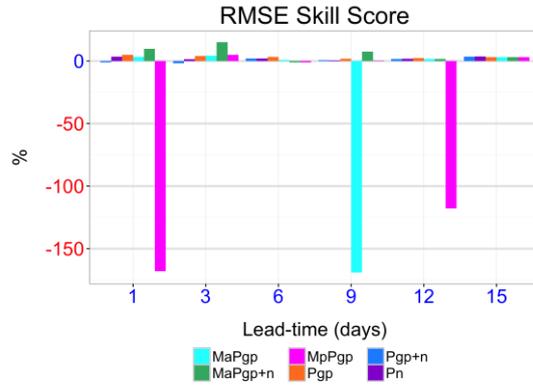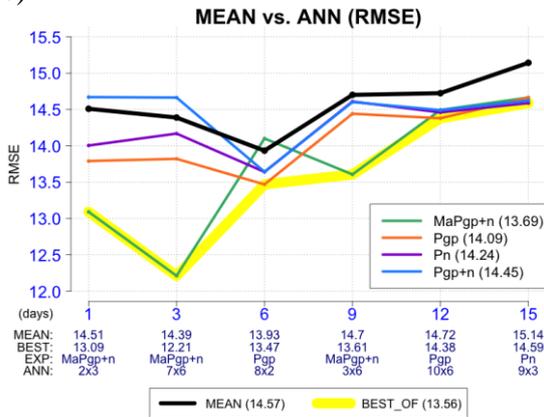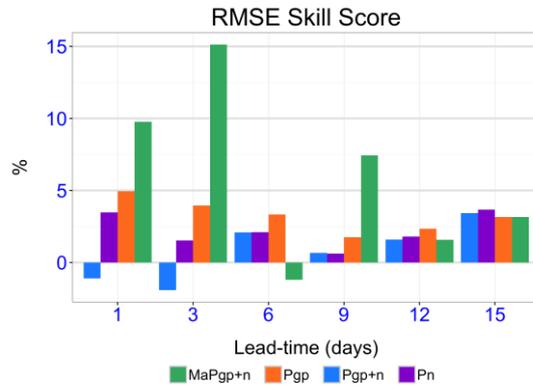
Four of six experiments that performed worst than MEAN were eliminated. Better improvements were verified at the 1st and 3rd days, reaching 15%. Overall performance was around 14%.

4.4.3. Bagé (RS) - BG

a)

MEAN vs. ANN (RMSE)

(days)	1	3	6	9	12	15
MEAN:	10.82	11.08	12.41	12.41	12.36	12.65
BEST:	10.09	10.49	11.65	11.52	12.06	12.13
EXP:	MpPgp	Pgp	Pn	MaPgp+n	MaPgp	MpPgp
ANN:	5x3	7x1	6x10	4x3	2x2	10x6

b)

RMSE Skill Score
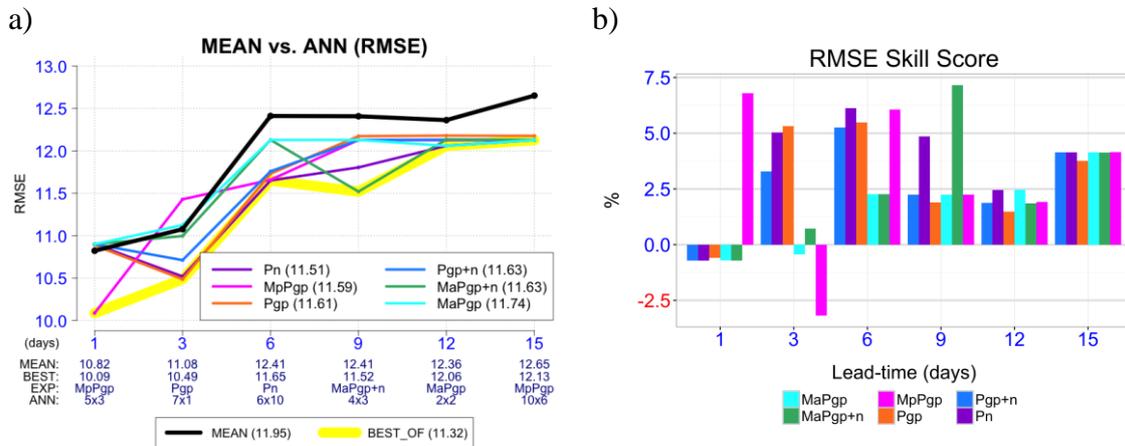
There was no experiment that performed worst than MEAN, so none was eliminated. In general, ANN performance followed the MEAN and improved around 9%.

4.4.4. Campo Grande (MS) - CG
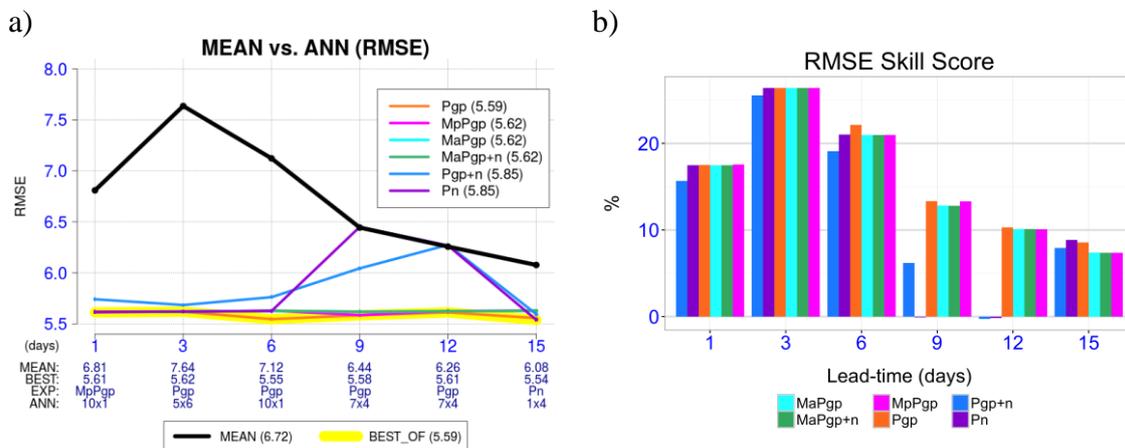
a)

MEAN vs. ANN (RMSE)

(days)	1	3	6	9	12	15
MEAN:	6.81	7.64	7.12	6.44	6.26	6.08
BEST:	5.61	5.62	5.55	5.58	5.61	5.54
EXP:	MpPgp	Pgp	Pgp	Pgp	Pgp	Pn
ANN:	10x1	5x6	10x1	7x4	7x4	1x4

b)

RMSE Skill Score

There was no experiment that performed worst than MEAN, so none was eliminated. A better performance is noticed in the 3rd day achieving 26%. In general, ANN performance was 17%.

4.4.5. Brasília (DF) - BR

a)

MEAN vs. ANN (RMSE)

Pn (7.74)
Pgp+n (7.83)
Pgp (7.97)
MaPgp (16.12)
MpPgp (18.05)
MaPgp+n (18.38)

(days)	1	3	6	9	12	15
MEAN:	8.58	12.22	10.8	13.58	13.02	14.37
BEST:	7.35	7.7	7.74	7.68	7.94	7.98
EXP:	Pn	Pn	Pgp+n	Pn	Pn	Pn
ANN:	2x2	2x8	1x1	1x3	3x10	4x1

MEAN (12.09)   BEST_OF (7.73)

b)

RMSE Skill Score

Lead-time (days)

MaPgp   MpPgp   Pgp+n
MaPgp+n   Pgp   Pn

c)

MEAN vs. ANN (RMSE)

Pn (7.74)
Pgp+n (7.83)
Pgp (7.97)

(days)	1	3	6	9	12	15
MEAN:	8.58	12.22	10.8	13.58	13.02	14.37
BEST:	7.35	7.7	7.74	7.68	7.94	7.98
EXP:	Pn	Pn	Pgp+n	Pn	Pn	Pn
ANN:	2x2	2x8	1x1	1x3	3x10	4x1

MEAN (12.09)   BEST_OF (7.73)

d)

RMSE Skill Score

Lead-time (days)

Pgp   Pgp+n   Pn

Three of six experiments that performed worst than MEAN were eliminated. ANN performed very well at all lead-times, especially at longer-range where improvements around 40% were achieved. Overall performance were around 36%.

4.4.6. Cuiabá (MT) - CB

a)

b)

c)

d)
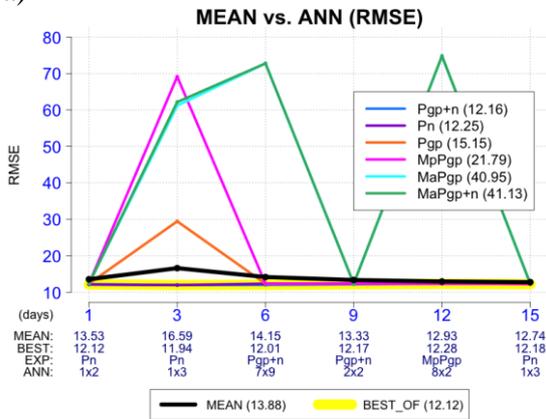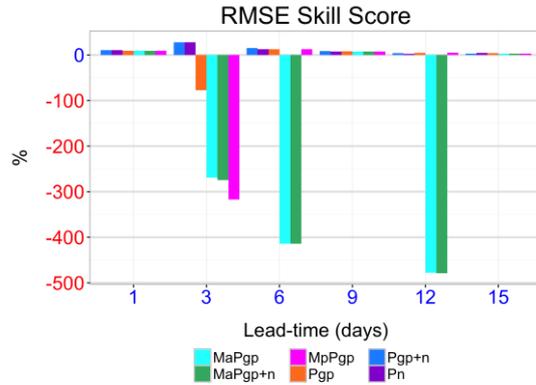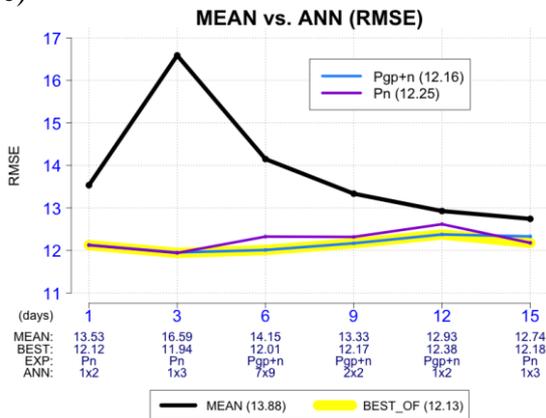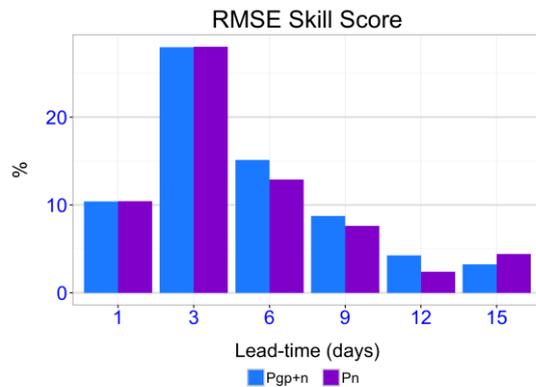
Four of six experiments that performed worst than MEAN were eliminated. A better performance was verified at the 3rd day of forecast, achieving almost 30%. A performance decay is noticed for all subsequent lead-times, which led the overall performance to be around 13%.

4.4.7. Curitiba (PR) - CT
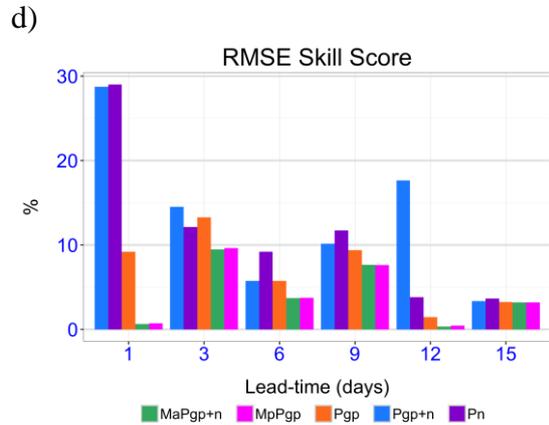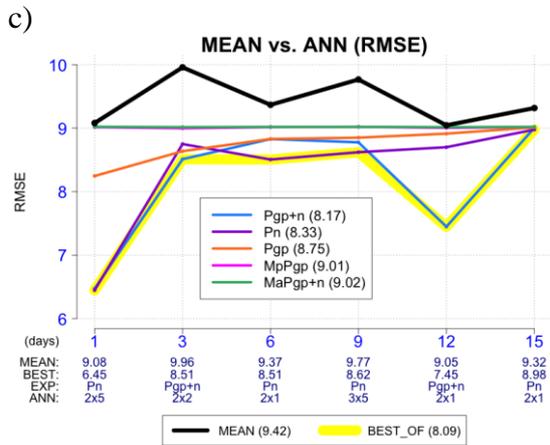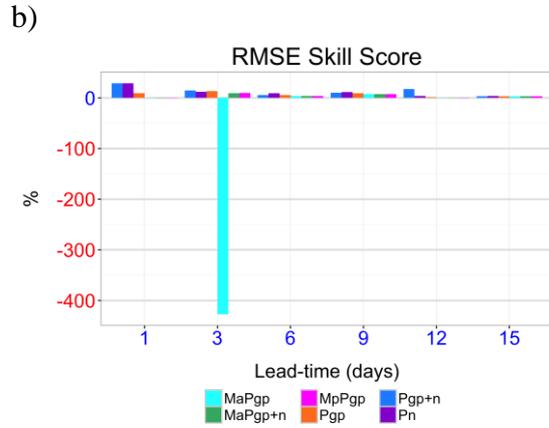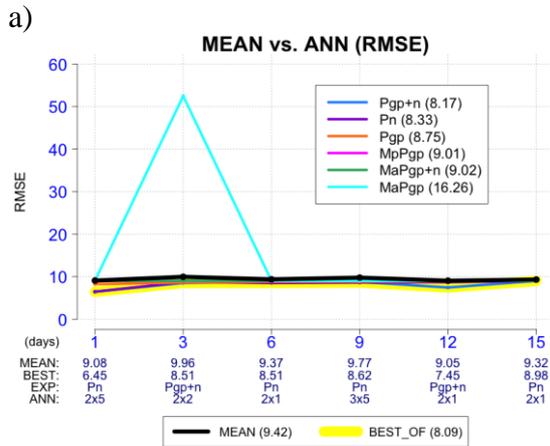
a)

b)

c)

d)

One of six experiments that performed worst than MEAN was eliminated. A better performance was verified at the 1st day of forecast, achieving almost 30%. A performance decay is noticed for all subsequent lead-times, which led the overall performance to be around 14%.

4.4.8. Montevideo (URU) - MV

a)
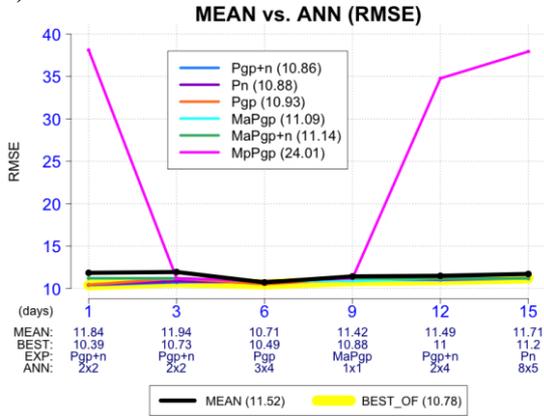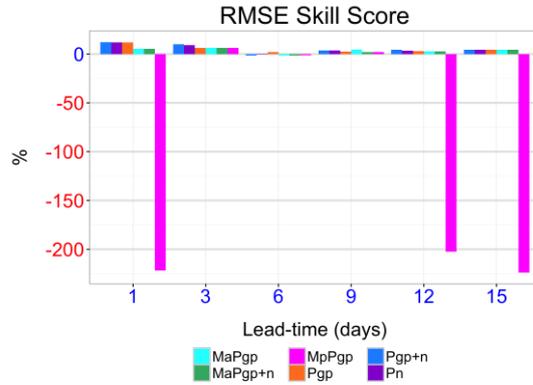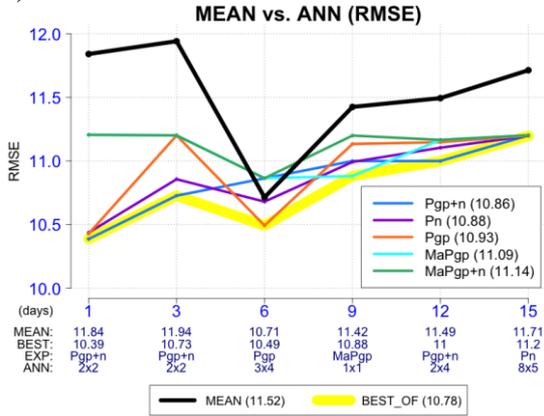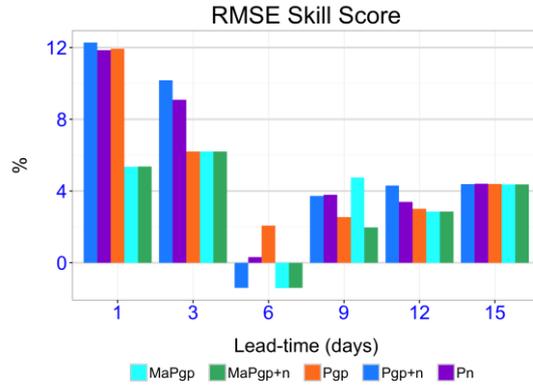
MEAN vs. ANN (RMSE)

(days)	1	3	6	9	12	15
MEAN:	11.84	11.94	10.71	11.42	11.49	11.71
BEST:	10.39	10.73	10.49	10.88	11	11.2
EXP:	Pgp+n	Pgp+n	Pgp	MaPgp	Pgp+n	Pn
ANN:	2x2	2x2	3x4	1x1	2x4	8x5

b)

RMSE Skill Score

c)

MEAN vs. ANN (RMSE)

(days)	1	3	6	9	12	15
MEAN:	11.84	11.94	10.71	11.42	11.49	11.71
BEST:	10.39	10.73	10.49	10.88	11	11.2
EXP:	Pgp+n	Pgp+n	Pgp	MaPgp	Pgp+n	Pn
ANN:	2x2	2x2	3x4	1x1	2x4	8x5

d)

RMSE Skill Score

There was no experiment that performed worst than MEAN, so none was eliminated. A better performance is noticed in the 1st day achieving 12%. In general, ANN performance was 6% only.
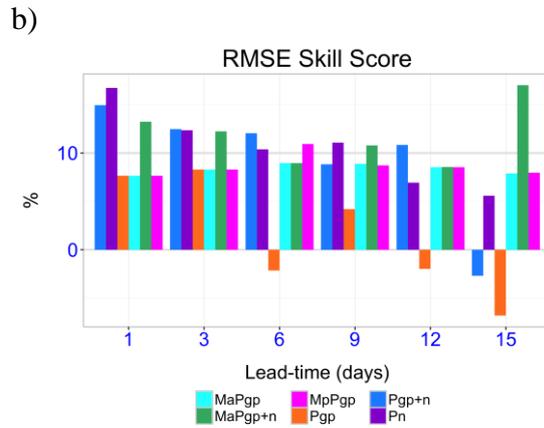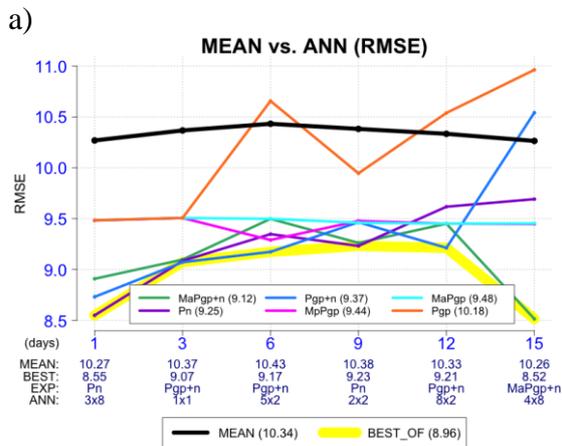
84

4.4.9. Salta (ARG) - ST

a)

MEAN vs. ANN (RMSE)

| MaPgp+n (9.12) | Pgp+n (9.37) | MaPgp (9.48) |
| Pn (9.25) | MpPgp (9.44) | Pgp (10.18) |

(days)	1	3	6	9	12	15
MEAN:	10.27	10.37	10.43	10.38	10.33	10.26
BEST:	8.55	9.07	9.17	9.23	9.21	8.52
EXP:	Pn	Pgp+n	Pgp+n	Pn	Pgp+n	MaPgp+n
ANN:	3x8	1x1	5x2	2x2	8x2	4x8

MEAN (10.34) BEST_OF (8.96)

b)

RMSE Skill Score

MaPgp MpPgp Pgp+n
MaPgp+n Pgp Pn

There was no experiment that performed worst than MEAN, so none was eliminated. A better performance is noticed in the 1st and last days of forecasting. In general, ANN performance was 13%.

4.4.10. Potosi (BOL) - PO
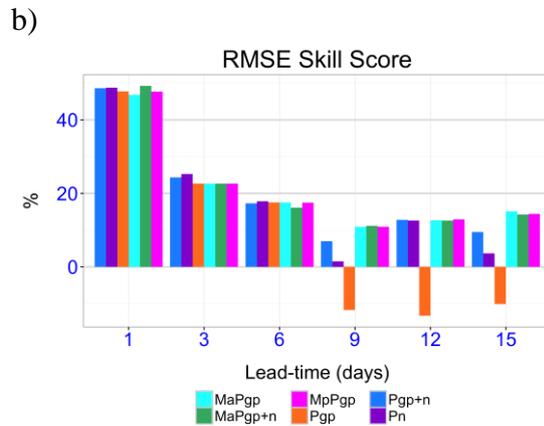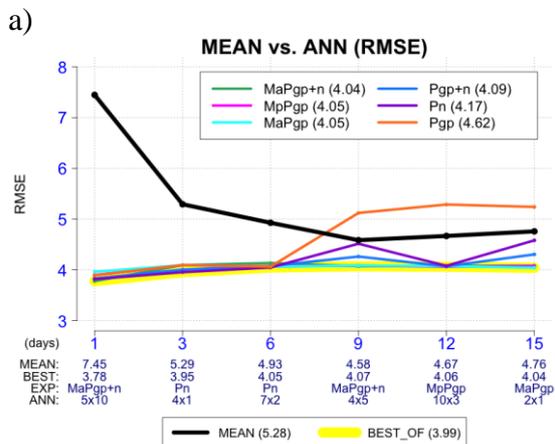
a)

MEAN vs. ANN (RMSE)

MaPgp+n (4.04)	Pgp+n (4.09)
MpPgp (4.05)	Pn (4.17)
MaPgp (4.05)	Pgp (4.62)

(days)	1	3	6	9	12	15
MEAN:	7.45	5.29	4.93	4.58	4.67	4.76
BEST:	3.78	3.95	4.05	4.07	4.06	4.04
EXP:	MaPgp+n	Pn	Pn	MaPgp+n	MpPgp	MaPgp
ANN:	5x10	4x1	7x2	4x5	10x3	2x1

MEAN (5.28) BEST_OF (3.99)

b)

RMSE Skill Score

MaPgp MpPgp Pgp+n
MaPgp+n Pgp Pn

There was no experiment that performed worst than MEAN, so none was eliminated. A better performance is noticed in the 1st forecasting day achieving around 30%. A performance decay is noticed for all subsequent lead-times, which led the overall performance to be around 24%.

4.4.11. São Paulo (SP) - SP

a)
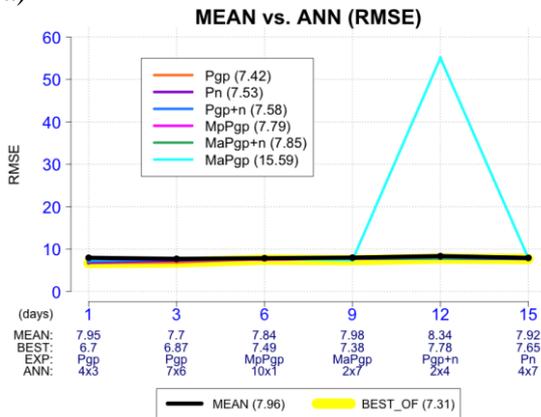
b)
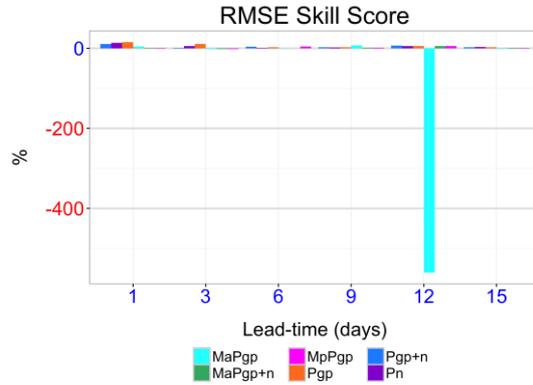
c)
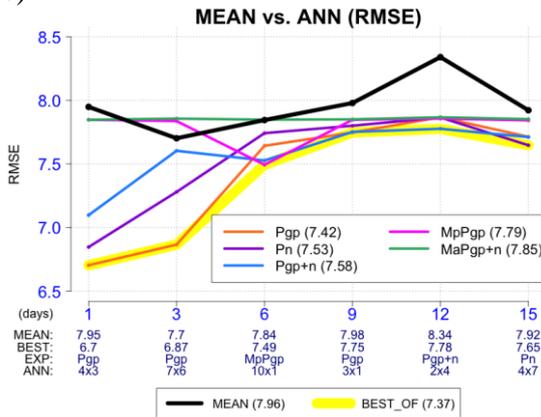
d)
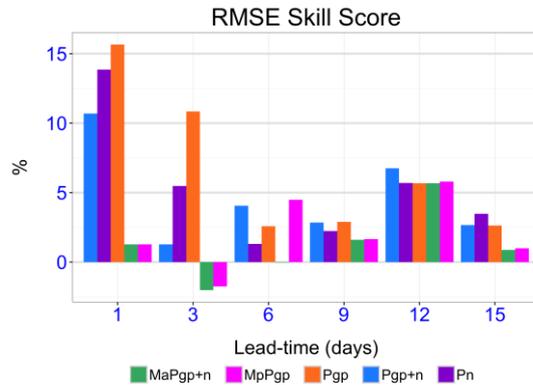
One of six experiments that performed worst than MEAN was eliminated. A better performance was verified at the 1st day of forecast, achieving almost 30%. A performance decay is noticed for all subsequent lead-times, which led the overall performance to be around 7% only.

4.5. **Results in numbers**

The same cutting criterion of Caxambu experiments was applied to the remaining locations, i.e., ignore every experiment whose overall performance was below MEAN overall performance for the location. Table 4.5 shows the ignored experiments of the remaining experiments.

Table 4.5   Ignored experiments at remaining locations.

LOC	EXP_ID	OVERALL RMSE	
		LOCATION	MEAN
AS	Pgp+n	13.71	13.68
	Pn	13.77	
	Pgp	14.52	
	MpPgp	14.17	
BA	MaPgp	18.39	14.57
	MpPgp	21.34	
BR	MaPgp	16.12	12.09
	MaPgp+n	18.38	
	MpPgp	18.05	
CB	Pgp	15.15	13.88
	MaPgp	40.95	
	MaPgp+n	41.13	
	MpPgp	21.79	
CT	MaPgp	16.26	9.42
MV	MpPgp	24.01	11.52
SP	MaPgp	15.59	7.96

LOC is the location where the experiment was run. EXP_ID is the experiment identifier.

After taking unperformed experiments out, a number of statistics can be done on the final resulting Table 4.6, which shows the best experiment and the ANN topology for each lead-time at every chosen location. The term "best experiment" means that it had the lowest RMSE value among all experiments, i.e., the winning experiment.

Table 4.6   Summarization of experiments vs. location vs. lead-time

LOC	1		3		6		9		12		15	
	N	L	N	L	N	L	N	L	N	L	N	L
CX	Pn		Pgp+n		Pgp		Pgp+n		Pgp+n		Pn	
	5	1	4	1	9	5	3	1	1	1	1	1
CT	Pn		Pgp+n		Pn		Pn		Pgp+n		Pn	
	2	5	2	2	2	1	3	5	2	1	2	1
SP	Pgp		Pgp		MpPgp		Pgp		Pgp+n		Pn	
	4	3	7	6	10	1	3	1	2	4	4	7
BR	Pn		Pn		Pgp+n		Pn		Pn		Pn	
	2	2	2	8	1	1	1	3	3	10	4	1
PO	MaPgp+n		Pn		Pn		MaPgp+n		MpPgp		MaPgp	
	5	10	4	1	7	2	4	5	10	3	2	1
CB	Pn		Pn		Pgp+n		Pgp+n		Pgp+n		Pn	
	1	2	1	3	7	9	2	2	1	2	1	3
CG	MpPgp		Pgp		Pgp		Pgp		Pgp		Pn	
	10	1	5	6	10	1	7	4	7	4	1	4
MV	Pgp+n		Pgp+n		Pgp		MaPgp		Pgp+n		Pn	
	2	2	2	2	3	4	1	1	2	4	8	5
ST	Pn		Pgp+n		Pgp+n		Pn		Pgp+n		MaPgp+n	
	3	8	1	1	5	2	2	2	8	2	4	8
BG	MpPgp		Pgp		Pn		MaPgp+n		MaPgp		MpPgp	
	5	3	7	1	6	10	4	3	2	2	10	6
BA	MaPgp+n		MaPgp+n		Pgp		MaPgp+n		Pgp		Pn	
	2	3	7	6	8	2	3	6	10	6	9	3
AS	MaPgp		MaPgp+n		MaPgp		MaPgp		MaPgp+n		MaPgp+n	
	8	2	7	4	7	2	8	5	4	2	6	8

LOC=location; N= # of neurons of the winner topology at the lead-time; L= # of layers

4.5.1. Experiments

Figure 4.1 illustrates the rank of winning experiments in general terms and how they are split by lead-time. Table 4.7 shows the rank in numbers.

Figure 4.1   Frequency chart: general winning experiments end by lead-time

a)   b)
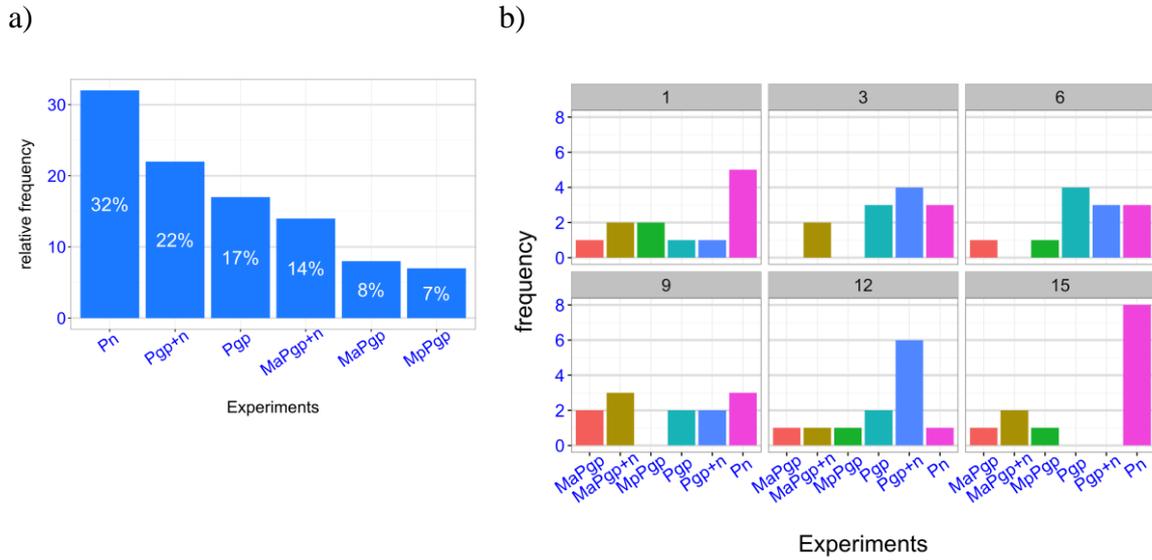
a) in general; b) by lead-time

Table 4.7   Rank table: winning experiments by lead-time.

EXP_ID	1	3	6	9	12	15	FREQ	REL.FREQ
Pn	5	3	3	3	1	8	23	32%
Pgp+n	1	4	3	2	6	0	16	22%
Pgp	1	3	4	2	2	0	12	17%
MaPgp+n	2	2	0	3	1	2	10	14%
MaPgp	1	0	1	2	1	1	6	8%
MpPgp	2	0	1	0	1	1	5	7%
Total	12	12	12	12	12	12	72	100%

FREQ=frequency, REL.FREQ=relative frequency.

4.5.2. Number of layers

Figure 4.2 illustrates the rank of the number of layers used in the ANN topology. It is sorted by the relative frequency of the number of hidden layers and summation is not 100% due to rounding issues. Figure 4.3 illustrates how it splits along lead-time.

Figure 4.2   Frequency chart: winning number of hidden layers.
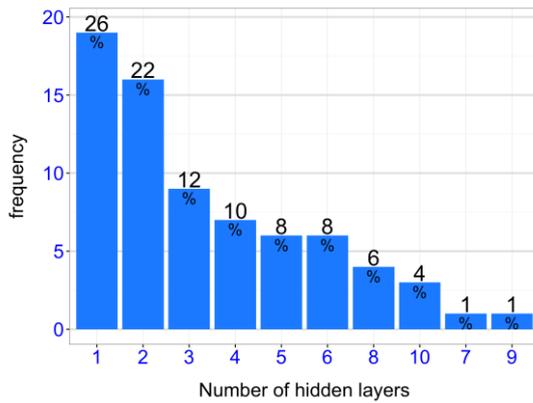
Figure 4.3   Frequency chart: winning number of hidden layers by led-time
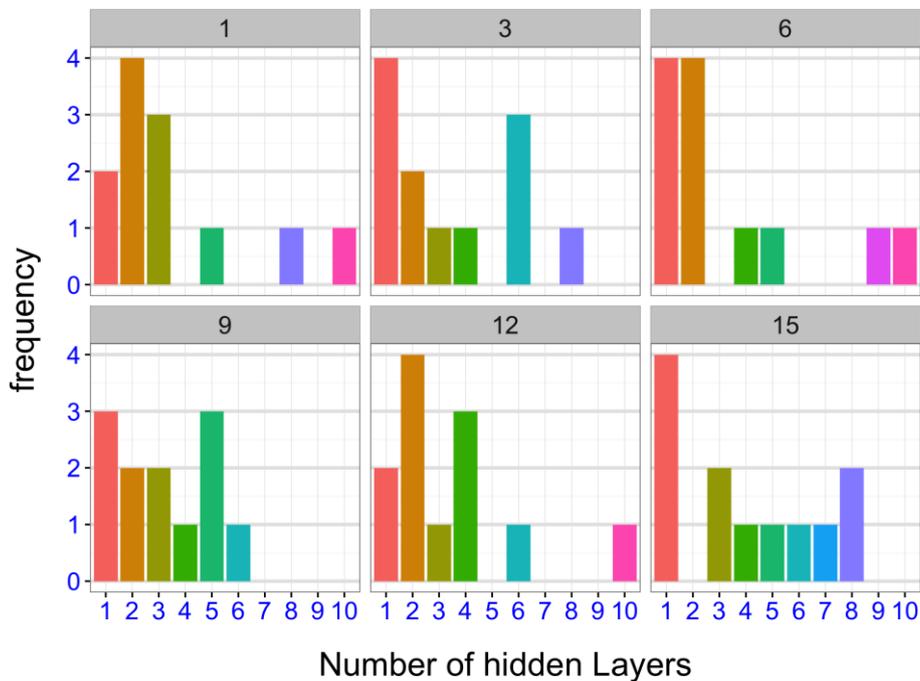
4.5.3. ANN topologies

Figure 4.4 illustrates the rank of ANN topologies (neuron-layer pairs) used in the experiments and Figure 4.5 illustrates how it is split by lead-time.

Figure 4.4   Frequency chart: winning ANN topologies.

Top 16 neurons x layers pairs (of 41)
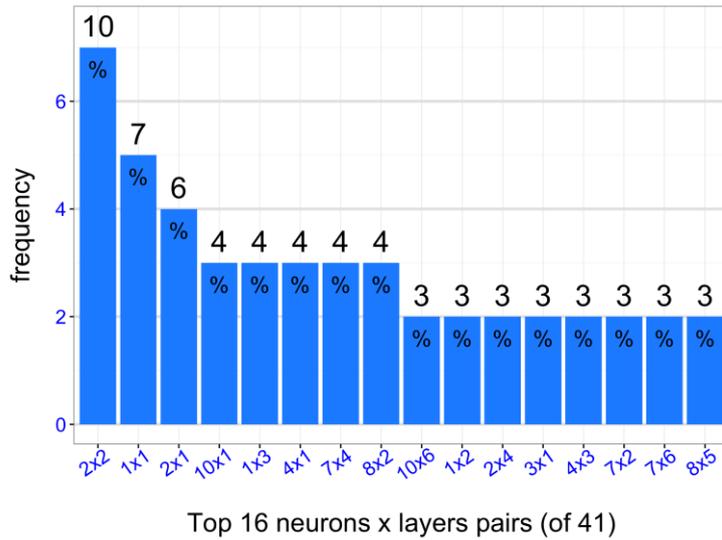
Figure 4.5   Frequency chart: winning ANN topologies by lead-time.

Top 16 ANN topologies (of 41)
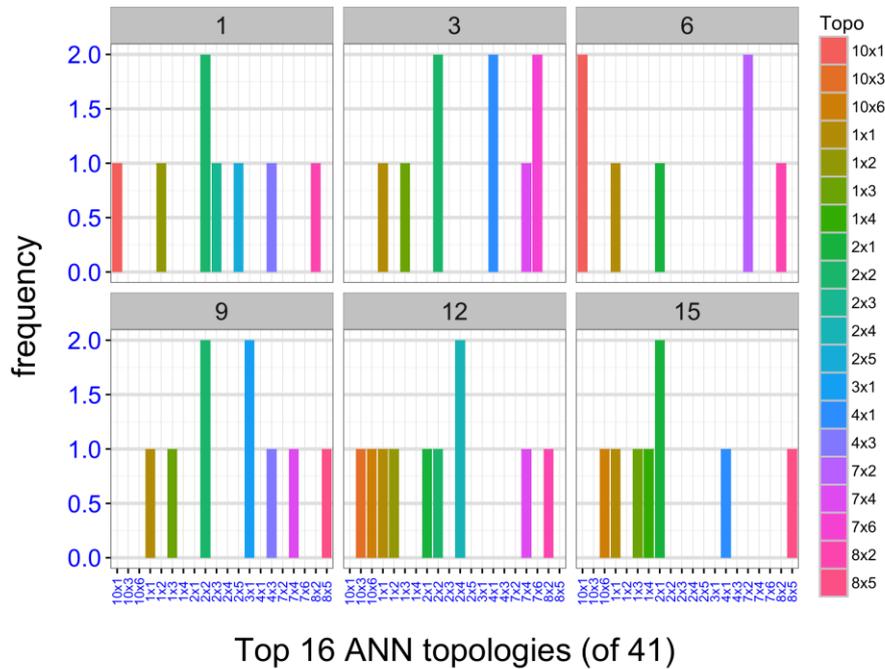
5 CONCLUSIONS

The quality of numerical weather prediction has been constantly improved. Much of this is undoubtedly credited to advances in intrinsic weather issues that are present in weather forecast modeling, but there is also a credit for computational post-processing systems, which are used to fine-tune the weather forecasting model outcome. The use of statistical techniques to post-process weather forecasting model's data is an essential requirement these days, once there are both renowned techniques as computational power to perform this processing. The purpose of using these techniques, such as neural networks, is to discover the best computational representation of the relationship among meteorological variables that are present in the system.

In this work, different experiments distinguishable by different arrangements of pure or derived CPTEC-EPS forecast variables were submitted as predictors to different structures of MLP neural networks. The goal was to build and assess different computational models of their relationship with OBS along the forecast horizon. The experiments were applied locally due to differences in weather behavior caused by different geographies.

Conclusions are now addressed from different perspectives.

5.1. **Predictors selection**

Results from Caxambu grid point (experiment set #2) showed that the simultaneous use of the variables vves, uves, tsfc, psnm, and prec led ANN to have a very poor performance, revealed by high RMSE rates, which indicates that they are not a good set of predictors, at least when used concomitantly. One possible explanation is the low level of correlation between chosen variables and OBS, as shown in the multi-charts in Figure 5.1.

Although WMO (2008) highly recommends Pearson's correlation coefficient as a score of forecasting precipitation amount, newer studies have found that its is

neither robust to deviations from linearity in a relationship nor resistant to outlying data (WILKS, 2011, p. 55-56). Thus, the correlation is measured by Spearman's rank correlation coefficient (or Spearman's rho, denoted by r_s or ρ).

Figure 5.1   Multi-info chart: correlation between variables and OBS at Caxambu.
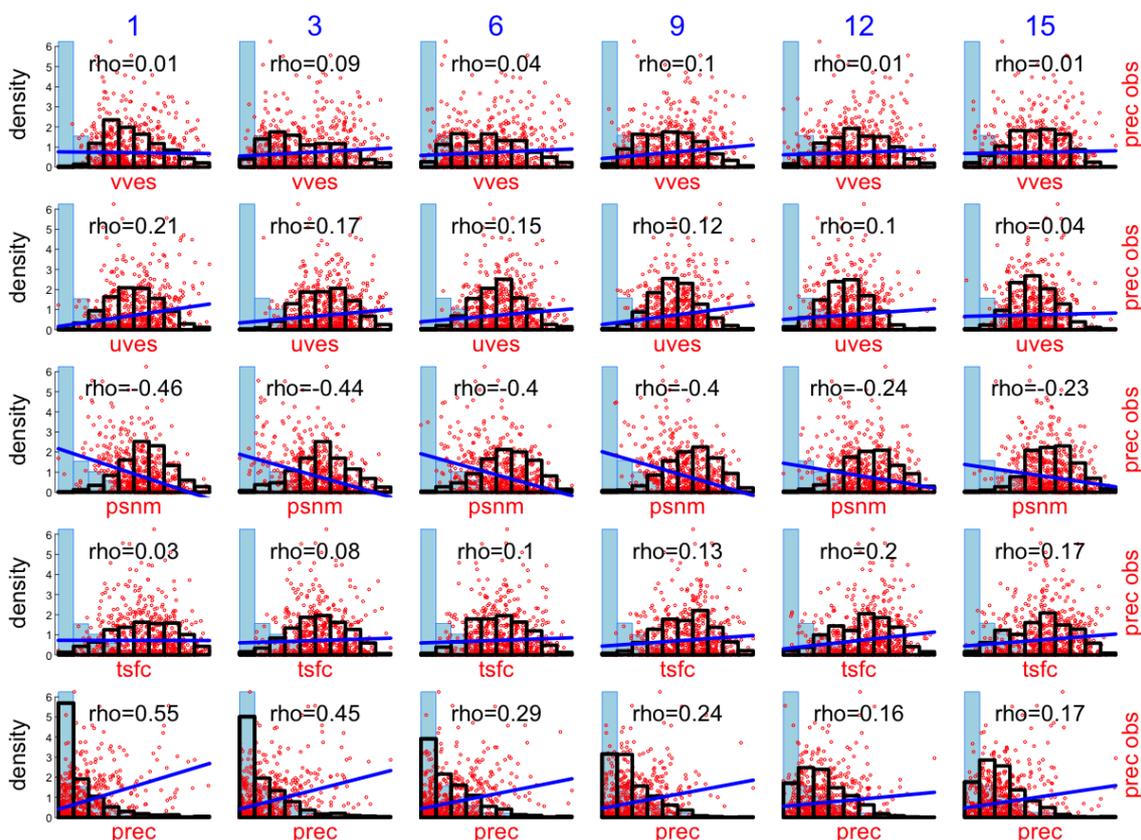
Each line is related to a variable, which is labeled in the x-axis; Each column is related to a lead-time, from the 1st to the 15th day, which is labeled in blue at the top of each column; Density histogram of OBS is in light blue, which remains unchanged in all plots; Thick black-lined rectangles are the density histogram of the independent variable; Red small circles are the scatter plot between the independent variable (x-axis) and OBS (y-axis); Blue thick line is the best-fitting straight line that represents the relationship between both vars. Spearman's correlation coefficient (rho) is shown in black at the top of each chart.

Source: author's production based on base::plot(), cor(), and lm().

According to the subjective classification proposed by Evans (1996) and Cohen (1988), the correlation is classified as very weak or weak for *rho < 0.3* and moderate for *rho ≈ 0.4*. Among the variables, only psnm and prec, which is actually the MEAN, have coefficients greater than 0.4, indicating that this set of predictors could not work well.

Developing a screening regression, i.e., selecting a good set of predictors from a pool of potential predictors, is not trivial and should be done with care. According to Wilks (2011), some issues must be taken into account, such as:

a) avoid selecting uncorrelated predictors with respect to the target variable because it increases variance of prediction and, thus, reduces precision;

b) although the tendency is to choose the most correlated set of predictors with respect to the target variable, choosing too many predictors can cause over fitting, i.e., good generalization in training but poorly generalization when facing new data;

c) avoid choosing mutually correlated predictors, because redundant information leads to poor estimates.

The final conclusion about the predictors selection method applied in this work is that the search for predictors should have been done in a more critical way. First of all, more variables should have been extracted from the prognostic fields of the CPTEC-EPS, then it could have been used some measurements such as Pearson or Spearman correlation or even apply the techniques reported by Applequist (2002) in order to asses the importance of the predictor to the target variable. In addition, the use of psnm should have been more explored, once it was already available and has revealed to be more correlated to OBS even than MEAN, in most lead-times.

5.2. Averaged vs. members variables

According to the rank of winning experiments (Figure 4.1), the experiment that used MEAN as predictors performed better than the experiments that used raw precipitation data from members. A possible reason can be credited to the fact that raw original member's data are, indeed, deterministic predictions and its variability can be too different from OBS, making the relationship to be even more complex. Figure 3.3 and its interpretation below show that MEAN is more correlated to OBS than any individual member.

The conclusion regarding this issue is that the complexity of deterministic predictions was carried to inside the ANN system, making MEAN to be much more correlated to OBS than data from members. To complement, other measurements of central tendency as median and mode could also be taken into account to generate additional derived data, since the arithmetic mean is highly sensitive to skewed data distributions, as are the precipitation data.

5.3. Data from adjacent grid points

Data from neighbors explain the relationship with OBS better than data from the target grid point. Figure 4.1 shows that, in general, adding data from adjacent grid points enhances the skill of the set of predictors to calibrate MEAN. It can be noticed in the first three winning experiments, which are the experiments composed by MEAN, and also in the last three worst experiments, which are the ones that used members' data as predictors.

As detailed in Figure 4.2, with a exception of the 6th day, the inclusion of data from adjacent grid points made the set of predictors perform better than using only data of the target grid point, highlighting days 1, 12, and 15, whose difference in the number of occurrence is greater.

The conclusion regarding this issue is that data from adjacent grid points should be used, however, the use of all eight neighbors together should be reevaluated

and applying screening regression to them should be considered, in order to filter the most important ones for each grid point and lead-time.

5.4. **Processing by location**

The location filter used to create the train and test datasets is considered correct, since the relationships between weather variables can vary according to the topography of the terrain, which affects the baroclinic stability. The disadvantage is that it makes the training set to have only 632 cases, which can be not enough to set the model with the most appropriate weights. To solve this problem, one solution to be explored in a future work can be to apply the clustering technique on the entire sample cases according to behavioral similarity. This would make the 513.184 samples cases (28 x 29 grid points x 632 days) to be grouped in few groups, increasing the number of training cases of the ANN.

5.5. **ANN**

Rules of thumb assert that ANN topologies composed by three hidden layers is enough to solve the majority of complex relationships. However, according to Figure 4.3 and detailed in Table 4.6, approximately 38% of winning ANN topologies were composed by at least four hidden layers.

Such expressive amount has lead the author to conclude that although the loop to create one hundred ANN topologies, reaching up to a 10x10 topology, can cause some impact, it was worth exploring ANN capabilities. In strict sense, tests on ANN topology had exploratory purposes and the approach was also motivated by the lack of consensus about the suitable ANN topology.

Another significant fact is the number of winning topologies composed by only one and two layers. According to Figure 4.3, such ANN topologies are responsible for the solution of about 50% of cases. According to Figure 4.4, this fact is emphasized at the 6th day, in which the factor increases to 66%. It can

also be seen that, in general, chart bars lay on the left side of lead-time charts, just where the ANNs composed by smaller number of hidden layers are. To the author, this scenario corroborates both assumptions that different problems require different solution as well as ANNs composed by few layers are powerful enough to solve highly complex relationships.

According to Figure 4.5, the total number of topologies used to solve 72 experiments (12 locations x 6 lead-times) sums up to 41, and the most used is a 2x2 topology, in 10% of cases, followed by the 1x1 topology, which is the minimum possible topology, with hidden layers.

The conclusion regarding this issue is that, on future work, an exploration of ANN topologies composed by a high amount of neurons up to 2 hidden layers, as proposed by Cybenko (1989), can be a good alternative to avoid de loop of ANN topologies and tune the mechanism in a more detailed way. Furthermore, a hint of the best configuration for each location+lead-time pair can be investigated as done in Anochi (2015), which aims to make the configuration phase easier.

The problem this work deals with is considered very difficult given that the chosen variables are very weakly correlated to the target observed value. In order to try to "help" the ANN, an architecture that keeps memory of prior events to simulate temporal behavior could be explored in a future work. This architecture is present in Recurrent Neural Networks (or Hopfield networks) (HOPFIELD 1982).

5.6. R platform experience

The choice of using R for the whole process is controversial. While providing high-level solutions avoiding the implementation of consolidated tasks such as analysis charts and neural network itself, it presents limitations in flexibility, and prevents certain needs to be achieved. The use of a third-part package in order to run ANNs has presented limitations in the following situations:

a) lack of information about poor performance: as described in experiment set #2, the reason for high RMSE values could not be identified and discussion is based on suppositions about getting into a local minimum or saturation;

b) trade-off between fast execution and error evolution during training: as one hundred ANN topologies were run for each experiment and lead-time, the output of the training process provided by the neuralnet function had to be disabled once the whole process was demanding too much time to complete, making the methodology to be not feasible;

c) lack of configuration on activation function: fine tuning issues such as the slope of the sigmoid function and even the use of alternative functions such as Piece-wise or ReLU is missing;

d) lack of *momentum* configuration: the official user manual says nothing about momentum and advanced configurations in order to avoid local minimum is missing;

e) the conclusion regarding this issue is that the tool that plays the role of protagonist in such a complex process of dealing with highly non-linear relationships present in this work should provide means to configure its working process, under penalty of lack of clarity in some details. The tool could be implemented even in R programming language, but should be flexible enough to provide full control of the operation for the researcher.

5.7. **Final remarks**

The main contribution of this work is that it allows an improvement in the ensemble mean forecast at every lead-time based only in computational effort, involving no physics concepts. This improvement is verified at every lead-time

and the difference of the performance among the locations can be explained by the amount of rainy days verified at each of them. As can be seem in Figure 5.2, in which X-axis is sorted by the amount of rain verified at each location from rainner to dryer, both tendency red lines indicate that the rainiest is the location the best the network performance is.

Figure 5.2   Multi-info chart: correlation between variables and OBS at Caxambu.
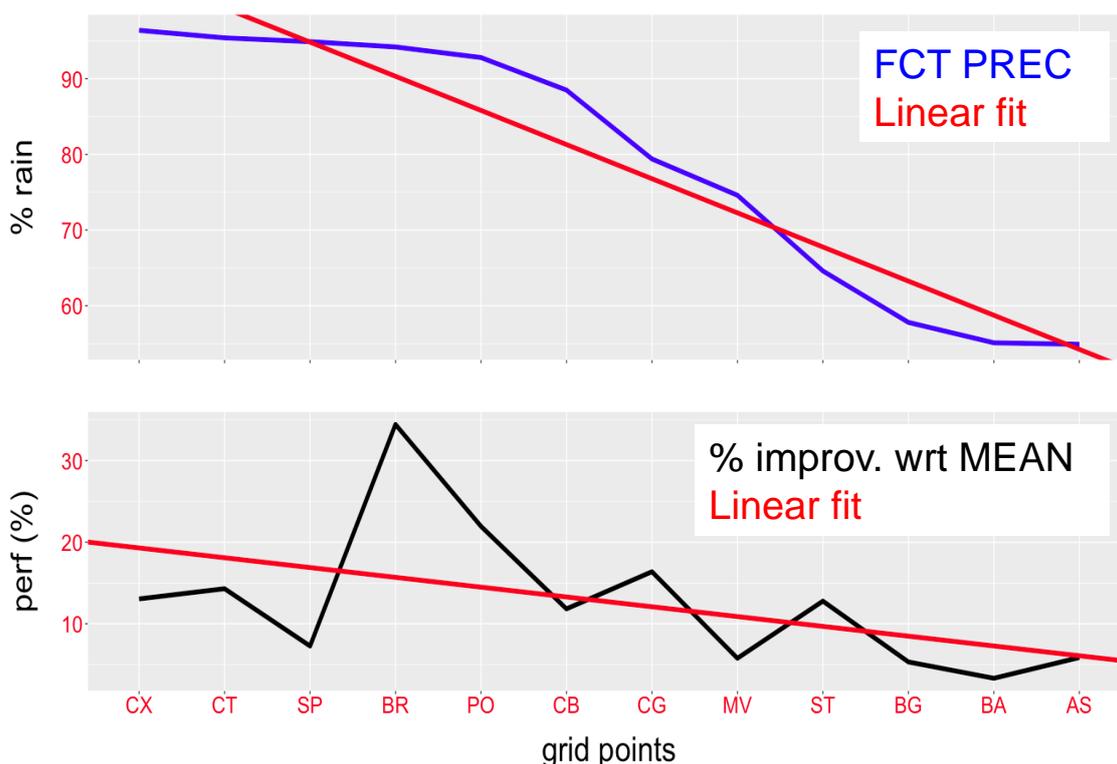
X-axis sorted by rainner location to dryer. Top box shows the percentage of rainy days at each location (blue line). Bottom box shows de percentage improvements in ANN models with respect to the MEAN at each location (black line). Tendency lines are in red.

Source: author's production.

Yet with respect to improvements, the greater one was achieved for the 24hs and 72hs forecast, as can be seem in Figure 5.3. This leads to a overall improvement of about 13% with respect to MEAN.

Figure 5.3   Percentage improvements at each lead-time
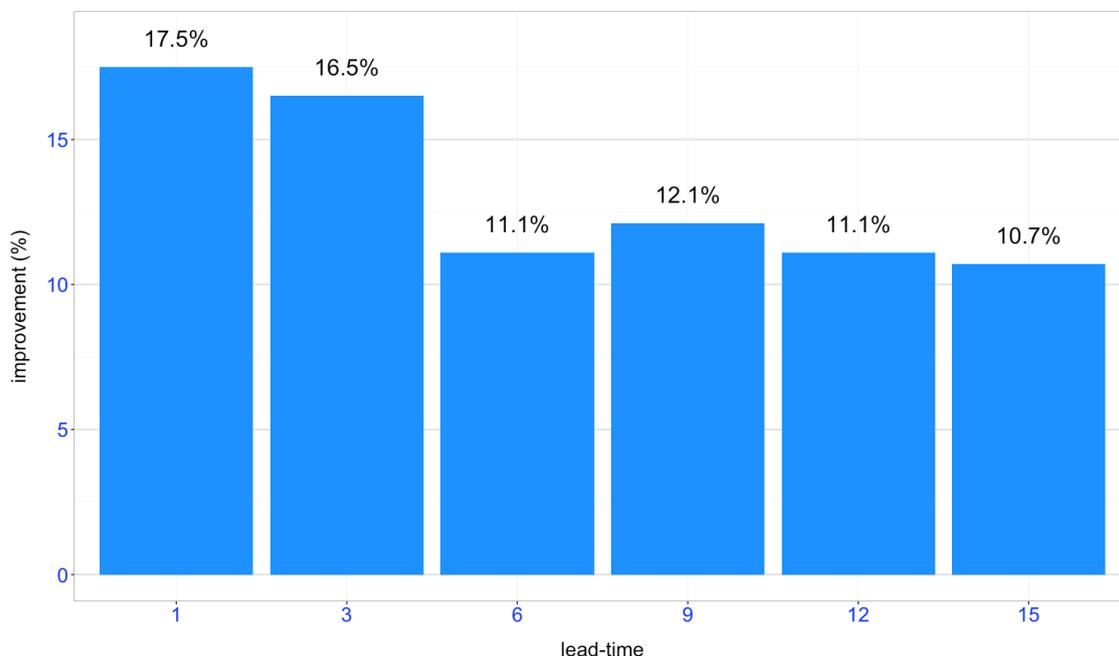
Source: author's production.

The specific goals of assessing the correlation between variables and OBS, investigating the best combination of predictors and ANN topology that best calibrate MEAN along lead-times were achieved. As also were the check of whether neighbor's data influence the precipitation at the target grid point.

Moreover, the general goal of discovering an optimal input-based model based on both set of predictors and ANN topology in order to improve MEAN forecast for the lead-times in the forecast horizon by using data from CPTEC-EPS was also achieved. As can be seen in Table 4.6 and in the charts of the experiments, location and lead-time pairs have different set of predictors and ANN topologies as most suitable solutions. This fact corroborates the assumption that weather variables have different relationships with each other for distinct location and lead-time. To solve this, different ways of thinking are required and, in this case, they are represented by predictors and topologies.

Directions of future work are based on addressing the shortcomings that have been identified in the methodology. Although it is considered promising, the exposed adjustments and further tests are necessary to consolidate it as an operational tool and help forecasters in the complex task of predicting weather. The mentioned screening regression, the issues about ANN fine tuning, and also providing more samples cases in order to feed the training process with more data compose the main requirements to improve this pioneer methodology. Figure 5.4 shows a simplified scheme that could serve as an operational workflow of the entire process of training and calibrating the MEAN according to the methodology here proposed.

Figure 5.4   Simplified operational workflow
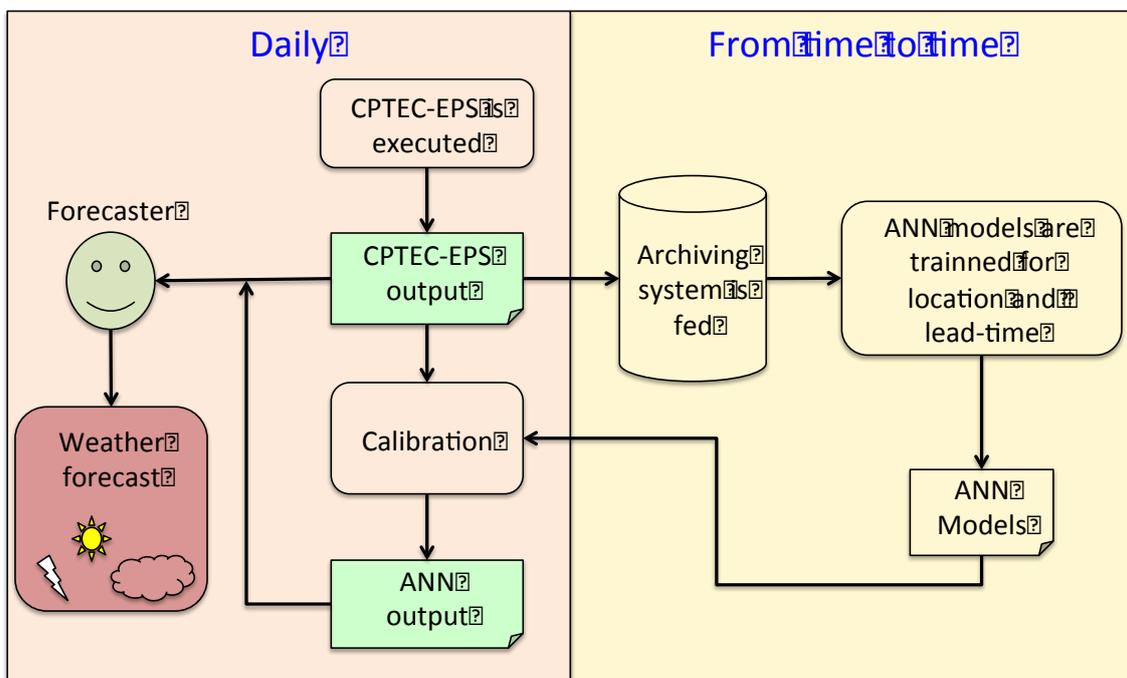
Salmon box represents daily forecasting procedures that start with the execution of the CPTEC-EPS. Its output is archived and used to build the ANN models from time to time (light yellow box). Back in daily routines, the CPTEC-EPS output is calibrated by the models generating new output and both are used by forecasters to predict the weather.

Source: author's production.

LIST OF REFERENCES

ABU-MOSTAFA, Y. S.; MAGDON-ISMAIL, M.; LIN, H-T. **Learning from data**, AML Book, 2012.

ABRAHART, R. J.; SEE, L. M. Neural network modeling of non-linear hydrological relationships. **Hydrology and Earth Systems Sciences**, v. 11, p. 1563:1579, 2007.

AHMAD, I.; ANSARI, M. A.; MOHSIN, S. Performance comparison between backpropagation algorithms applied to intrusion detection in computer network systems. In: WSEAS INTERNATIONAL CONFERENCE ON NEURAL NETWORKS (NN'08), 9., 2008, May 2-4, Sofia, Bulgaria. **Proceedings**… Stevens Point, Wisconsin, USA (WSEAS, 2008). p. 231-236.

AMERICAN METEOROLOGICAL SOCIETY (AMS). **Glossary of meteorology**. 2015. Available online at: http://glossary.ametsoc.org. Accessed in: Apr 2016.

ANOCHI, J. A. **Previsão climática de precipitação por redes neurais autoconfiguradas**, 2015. pp. 161, Tese (Doutorado em Computação Aplicada) - Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 2015.

APPLEQUIST, S.; GAHRS, E. G.; PFEFFER, R.L.; NIU, X-F. Comparison of methodologies for probabilistic quantitative precipitation forecasting. **Weather and Forecasting**, v. 17, p. 783–799, 2002.

BARNES, S. L. **Mesoscale objective analysis using weighted time-series observations**. NOAA Tech. Memo. ERL NSSL-62, National Severe Storms Laboratory, Norman, OK, 60 pp. [NTIS COM-73-10781], 1973.

BARROS V.; CLARKE R.; DIAS P. S. **Climate change in the La Plata Basin**, Project SGP II 057: Trends in the hydrological cycle of the Plata basin: Raising awareness and new tools for water management of Inter American Institute On Global Change (IAI). 1st ed., CIMA/CONICET-U.B.A., Buenos Aires, 2006.

BARROS V.;; DOYLE, M.; CAMILLONI, I. **Potential impacts of climate change in the Plata Basin**, Regional Hydrological Impacts of Climatic Variability and Change, In: IAHS SCIENTIFIC, 7., 2005, Foz do Iguaçu, Brazi, Proceedings....

BERBERY, E. H.; BARROS, V. The Hydrologic Cycle of the La Plata Basin in South America, **Journal of Hydrometeorology**, v. 3, p.630-645, 2002.

BISHOP, C. M. **Neural networks for pattern recognition**. Oxford University Press, Inc. New York, NY, USA, 1995.

BJERKNES V. Das Problem der Wettervorhersage, betrachtet vom Standpunkte der Mechanik und der Physik (The problem of weather forecasting as a problem in mechanics and physics). **Meteorologische Zeitschrift**, 21, 1–7, 1904. (English translation by Y. Mintz, 1954, reproduced in The Life Cycles of Extra tropical Cyclones, 1999, Amer. Meteor. Soc.).

BOCCHIERI, J. R. **The use of Model Output Statistics (MOS) for predicting the probability of heavy snow**. NOAA, TDL Office Note 77-18, 1977.

BRAVO, J. M.; PAZ A. R.; COLLISCHONN W.; UVO C. B.; PEDROLLO O. C.; CHOU S. C. Incorporating forecasts of rainfall in two hydrologic models used for medium-range stream flow forecasting. **Journal of Hydrologic Engineering**, v. 14, n. 5, p. 435-445, 2009.

BUIZZA, R.; HOUTEKAMER P. L.; TOTH Z.; PELLERIN G.; WEI M.; ZHU Y. A comparison of the ECMWF, MSC, and NCEP global ensemble prediction systems. **Monthly Weather Review**, v. 133, p. 1076–1097, 2005.

CALAWAY, R.; WESTON, S. **Package "foreach". Providing foreach looping construct for R**, v1.4.3, 2015. Available online at https://cran.r-project.org/web/packages/foreach/, Accessed in Apr 2016.

CALAWAY, R.; WESTON, S.; TENENBAUM, D. **Package "doParallel". oreach Parallel Adaptor for the 'parallel' Package**, 1.0.10, 2015. Available online at https://cran.r-project.org/web/packages/doParallel, Accessed in Apr 2016.

CARTER, G. M.; DALLAVALLE J. P.; GLAHN, H. R. Statistical forecasts based on NMC's numerical weather prediction system. **Weather and Forecasting**, v. 4, p. 401-412, 1989.

CAVALCANTI, I. F. A.; MARENGO, J. A.; SATYAMURTY, P.; NOBRE, C. A.; TROSNIKOV, I.; BONATTI, J. P.; MANZI, A. O.; TARASOVA, T.; PEZZI, L. P.; D'ALMEIDA, C.; SAMPAIO, G.; CASTRO, C. C.; SANCHES, M. B.; CAMARGO, H. Global Climatological Features in a Simulation Using the CPTEC–COLA AGCM. **Journal of Climate**, v. 15, n. 21, 2002.

CHADWICK, R.; COPPOLA, E.; GIORGI, F. An artificial neural network technique for downscaling GCM outputs to RCM spatial scale. **Nonlinear Processes in Geophysics**, v. 18, p. 1013–1028, 2011

CHARNEY, J. G. On the scale of the atmospheric motions. **Geofysiske publikasjoner**, v. 17 :2, 1948.

_____. Progress in dynamic meteorology. **Bulletin of the American Meteorological Society**, v. 31, p. 231-236, 1950.

_____. **Dynamic forecasting by numerical process**. Compendium of Meteorology, American Meteorological Society, Boston, p. 470–482, 1951.

CHEN, C.; KAMENKOVICH, I. Effects of Topography on Baroclinic Instability. **Journal of Physical Oceanography**, v. 43, p. 790–804, 2013.

INTERNATIONAL CENTER OF TROPICAL AGRICULTURE (CIAT). **Atlas de indicadores ambientales y sustentabilidad para América Latina y el Caribe**. Bogotá, PNUMA. Available online at http://www.cima.fcen.uba.ar/ClarisLPB/RDA_list.php?ti=ob&su=geo, Accessed in Apr 2016, 1998.

COHEN, J. **Statistical power analysis for the behavioral sciences**, 2nd ed. Hillsdale, NJ, 1988.

CORONEL, G.; MENÉNDEZ A. Physiography and hydrology. In **Barros et al. (2006)**, 2006.

CORTIVO, F. D.; CHALHOUB, E. S.; VELHO, H. F. C. **Comparison of two learning strategies for a supervised neural network**. In: International Symposium on Uncertainty Quantification and Stochastic Modeling, 1., 2012, São Sebastião, São Paulo. **Proceedings**..., p. 366-380.

COUTINHO, M. M. **Previsão por conjuntos utilizando perturbações baseadas em componentes principais**. 1999. 136 p.. Dissertação (Mestrado em Meteorologia) - Instituto Nacional de Pesquisas Espaciais (INPE), São José dos Campos, 1999.

CYBENKO, G. Approximaton by superpositions of a sigmoidal function, **Mathematics of Control, Signals, and Systems**, v. 2, pp. 303-314, Springer-Veriag New York Inc., 1989.

DA SILVA, A. P. A.; REIS, A. J. R.; EL-SHARKAWI, M. A.; MARKS, R. J. Enhancing neural network based load forecasting via preprocessing. In: International Conference on Intelligent System Application to Power Systems, 2001, Budapest, Hungary, **Proceedings**..., p. 118-123.

DALLAVALLE, J. P.; JENSENIUS JR J. S.; GILBERT S. A. **NGM based MOS guidance**. The FOUS14/FWC message. Technical Procedures Bulletin 408, NOAA/National Weather Service, Silver Spring, 9 pp., 1992.

DASTORANI, M. T.; AFKHAMI, H.; SHARIFIDARANI, H.; DASTORANI M. Application of ANN and ANFIS models on dry land precipitation prediction (Case Study: Yazd in Central Iran). **Journal of Applied Sciences**, v. 10, p. 2387–2394, 2010.

DAWSON, C. W.; WILBY, R. L. Hydrological modeling using artificial neural networks. **Progress in Physical Geography**, v. 25:1, p. 80–108, 2001.

DENNIS, J. E.; MORÉ, J. J. Quasi-Newton methods, motivation and theory. **Society for Industrial and Applied Mathematics Review**, v. 19, p. 46–89, 1977.

EADY, E. **Long waves and cyclone waves**. Tellus, v. 1, p. 33–52, 1949.

EBERT, E. E. Ability of a poor man's ensemble to predict the probability and distribution of precipitation. **Monthly Weather Review**, v. 129, p. 2461–2480, 2001.

ECKEL, F. A.; WALTERS M. K. Calibrated probabilistic quantitative precipitation forecasts based on the MRF ensemble. **Weather and Forecasting**, v. 13, p. 1132–1147, 1998.

EL-SHAFIE A.; JAAFER O.; SEYED A. Adaptive neuro-fuzzy inference system based model for rainfall forecasting in Klang River, Malaysia. **International Journal of Physical Sciences**, v. 6:12, p. 2875–2888, 2011.

EPSTEIN, E. S. **Stochastic dynamic prediction**. Tellus, v. 21, p. 739–759, 1969.

EUROPEAN VIRTUAL ORGANISATION FOR METEOROLOGICAL TRAINING (EUMETCAL). **Verification of probability and ensemble forecasts - The Rank Histogram** – Definition and use, Available online at http://www.eumetcal.org/resources/ukmeteocal/verification/www/english/msg/ver_prob_forec/uos4b/uos4b_ko1.htm, 2011. Accessed in Apr 2016.

EVANS, J. D. **Straightforward statistics for the behavioral sciences**. Pacific Grove, CA: Brooks/Cole Publishing, 1996.

FALL, S.; NIYOGI, D.; MOHANTY, U.; KUMAR, A. Application of weather prediction models for hazard mitigation planning: a case study of heavy off-season rains in Senegal, Springer, **Natural Hazards**, v. 41, n.1, p. 227-243, 2007.

FAUSETT, L. **Fundamentals of neural networks**: architectures, algorithms, and applications, London, Prentice Hall Inc., 1994.

FREIWAN, M.; CIGIZOGLU, H. K. **Prediction of total monthly rainfall in Jordan using feed forward backpropagation method**. Fresenius Environment Bulletin, v. 14:2, p. 142–15, 2005.

FRENCH, M. N.; KRAJEWSKI, W. F.; CUYKENDAL, R. R. Rainfall forecasting in space and time using a neural network. **Journal of Hydrology**, v. 137, p. 1–37, 1992.

FRITSCH, S.; GUENTHER, F. **Package "neuralnet"**. Training of neural network, v1.32, 2012. Available online at https://cran.r-project.org/web/packages/ neuralnet/, Accessed in Apr 2016.

GAGNE, D. J.; MCGOVERN, A.; XUE, M. Machine Learning Enhancement of Storm-Scale Ensemble Probabilistic Quantitative Precipitation Forecasts, **Weather and Forecasting**, v. 29, p. 1024-1043, 2013.

GALLUS, W. A.; SEAGAL, M. Does increased predicted warm-season rainfall indicate enhanced likelihood of rain occurrence? **Weather and Forecasting**, v. 19, p. 1127–1135, 2004.

GARDNER, M.W.; DORLING, S. R. Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences, **Atmospheric Environment**, v. 32, Issues 14–15, p. 2627-2636, 1998.

GLAHN, H. R.; DALLAVALLE, J. P. **MOS-2000**. TDL Office Note 00-1, National Weather Service, NOAA, U.S. Department of Commerce, 2000.

GLAHN, H. R.; LOWRY, D.A. The use of model output statistics (MOS) in objective weather forecasting. **Journal of Applied Meteorology**, v. 11, p. 1203-1211, 1972.

GNEITING, T.; RAFTERY A. E.; Weather forecasting with ensemble methods. **Science**, v. 310, p. 248–249, 2005.

GNEITING, T.; BALABDAOUI, F.; RAFTERY, A. E. Probabilistic forecasts, calibration and sharpness. **Journal of the Royal Statistical Society**. B 69, *Part* 2, *pp.* 243–268, 2007.

GRIMIT, E. P.; MASS, C. F. Initial results of a mesoscale short-range ensemble forecasting system over the Pacific Northwest. **Weather and Forecasting**, v. 17, p. 192–205, 2002.

GUARNIERI R. A.; PEREIRA E. B.; MARTINS F. R.; CHAN, C. S. Previsões de radiação solar utilizando modelo de mesoescala: refinamento com redes neurais. In: Congresso Brasileiro de Energia Solar - CBENS, Fortaleza, Brazil. **Proceedings**... 2007.

HALL, T.; BROOKS, H. E.; DOSWELL III, C. A. Precipitation Forecasting Using a Neural Network. **Weather and Forecasting**, v. 14, p. 338-345, 1999.

HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, I. H. The WEKA Data Mining Software: An Update; **SIGKDD Explorations**, Volume 11, Issue 1, 2009.

HAMILL T. M.; COLUCCI S. J.; Verification of Eta-RSM short-range ensemble forecasts. **Monthly Weather Review**, v. 125, p. 1312–1327, 1997.

_____. Evaluation of Eta-RSME ensemble probabilistic precipitation forecasts. **Monthly Weather Review**, v. 126, p. 711–724, 1998.

HAMILL, T. M.; SNYDER, C.; MORSS, R. E. A comparison of probabilistic forecasts from bred, singular-vector, and perturbed observation ensembles. **Monthly Weather Review**, v. 128, p. 1835–1851, 2000.

HAMILL, T. M.; WHITAKER, J. S. Ensemble Calibration of 500-hPa Geopotential Height and 850-hPa and 2-m Temperatures Using Reforecasts. **Monthly Weather Review**, v. 135, p. 3273-3280, 2006.

HAMILL, T. M.; WHITAKER, J. S.; WEI, X. Ensemble reforecasting: Improving medium-range forecast skill using retrospective forecasts. **Monthly Weather Review**, v. 132, p. 1434–1447, 2004.

HAYKIN, S. **Neural Networks: A Comprehensive Foundation**, 842 pp., Prentice-Hall, Old Tappan, N. J., 1999.

HONG, W. C. Rainfall forecasting by technological machine learning models. **Applied Mathematics and Computation**, v. 200:1, p. 41–57, 2008.

HONG, S.-Y. **Next-Generation Numerical Weather Prediction**. Bulletin of the American Meteorological Society, v. 93, p. ES6-ES9, 2010

HOPFIELD, J. J. **Neural networks and physical systems with emergent collective computational properties**. Proceedings of the National Academy of Sciences of the United States of America, v. 79, 2554-2558, 1982.

HOU, D.; KALNAY, E.; DROEGEMEIER, K. K. Objective verification of the SAMEX'98 ensemble forecasts. **Monthly Weather Review**, v. 129, p. 73–91, 2001.

HSIEH, W. W. **Machine Learning Methods in the Environmental Sciences**: Neural Networks and Kernels. University of British Columbia, Vancouver, 2009.

HUFFMAN, G. J.; COAUTHORS The TRMM: Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. **Journal of Hydrology and Meteorology**, v. 8, p. 38–55, 2007.

HUNG, N. Q.; BABEL, M. S.; WEESAKUL, S.; TRIPATHI, N.K. An artificial neural network model for rainfall forecasting in Bangkok, Thailand. **Hydrology and Earth Systems Sciences**, v. 13, p. 1413–1425, 2009.

INGSRISAWANG, L.; INGSRISAWANG, S.; SOMCHIT, S.; AUNGSURATANA, P.; KHANTIYANAN, W. Machine learning techniques for short-term rain forecasting system in the northeastern part of Thailand. World Academy of Science, Engineering and Technology, International **Journal of Computer,**

Electrical, Automation, Control and Information Engineering, v. 2, n. 5, p. 1422–1427, 2008.

INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS - CENTRO DE PREVISÃO DE TEMPO E ESTUDOS CLIMÁTICOS (INPE.CPTEC). **Explicativo Ensemble**: probabilidades de ocorrência de temperatura à superfície baseada em previsões por conjunto (ensemble). Cachoeira Paulista, 2016. Available online at: http://agricultura.cptec.inpe.br/exp_ensemble.shtml, Accessed in Apr. 2016.

JAIN, A. K.; MAO, J. **Artificial Neural Network: A tutorial**. Special Report, IEEE, Computer - Special issue: neural computing, v. 29, Issue 3, p. 31-44, 1996.

JIAWEI H.; KAMBER, M.; PEI J. **Data mining: concepts and techniques**, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 3rd ed., 740 pp., 2012.

KALNAY, E. **Atmospheric modeling, data assimilation and predictability**. Cambridge University Press, 341 pp., 2003.

KALTEH, A.M. Monthly river flow forecasting using artificial neural network and support vector regression models coupled with wavelet transform. **Computers & Geosciences**, V. 54, p. 1–8, 2013.

KISI, O. River flow modeling using artificial neural networks. **Journal of Hydrology Engineering**, ASCE, v. 9, issue 1, p. 60-63, 2004.

LEITH C. E. Theoretical skill of Monte Carlo forecasts. **Monthly Weather Review**, v. 102, p. 409-418, 1974.

LEWIS, J. M. Roots of Ensemble Forecasting. **Monthly Weather Review**, v. 133:7, p. 1865-1885, 2005.

LORENZ, E. N. Deterministic nonperiodic flow. **Journal of the Atmospheric Sciences**, v. 20, p. 130-141, 1963.

_____. **A study of the predictability of a 28-variable atmospheric model**. Tellus, 17, p. 321–333, 1965a.

_____. **On the possible reasons for long-period fluctuations of the general circulation**. WMO-IUGG Symposium on Research and Development Aspects of Long-Range Forecasting, Boulder, CO. Tech. Note 66, 345 pp., 1965b.

_____. **The predictability of a flow that possesses many scales of motion**. Tellus, v. 21, p. 289–307, 1969.

LYNCH, P. The origins of computer weather prediction and climate modeling. **Journal of Computational Physics**, v. 227:7, p. 3431-3444, 2007.

MCCULLOCH, W. S.; PITTS, W. **A logical Calculus of Ideas Immanent in Nervous Activity**, Bulletin of Mathematical Biophysics, v. 5, p. 115-133, 1943.

MENDONÇA, A. M.; BONATTI, J.P. Experiments with EOF-Based perturbation Methods and their impact on the CPTEC/INPE Ensemble Prediction System. **Monthly Weather Review**, v. 137, n. 4, p. 1438-1459, 2009.

MINSKY, M.; PAPERT, S. **Perceptrons: An Introduction to Computational Geometry**. M.I.T. Press, Cambridge, Mass., 258 pp., 1969.

MOUSTRIS, K. P.; LARISSI, I. K.; NASTOS, P. T. Precipitation Forecast Using Artificial Neural Networks in Specific Regions of Greece, **Water Resources Management**, v. 25, Issue 8, p. 1979-1993, 2011.

MURPHY J. M. The impact of ensemble forecasts on predictability, Quarterly **Journal of the Royal Meteorological Society**, v. 114, p. 463-493, 1988.

NAIR, V.; HINTON, G. E. Rectified linear units improve restricted Boltzmann machines. In: INTERNATIONAL CONFERENCE ON MACHINE LEARNING, 27., 2010, **Proceedings**...

NAVONE, H. D.; CECCATTO, H. A. Predicting Indian monsoon rainfall: a neural network approach. **Climate Dynamics**, v. 10, p. 305–312, 1994.

NICULESCU, S.P. Artificial neural networks and genetic algorithms in QSAR. **Journal of Molecular Structure**, v. 622:1-2, p. 71–83, 2003.

NACIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION (NOAA). **National Weather Service Glossary**. 2016. Available online at http://w1.weather.gov/glossary/. Accessed in Apr 2016.

PEÑA, M.; VAN DOOL, H. Consolidation of Multimodel Forecasts by Ridge Regression: Application to Pacific Sea Surface Temperature. American Meteorological Society, **Journal of Climate**, v. 21, p. 6521-6538, 2008.

PETROLIAGIS T. I.; PINSON P. **Early indication of extreme winds utilizing the Extreme Forecast Index**. ECMWF Meteorology Newsletter, n. 132, p. 13-19, 2013.

R CORE TEAM. **R: A language and environment for statistical computing**. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/, 2015.

RADZUAN, N. F. M.; PUTRA, A.; OTHMAN, Z.; BAKAR, A. A.; HAMDAN, A. R. Three Artificial Intelligence Techniques for Rain Domain in Precipitation

Forecast: A Comparative Study, World Academy of Science, Engineering and Technology, **International Journal of Environmental, Chemical, Ecological, Geological and Geophysical Engineering**, v. 7, n. 12, 2013.

RAFTERY, A. E.; GNEITING, T.; BALABDAOUI, F.; POLAKOWSKI, M. Using Bayesian model averaging to calibrate forecast ensembles. **Monthly Weather Review**, v. 133, p. 1155–1174, 2005.

RAMIREZ M. C. V.; VELHO H. F. C.; FERREIRA N. J. Artificial neural network technique for rainfall forecasting applied to the São Paulo region, **Journal of Hydrology**, v. 301, n. 1-4, p. 146-162, 2005.

REYNOLDS, C. A.; WEBSTER, P. J.; KALNAY, E. Random error growth in NMC's global forecasts. **Monthly Weather Review**, v. 122, p. 1281-1305, 1994.

RIAD, S.; MANIA, J.; BOUCHAOU, L.; NAJJAR, Y. Rainfall-Runoff model using an Artificial Neural Network Approach. **Mathematical and Computer Modeling**, v. 40:7–8), p. 839–846, 2004.

RICHARDSON, L. F. **Weather Prediction by Numerical Process**, Cambridge (The University press), 256p, 1922.

RIEDMILLER, M.; BRAUN, H. **A direct adaptive method for faster backpropagation learning: the RPROP algorithm**. Neural Networks, IEEE International Conference on, San Francisco, CA, v.1, p. 586-591, 1993.

RIPLEY, B.; VENABLES, W. **Package "nnet"**. Feed-Forward Neural Networks and Multinomial Log-Linear Models, v7.3-12. 2016. Available online at https://cran.r-project.org/web/packages/nnet/, Accessed in Apr 2016.

ROSENBLATT, F. The perceptron: a probabilistic model for information storage and organization in the brain. **Psychological Review**, v. 65, p. 386–408, 1958.

ROZANTE, J. R.; MOREIRA, D. S., GONÇALVES, L. G. G., VILA, D A. Combining TRMM and Surface Observations of Precipitation: Technique and Validation over South America, **Weather and Forecasting**, AMS, v. 25, p. 885-894, 2010.

RUIZ, J. J.; SAULO, C. How sensitive are probabilistic precipitation forecasts to the choice of calibration algorithms and the ensemble generation method? Part I: sensitivity to calibration methods. **Meteorological Applications**, v. 19, p. 302–313, 2012.

RUIZ, J.; SAULO, C.; KALNAY, E. Comparison of methods to generate probabilistic quantitative precipitation forecasts over South America. **Weather and Forecasting**, v. 24, p. 319–336, 2009.

RUMELHART, D. E., MCCLELLAND, J. L. **Parallel Distributed Processing: Exploration in the Microstructure of Cognition**, MIT Press, Cambridge, Mass., 1986.

SAMUEL, A. L. Some studies in machine learning using the game of checkers. **IBM Journal of Research and Development**, v. 44, n. 1.2, p. 206-226, 1959.

SARLE, W.S. **Neural Network FAQ**, part 1 of 7: Introduction, periodic posting to the Usenet newsgroup comp.ai.neural-nets, Available online at ftp://ftp.sas.com/pub/neural/FAQ.html, Accessed in Apr 2016, ed. 2002.

SAROHA, S.; AGGARWAL, S. K. Multi Step Ahead Forecasting of Wind Power by Different Class of Neural Networks. RAECS UIET Panjab University Chandigarh, **Proceedings**…, 2014.

SHARMA, S. K.; CHANDRA, P. An Adaptive Slope Sigmoidal Function Cascading Neural Networks Algorithm. In: International Conference on Emerging Trends in Engineering and Technology (ICETET), 3., 2010, Goa, **Proceedings**…, p. 531-536.

SIMMONS, A. J.; MUREAU, R.; PETROLIAGIS, T. Error growth and estimates of predictability from the ECMWF forecasting system. **Quarterly Journal of the Royal Meteorological Society**, v. 121, p. 1739–1771, 1995.

SLOUGHTER, J. M.; RAFTERY, A.; GNEITING, T.; FRALEY, C. Probabilistic quantitative precipitation forecasting using Bayesian model averaging. **Monthly Weather Review**, v. 135, p. 3209–3220, 2007.

STENSRUD, D.; YUSSOUF, N. Reliable probabilistic quantitative precipitation forecasts from a short-range ensemble forecasting system. **Weather and Forecasting**, v. 22, p. 3–17, 2007.

_____. Short-range ensemble predictions of 2-m temperature and dew point temperature over New England. **Monthly Weather Review**, v. 131, p. 2510–2524, 2003.

TAYLOR J. W.; BUIZZA R. Neural Network Load Forecasting with Weather Ensemble Predictions. **IEEE Trans. on Power Systems**, v. 17, p. 626-632, 2002.

THE WORLD METEOROLOGICAL ORGANIZATION (WMO). **A guide to the code form FM 92-IX EXT. GRIB**. Available online at http://www.wmo.int/pages/prog/www/WDM/Guides/ Guide-binary-2.html, Accessed in Apr 2016.

_____. **Recommendations for the verification and intercomparison of QPFS and PQPFS from operational NWP models**, World Weather Research Programme, WWRP 2009 – 1, Rev. 2, 2008.

UNIVERSITY CORPORATION FOR ATMOSPHERIC RESEARCH (UCAR) - COMET/METED PROGRAM. **Model Fundamentals - Version 2**. 2009. Available online at http://www.meted.ucar.edu/nwp/model_fundamentals, Accessed in Apr 2016.

_____. **Ensemble Forecasting Explained**. 2009. Available online at http://www.meted.ucar.edu/nwp/pcu1/ensemble/, Accessed in Apr 2016.

VEENHUIS, B. A. Spread Calibration of Ensemble MOS Forecasts. **Monthly Weather Review**, v. 141, p. 2467–2482, 2013.

VILLELA, A. P. **Tratado da Bacia do Prata** – Suplemento 149, R. Inf. legisl. Brasília a. 21 n. 81 Jan./Mar, 1984.

WEI M.; TOTH Z.; WOBUS R.; ZHU Y. **Initial perturbations based on the ensemble transform (ET) technique in the NCEP global operational forecast system**. Tellus, v. 60A, p. 62–79, 2008.

WESSEL A.; JIANG J.; DOBSCHINSKI J.; LANGE B. **Improving short-term forecast with online wind measurements**. European Wind Energy Conference and Exhibition (EWEC), 2009.

WILKS D. S. **Statistical methods in the atmospheric sciences**. v. 100, 3rd ed., International Geophysics Series, Academic Press, 2011.

WILKS, D. S. Extending logistic regression to provide full probability distribution MOS forecasts. **Meteorological Applications**, v. 16, p. 361–368, 2009.

WU, C. L.; CHAU, K. W.; FAN C. Prediction of rainfall time series using modular artificial neural networks coupled with data-preprocessing techniques. **Journal of Hydrology**, v. 389, p. 146–167, 2010.

YILMAZ, I.; KAYNAR, O. Multiple regression: ANN (RBF, MLP) and ANFIS models for prediction of swell potential of clayey soils. **Expert Systems with Applications**, Elsevier, v. 38, p. 5958-5966, 2011.

YUAN H.; GAO, X.; MULLEN, S. L.; SOROOSHIAN, S.; DU, J.; JUANG, H-M. H. Calibration of probabilistic quantitative precipitation forecasts with an artificial neural network. **Weather and Forecasting**, v. 22, p. 1287–1303, 2007

ZHANG, Z.; KRISHNAMURTI, T.N. A perturbation method for hurricane ensemble predictions. **Monthly Weather Review**, v. 127, p. 447-469, 1999.

ZUNRNDORFER, E. A. **Verification of MOS, PE, and LFM Quantitative Precipitation Forecasts for the 1979-80 Cool Season**, NOAA, TDL office Note 80-9, 1980.

APPENDIX A – MULTI-INFO CHARTS OF ALL LOCATIONS

Figure A.1   Multi-info chart: correlation of variables and OBS at Assuncion (PAR).
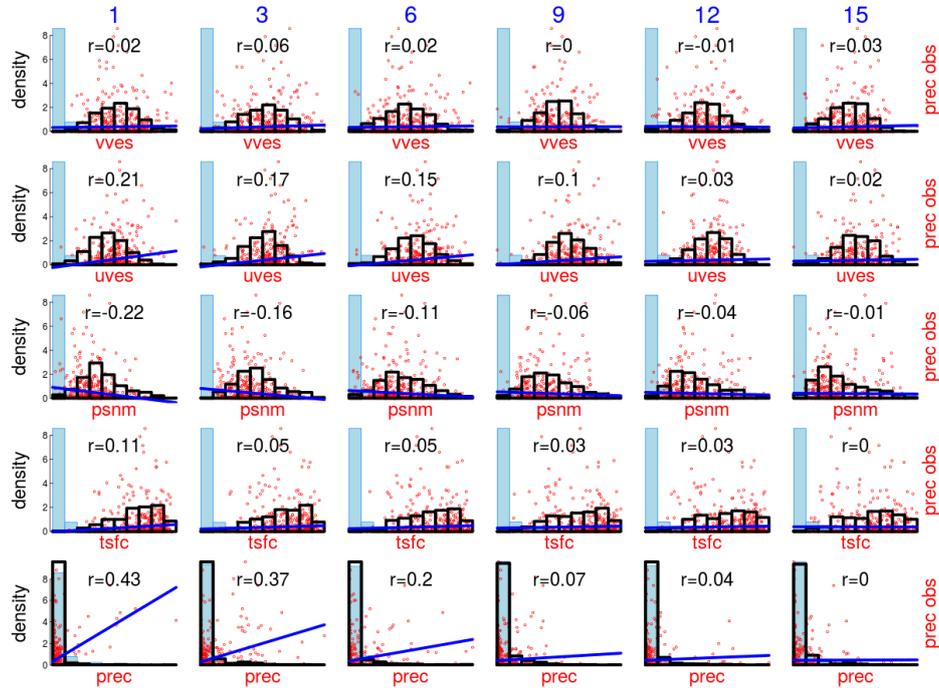

Figure A.2   Multi-info chart: correlation of variables and OBS at B. Aires (ARG).
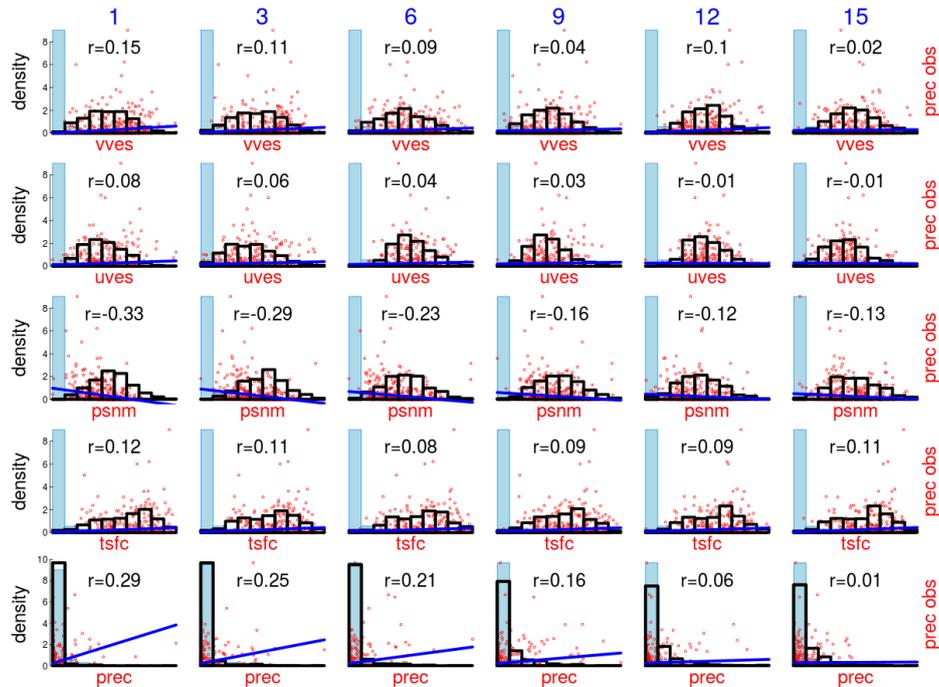
Figure A.3   Multi-info chart: correlation of variables and OBS at Bagé (RS).
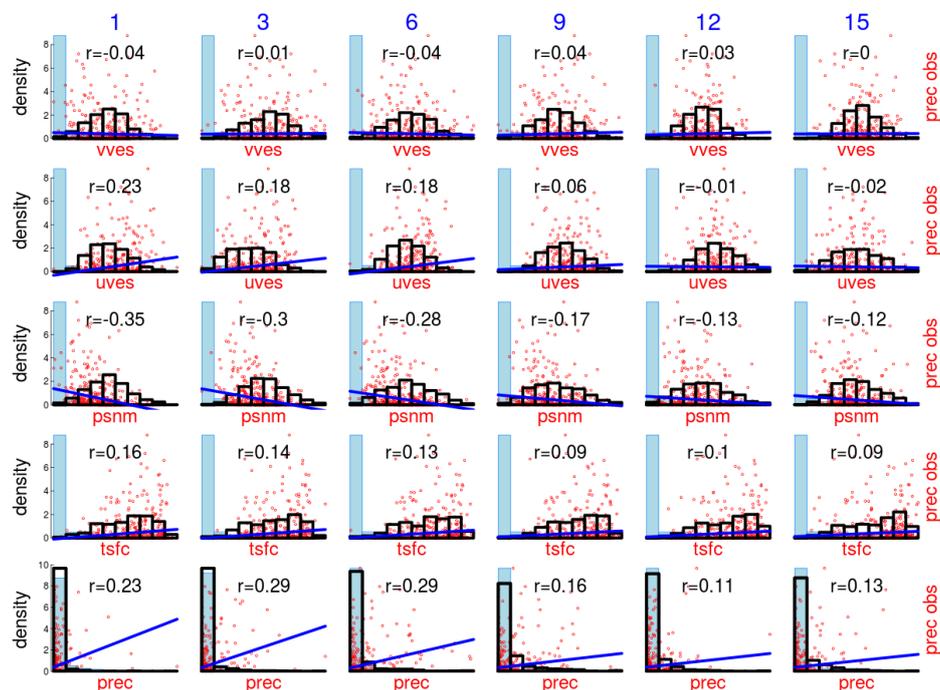

Figure A.4   Multi-info chart: correlation of variables and OBS at Brasília (DF).

116
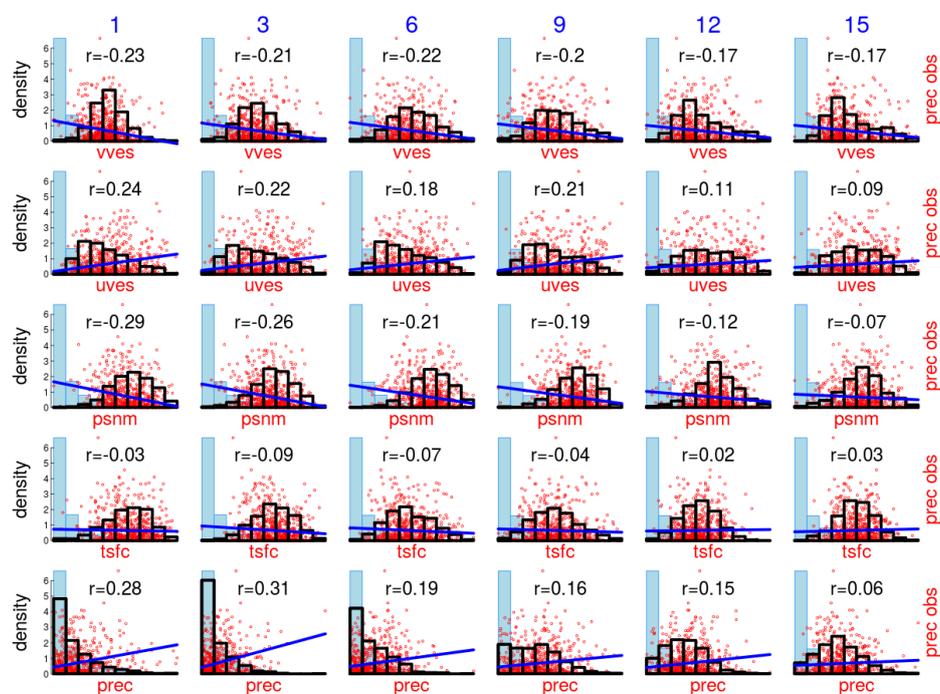
Figure A.5   Multi-info chart: correlation of variables and OBS at Cuiabá (MT).
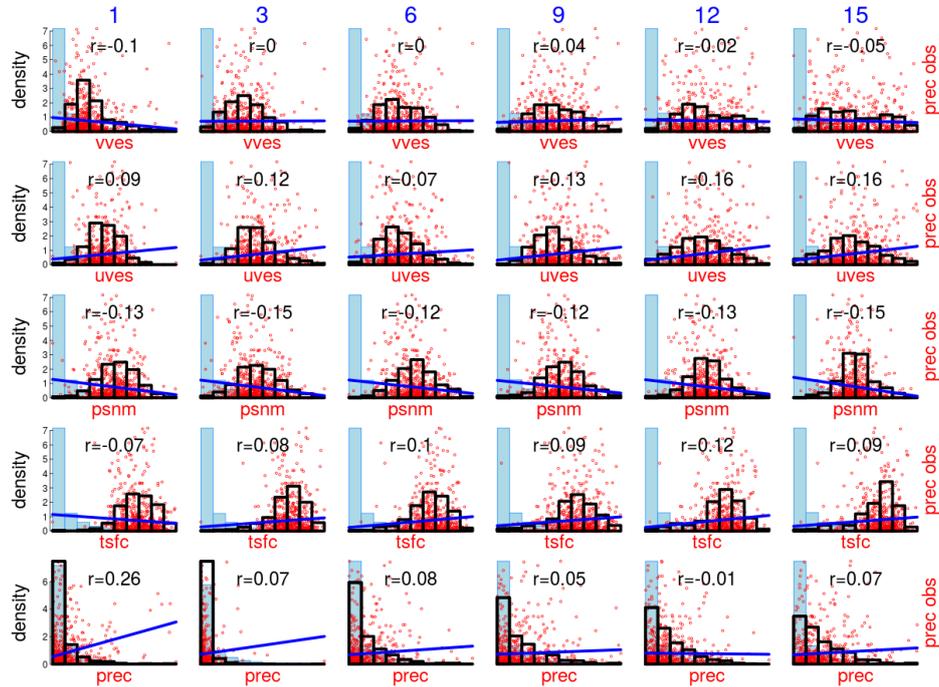
Figure A.6   Multi-info chart: correlation of variables and OBS at C. Grande (MS).
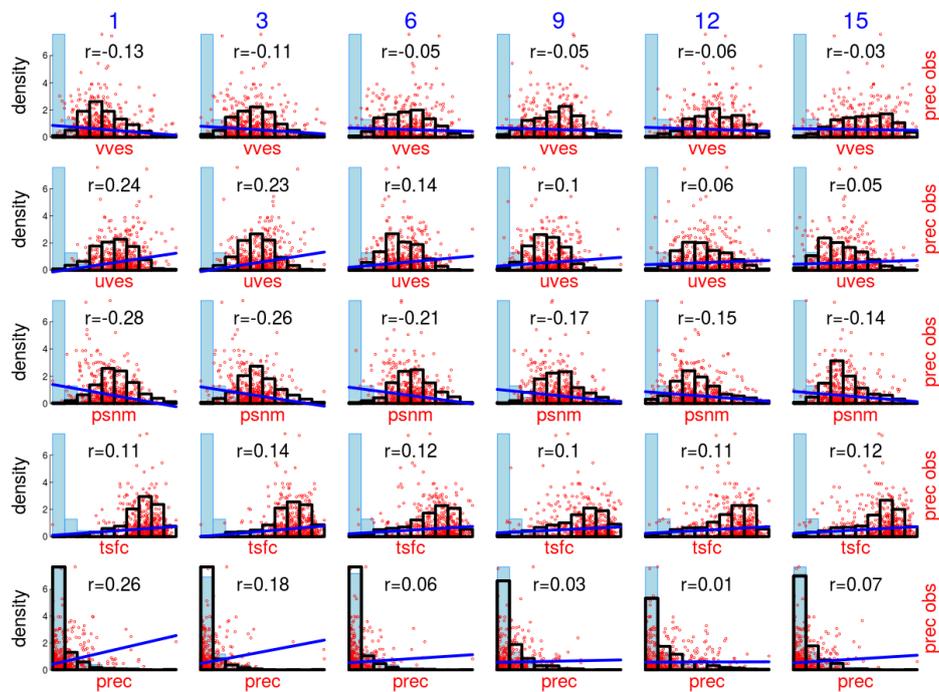
Figure A.7   Multi-info chart: correlation of variables and OBS at Curitiba (PR).
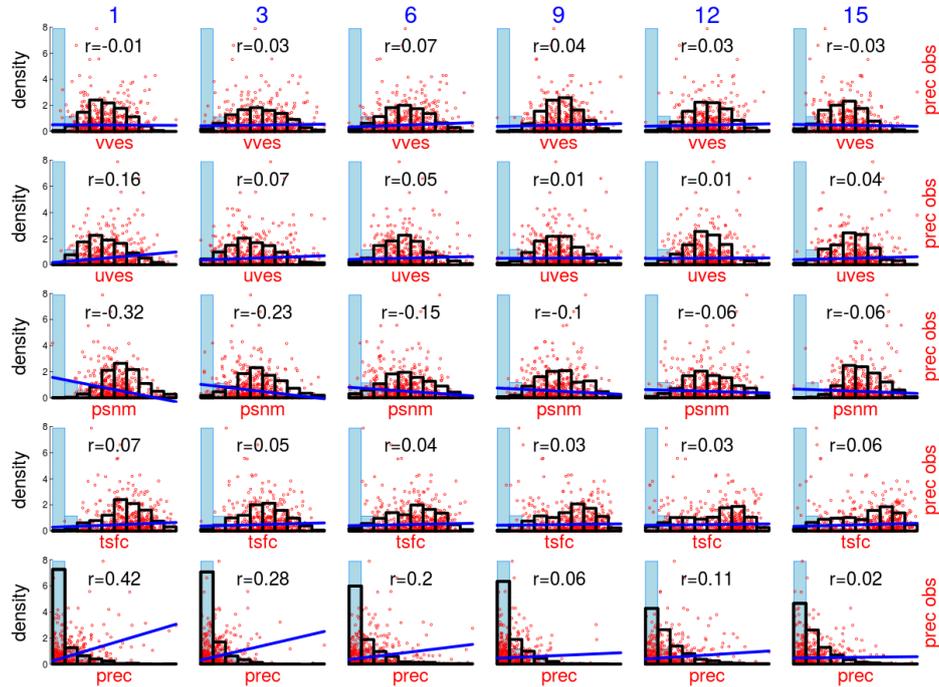
Figure A.8   Multi-info chart: correlation of variables and OBS at Caxambu (MG).
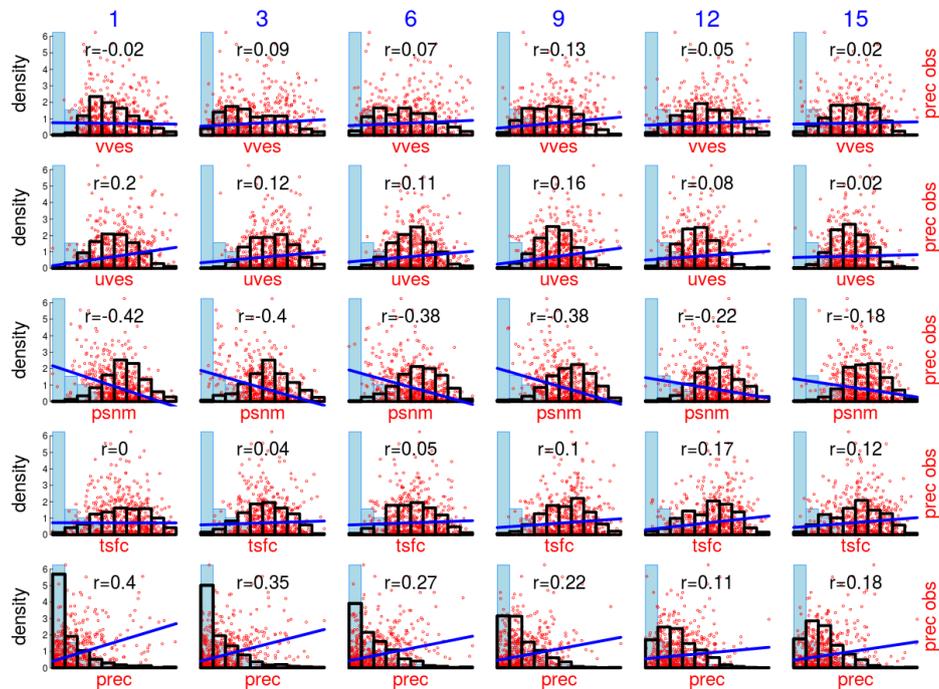
Figure A.9   Multi-info chart: correlation of variables and OBS at Montevideo (URU).
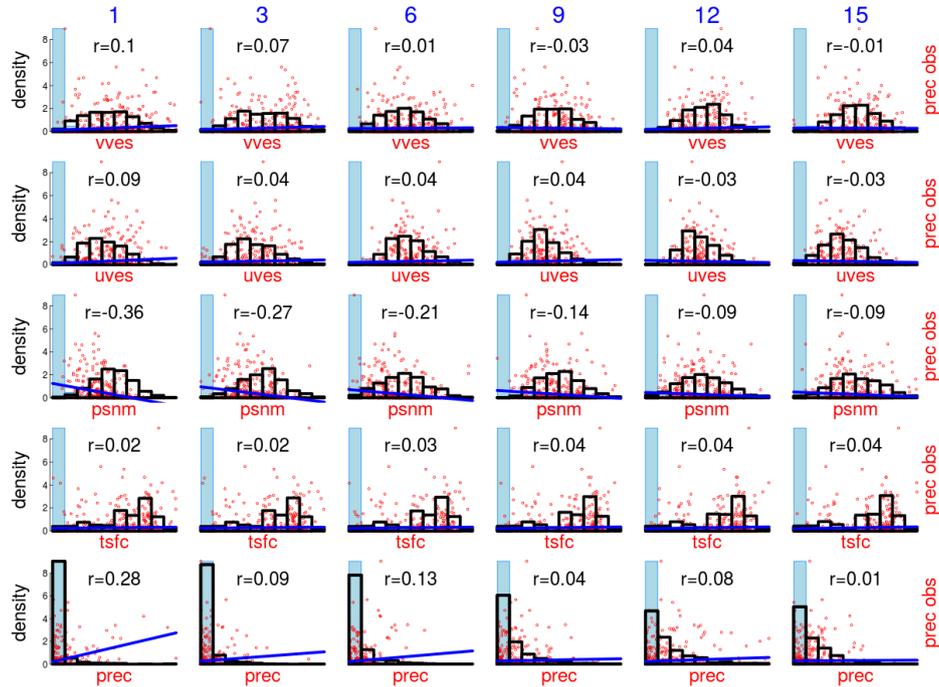

Figure A.10   Multi-info chart: correlation of variables and OBS at Potosi (BOL).
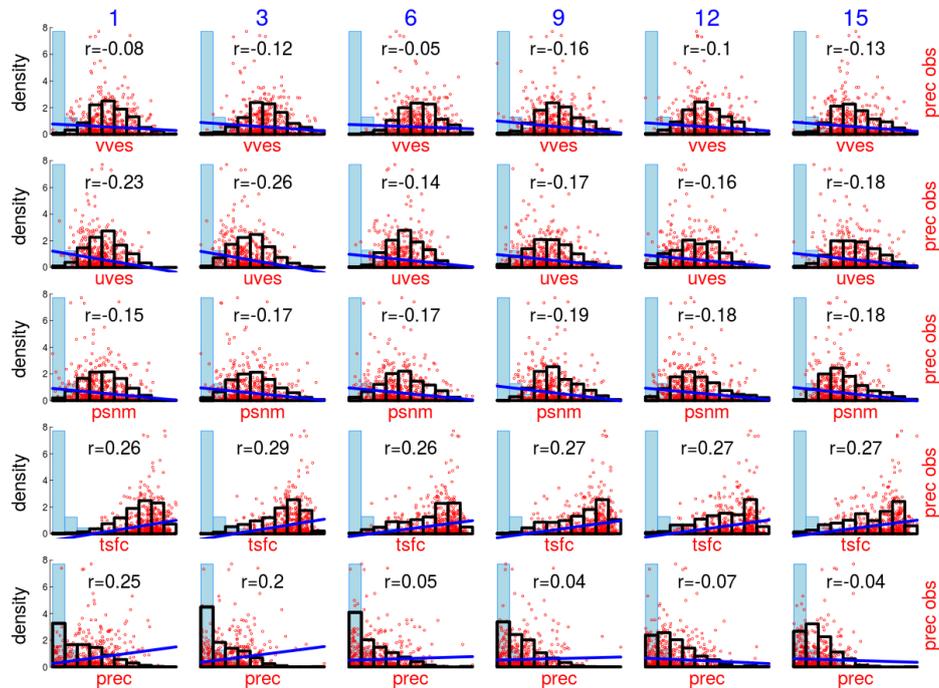
Figure A.11   Multi-info chart: correlation of variables and OBS at São Paulo (SP).
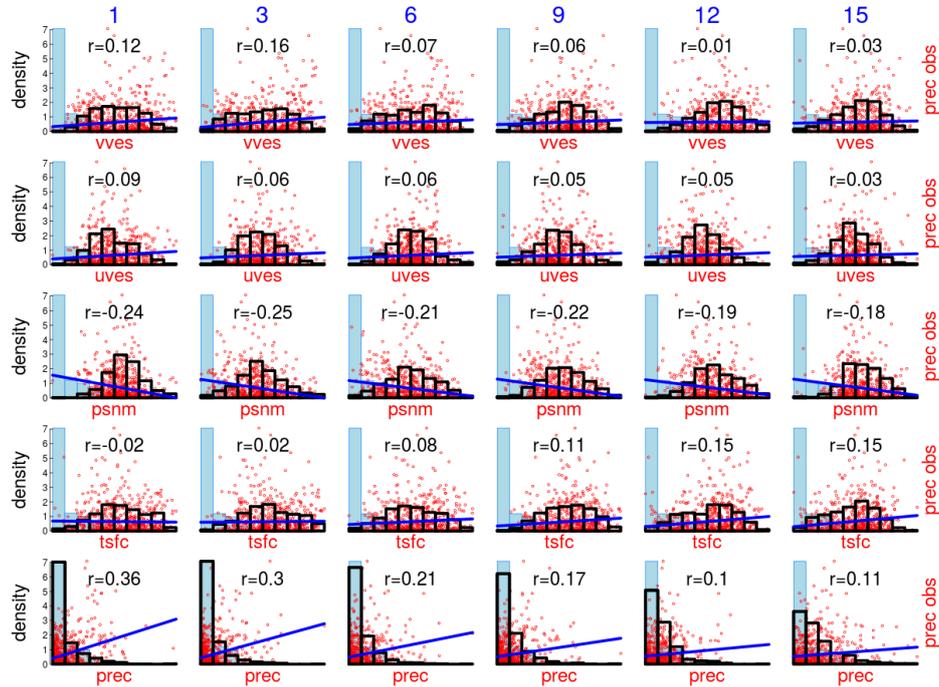

Figure A.12   Multi-info chart: correlation of variables and OBS at Salta (ARG).