Alma Mater Studiorum · Università di Bologna

Scuola di Scienze
Corso di Laurea Magistrale in Fisica del Sistema Terra

Application of Neural Networks in Atmospheric Rivers forecasting

Relatore:

Prof. Nadia Pinardi

Correlatori:

Dott. Arthur J. Miller
Dott. Luca Delle Monache
Dott. William Chapman

Presentata da:

Giacomo Lucidi

Sessione III
Anno Accademico 2019/2020

Abstract

The goal of the study is to investigate the skills of two Convolutional Neural Networks as postprocessing frameworks for improving the prediction of the vertically Integrated Horizontal Water Vapor Transport (IVT) field provided by the 30-years high resolution reforecast based on WRF and developed by the Centre for Western Weather and Water Extremes (CW3E) of UCSD San Diego. IVT is the main feature used to detect the Atmospheric Rivers (ARs). ARs are large scale filamentary regions associated with the warm sector of extratropical cyclones observable at midlatitudes on both the hemispheres. ARs are characterized by the poleward transport of large amount of water vapor in the low troposphere and they are responsible for extreme precipitation events, particularly on the North American West Coast, playing an important role for the water supply of the regions where they occur.

Numerical Weather Prediction (NWP) models, based on the numerical integration of a discretized version of the Navier-Stokes equations, show large forecasting errors in the landfalling of ARs. In particular, the ensemble forecasts provided by five main forecasting centres were analysed. The models resulted able to predict the presence of ARs, however, they failed in the timing and position of the landfall.

Starting from a study conducted at CW3E, we used a Machine Learning approach to improve the forecasts of ARs and to investigate the potential utility of data-driven models in weather forecasting. Two Convolutional Neural Networks were trained with 30-years reforecast IVT data based on

WRF over a region of interest from 25°N to 60°N latitude, and 150°W to 115°W longitude. After a large number of experiments to optimize the networks, the networks output was tested against MERRA2 reanalysis, used as ground-truth, in terms of root mean squared error, bias and correlation.

Italian version

L'obiettivo di questo studio è quello di analizzare le prestazioni di due Convolutional Neural Networks utilizzate come strumento di postprocessing per migliorare le previsoni del Trasporto Orizzontale di Vapore Acqueo verticalmente Integrato (IVT) ricavato dal reforecast trentennale messo a disposizione dal Centre for Western Weather and Water Extremes (CW3E) dell'Università della California di San Diego. L'IVT è la principale variabile utilizzata per le previsioni di Atmospheric Rivers (ARs). Gli ARs sono regioni a grande scala associati al settore caldo dei cicloni extratropicali e sono osservabili alle medie latitudini in entrambi gli emisferi. Gli Atmospheric Rivers sono caratterizzati da un importante trasporto di vapore acqueo nella bassa troposfera dalle medie latitudini verso i poli e sono responsabili di eventi di precipitazione molto intensa, in particolare sulla parte settentrionale della costa ovest Americana. Gli ARs rappresentano inoltre un'importante risorsa di acqua per le regioni lungo le quali si manifestano.

I modelli numerici di previsione basati sulla risoluzione delle equazioni di Navier-Stokes mostrano un errore significativo nel predire la precipitazione dovuta agli ARs. In particolare, sono state analizzate le previsioni ensemble di cinque principali centri di previsione che risultano attendibili nel predire la presenza degli ARs ma falliscono nella previsione della zona e del momento dell'evento di precipitazione.

Partendo da uno studio effettuato presso il CW3E, in questa tesi viene utilizzato il Machine Learning per migliorare le previsioni degli Atmospheric Rivers. Sono state utilizzate due Convolutional Neural Networks il cui training è stato effettuato con 30 anni di dati ottenuti dal reforecast di CW3E su una regione di interesse da 25°N a 60°N latitudine, e 150°W a 115°W longi-

tudine. Dopo un esaustivo numero di esperimenti per ottimizzare le due reti neurali i risultati ottenuti sono stati comparati con le reanalisi MERRA2 in termini di scarto quadratico medio, bias e correlazione.

Contents

List of Figures

Chapter 1

Introduction

1.1 What are atmospheric rivers

Atmospheric rivers (ARs) are narrows ($\leq$ 1000 km) and long ($\geq$ 2000 km) region of high water vapor content (Chapman et al., 2019 [50]). These corridors are often identified inside the warm conveyor belt (WCB), a warm and humid region located along and ahead the cold front of an extratropical cyclone and characterized by high water vapor content and strong low level winds in the low troposphere (Ralph et al., 2004 [37]; Wick et al., 2013 [10]; Rutz et al., 2014 [17]).

We can find ARs at midlatitudes on both the hemispheres (see figure 1.2) and it's estimated that these narrow corridors cover 10% of the hemispheric circumference (Ralph et al., 2006 [39]; Zhu et al., 1998 [53]). Associated with synoptic-scale extratropical cyclones, ARs are more numerous during the winter time (Gimeno et al., 2014 [25]; Waliser et al., 2012 [49]) and they play an important role in the heat poleward transport. It is shown that the 90% and more of water vapor travels across the midlatitudes meridionally through these filamentary regions (Zhu et al., 1998 [53]; Gimeno et al., 2014 [25]). ARs take their name by the water volume transported and their flow rate that is similar to the world's largest rivers (Zhu et al., 1998 [53]). A typical AR could carry as much water as $\sim$ 10 Mississipi River and there are three to five

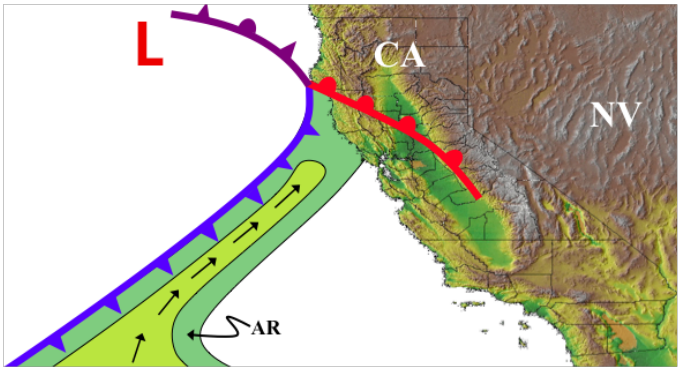


Figure 1.1: Schematic illustration of an atmospheric river (AR) in the warm sector of an occluded extratropical cyclone. Arrows represent the water vapor transport (Moore et al., 2011 [33]).

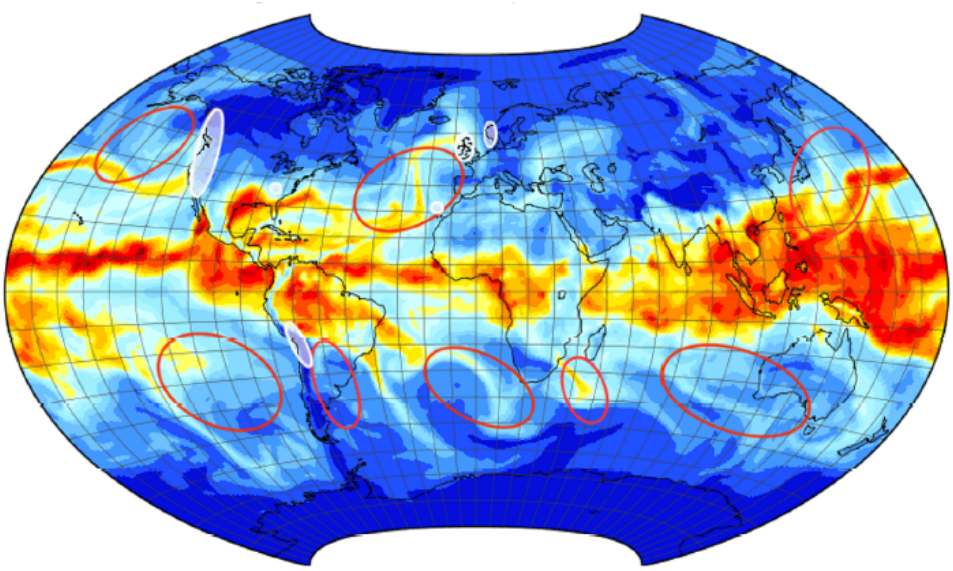


Figure 1.2: General ARs global distribution (red contours) and main continental areas (white contours) where extreme rain events caused by ARs have been reported (Gimeno et al., 2014 [25]).

ARs in each hemisphere at any given time (Ralph and Dettinger, 2011 [9]). There is some disagreement on the term AR, with alternatives like “tropical moisture exports ”or “moisture conveyor belt ”(Knippertz and Wernli, 2010 [19]; Bao et al., 2006 [3]), the last term because of their connection with the warm sector of extratropical cyclones.

1.1.1 Atmospheric rivers as a source of waters and flooding

ARs carry a big amount of water vapor in the lower troposphere and combining it with horizontal wind and moist neutrality they can be an important source of precipitation (Ralph et al., 2006 [39]). The main mechanism of precipitation is the orographic lifting so when an AR meet a topographic barrier (Wick et al., 2013 [10]) the AR is forced upward producing in some cases extreme precipitation events and flooding, decreasing the content of water vapor as the AR penetrates inland (Wick et al., 2013 [10]). ARs contribute 60-100% of the most extreme North American West Coast (NAWC) hydrometeorological events (Lamjiri et al., 2017 [21]; Gershunov et al., 2017 [11]). At the same time ARs play an important role in the regional water cycle as a significant seasonal water supply. With a focus on the western United States, it’s estimated that ARs contribute 30-65% of annual precipitation on the U.S. West Coast (Lamjiri et al., 2017 [21]; Gershunov et al., 2017 [11]), particularly in California the six or more ARs that produce landfalling have contributed 35-50% of all the state’s precipitation, with a single AR storm providing 2.5 to 5 cubic kilometres of precipitation (Ralph and Dettinger, 2011 [9]).

1.2 Characteristics of Atmospheric Rivers

An Atmospheric River is typically observed within the WCB in front of the cold region of an extratropical cyclone and from this disposition three

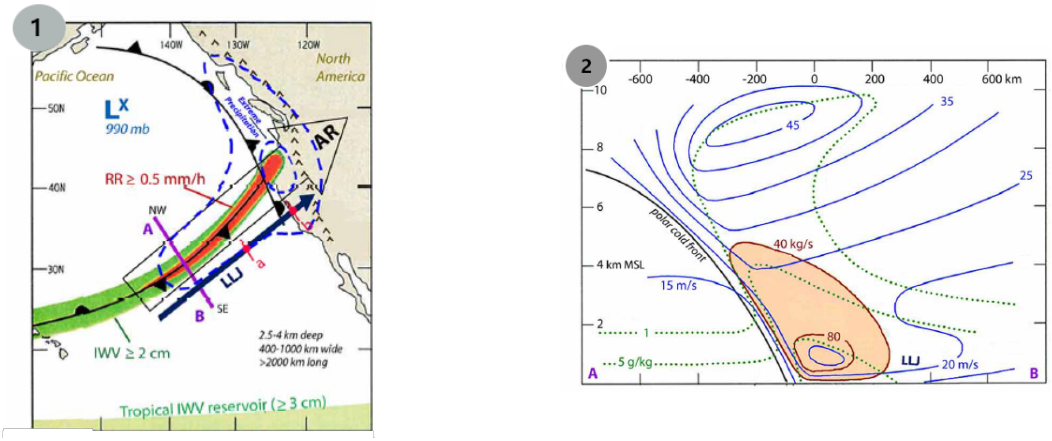


Figure 1.3: In figure (1) it showed the plan structure of an AR with the cold front and the low level wind defined as Low Level Jet (LLJ). In red it showed the rain rate enhancement ($RR \geq 0.5mmh^{-1}$) and in green is represented the thickness of the (IWV) if it were condensed. The line A-B (magenta) represents the potential positioning of the cross section of figure (2) and the pink point a-b defines a potential aerea for the integration of the profiles of wind speed and humidity flux. The broken line in blue identifyes a potential precipitation region based on the AR's position. The picture is obtained from (Gimeno et al., 2014 [25]) and the values showed represent the regional trend. In figure (2) is represented the cross section of the AR along A-B in figure (1). In the cross section in green is defined the specific humidity, in red the horizontal moisture flux and in blue the wind speed (Gimeno et al., 2014 [25]; Ralph et al., 2004 [37]).

main properties are identified: 1) Enhanced low-level specific humidity with a vertical expansion due to the frontal convergence that constraints air to ascend. 2) Presence of a strong wind in the first two kilometres of atmosphere in front of the cold frontal region originated by the temperature gradient across the cold front. 3) Low static stability (Ralph et al., 2004, 2005 [37], [38]; Wick et al., 2013 [10]). Furthermore two processes are identified as origin of the water vapor inside an AR: 1) Convergence of the humidity along the cold front of an extratropical cyclone. 2) Poleward flux of tropical moisture (Gimeno et al., 2014 [25]; Bao et al., 2006 [3]; Ralph et al., 2004 [37]).

1.3 Artificial Neural networks and weather forecasts: state of art

Weather forecasts are based on Numerical Weather Prediction (NWP) models that solving a numerical integration of a discretized version of the Navier-Stokes equations provide the evolution of the atmospheric state over the time (Chapman et al., 2019 [50]; Dueben and Bauer, 2018 [36]; Scher, 2018 [43]). NWP models still represent an essential tool for predicting the weather and the Earth's climate (Scher, 2018 [43]). However, NWP models forecast skills decrease with model time integration due to uncertainty in initial conditions, numerical approximations and model deficiencies (Chapman et al., 2019 [50]).

1.3.1 Statistics and Machine Learning for improving weather and climatological forecast: toward Artificial Neural Networks

Statistical forecast postprocessing techniques have been used to improve the forecast skills of NWP models. These algorithms correct the errors in the current predictions using historical forecasts and observations (Chapman et al., 2019 [50]). Algorithms developed for this purpose include: model out-

put statistics approaches (Glahn and Lowry, 1972 [12]; Carter et al., 1989 [5]; Wilks and Hamill, 2007 [51]), running mean techniques (Stensrud and Skindlov, 2002 [45]; Stensrud and Yussouf, 2003 [46]; Hacker and Rife, 2008 [13]), algorithms based on Kalman Filtering (Homleid, 1995 [15]; Roeger et al., 2003 [42]; Delle Monache et al., 2006 [6]; McCollor and Stull, 2008 [27]), and analog-based methods which draw from past events to match designed features of the current forecast to correct the error (Delle Monache et al., 2011 [7]). Recently, Machine Learning has been used to improve NWP models too. Machine Learning techniques are computer algorithms that improve automatically through experience (Mitchell, 1997 [31]), so they provide the possibility to build models based on sample data, known as "training data", capable of making predictions or taking decisions without explicitly knowing any type of connection between data and the physical process they are used to predict or taking decision over (Koza et al., 1996 [20]). They are also defined as *data-driven* models and some of them have been proposed in connection with NWP models. They include: studying orbital parameters and climate fields from a climate model (Holden et al., 2015 [14]), learning from high-resolution simulations in order to improve predictions (Anderson and Lucas, 2018 [2]; Rasp et al., 2018 [40]), helping in decision making in extreme weather situations (McGovern et al., 2017 [28]), detecting extreme weather in data set (Liu et al., 2016 [24]), and predicting the uncertainty of weather forecasts (Scher and Messori, 2018 [44]). The mentioned algorithms aim to extract and process information from models and add it into others, this means that NWP models involve Machine Learning techniques; they are linked. In recent years, however, a new challenge has been proposed from the scientific community: can data-driven models compete with NWP models? Broadly: can an algorithm trained with atmospheric data replace a model based on physical equations?

1.3.2 Artificial Neural Networks: a short introduction

Scientists are trying to investigate if data-driven models can beat NWP models with Artificial Neural Networks (ANN), usually simply called Neural Networks (NNs). NNs are a class of Machine Learning techniques that are used in applications such computer vision, speech recognition and data filtering, representing an important features of Artificial Intelligence (Dueben and Bauer, 2018 [36]). Their structure is inspired to the human brain, they are composed from artificial neurons organized in layers: input, hidden and output. Input data are weighted through hidden layers and returned as output.

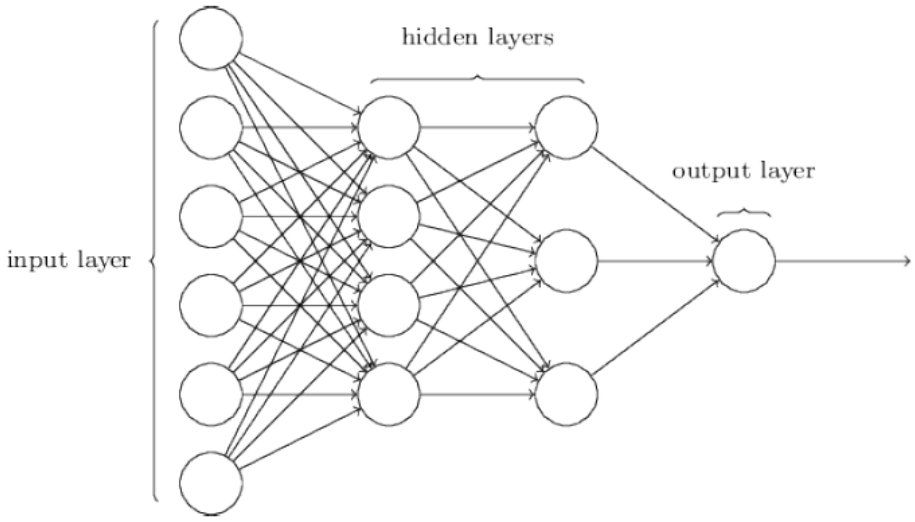


Figure 1.4: Basic structure of a Nerual Network layers (Nielsen, 2015 [29]).

The peculiarity of these algorithms is that they are capable to improve using their output information in relation to a desired outcome through a learning process. Two main types of learning process are defined: 1) Supervised learning: providing the NN with input and matching output patterns; 2) Unsupervised learning: the NN is supposed to discover statistically salient features of the input population without any predefined output patterns (Dongare et al., 2012 [1]). NNs exploit large amount of data to predict

the evolution of non-linear processes (Dueben and Bauer, 2018 [36]). Training (with training we define the action of the NN to read the input data on whose basis it will make the prediction) NNs with atmospheric data allow us to compute the evolution in time of the atmospheric states (Scher, 2018 [43]) emulating the physics and the dynamics of NWP models without using any physical equations. (Dueben and Bauer, 2018 [36]).

1.3.3 Artificial Neural Networks compared with NWP models in literature

Among the numerous works that have been done trying to test if a NN can compete with a NWP model we report the conclusions reached of four of them.

Dueben and Bauer, (2018) [36] developed a NN "toy model"with the purpose of predicting the geopotential height 500hPa (Z500) globally, based on 6° longitude - latitude grid. The goal of the study was to investigate trough the "toy model", challenges and possible design choices for NNs in order to obtain significant results on weather prediction. The NN has been compared with the Lorenz '95 low-complexity model that was presented in (Thornes et al., 2017 [48]). The study stated that although NNs work without any knowledge about the Earth system the results obtained from the toy model suggest that a physical knowledge of the Earth system is required considering that NNs and conventional models share the same challenges in terms of nonlinear interactions between model components, interaction among different scales, uncertainty in initial conditions and model biases in long-term simulations. This, the study specified, even with a possible big capacity of training data. Particularly, it would be difficult for NNs to compete with NWP models in climate forecasts and long-term weather forecast, hence seems more reasonable to use NNs for short-term and regional predictions and for the study of large-scale events such El Nino that does not require the prediction of the full atmosphere.

Scher, (2018) [43] presented an experiment never tried before. The

study proposed the comparison between a NN and a general circulation model (GCM) and showed that is possible, in principle, to predict the evolution in time and so the complete dynamics of a weather/climate system with a data-driven model. The results were encouraging, the NN showed stability even in the long-term predictions (almost 1000 years). The experiment obtained positive results respect to Dueben and Bauer, (2018) [36] due to various aspects as the simplicity of the GCM and the training of the networks with the GCM data, contrary to Dueben and Bauer, (2018) [36] that used observation/reanalysis data which are more complex. On the other hand Scher, (2018) [43] used training data with a daily time step while Dueben and Bauer, (2018) [36] used a hourly time step. In the end another possible cause of difference results between the two experiment could be the different architecture of the two NNs. Scher, (2018) [43] proposed as ideas for future works to use models with a more complex dynamics, in particular, to include external forcing and higher amount of data to face an improvement of the resolution and the requirements of more training data by more complex networks to simulate more complex weather/climate system. Scher, (2018) [43] proposed as NN a Convolutional Neural Network (CNN), a type of NN that we'll explain in section 2.5 since it has been used for the experiments of this thesis. Therefore, before go into details we would like to quote Larraondo et al., (2019) [35] since it showed with their results important features about CNNs.

Convolutional Neural Networks and weather interpretation

In a NWP model the state of the atmosphere and its evolution in space and time is defined by the values of different variables or fields and some of these variables are computed by other variables. All the variables are connected by physical equations. Due to a lack of spatial and temporal resolution and to the complexity of some processes these fields need to be interpretated by humans (Larraondo et al., 2019 [35]). Larraondo et al., (2019) [35] introduced the possibility to perform these "interpretations"with CNNs. The

purpose of their study was to predict precipitation events over a period of 5 years on 5 different locations in the European area. For the experiments they used geopotential data as input data relating to three different pressure levels (1000,700,500hPa) and precipitation observation data about the 5 different locations to define if the predicted rain events or dry events (no rain) correspond to reality. As stated in the study the used data aren't directly correlated to the precipitation variable, indeed, the study aimed to prove that CNNs can predict the state of the atmosphere in a way that humans would perform intuitively. Larraondo et al., (2019) [35] showed that CNNs allow to acquire spatial information from images (each state of the atmosphere can be seen as an image) and perform local weather prediction through a 3 dimensional CNN that incorporates as a third axis the temporal component. They showed, at least in principle, the capability of the CNN to exploit the pressure field to predict precipitation events and they suggested CNNs as method to compute variables that in NWP models would be computed with parameterizations or statistical models.

1.4 Thesis objectives

The objective of this thesis is to test and analyse the application of Convolutional Neural Networks as postprocessing tools to improve the forecast of ARs of a Numerical Weather Prediction model. In particular, to improve the performance of the West-WRF high-resolution reforecast provided by the Center For Western Weather and Water Extremes (CW3E) department of UCSD, San Diego. This study started from the work done at CW3E by Chapman et al., (2019) [50] in which a Convolutional Neural Network has been successfully used as postprocessing framework to improve the prediction of the vertically Integrated Horizontal Water Vapor Transport (IVT) field of GFS used to predict and detect the presence of Atmospheric Rivers on the East Pacific and the Northern West Coast of the United States. Two CNNs whose architectures will not be discussed in this study have been entirely

provided by Chapman et al., (2019) [50]. This thesis aims to apply the two networks as postprocessing tools to improve the forecast of the IVT field provided by the West-WRF reforecast and evaluate it against the MERRA-2 reanalyses, used as "ground-truth", in terms of three valuation metrics: root mean squared error, bias and pearson correlation coefficient. The application of CNN to Atmospheric Rivers prediction has been chosen because of the enormous practical importance of the potential improvements. This work has been started and continued under the supervision of the CW3E team at Scripps Institution of Oceanography during a Marco Polo stage.

Chapter 2

Introduction to Artificial Neural Networks

An Artificial Neural Network is a group of artificial neurons that work together. Neurons are organized in layers: input layer, hidden layers and output layer. As shown in figure 1.4, the input layer is the leftmost layer in the network and it's composed by input neurons which represent the entry values in the network. Every input neuron gives his value (we could consider it as a neuron with an output and without an input) to each neuron in the next layer. The output layer is the rightmost layer, in figure 1.4 it is composed by only one neuron. The number of neurons in each layer and the number of hidden layers (between input and output layer) it's chosen in reference to the problem we are approaching to and it's based on some rules. (Nielsen, 2015 [29]).

2.1 Neural Network fundamental unit: artificial neuron

Artificial Neurons (ANs) are the fundamental units of NNs. The task of an AN is to process information called inputs to provide outputs. The simplest model of AN is called *perceptron*. Perceptrons aren't used anymore because

their simplicity isn't suitable for solving complex problems, however they're a good example to understand how an AN works. Perceptrons take binary (0 or 1) inputs and produce binary outputs. Each inputs is weighted with a weight, a real number expressing the importance of the respective input to the output. The output, 0 or 1, is computed defining if the weighted sum $\sum_i w_i x_i$ is less than or greater than a threshold value. (Nielsen, 2015 [29]).

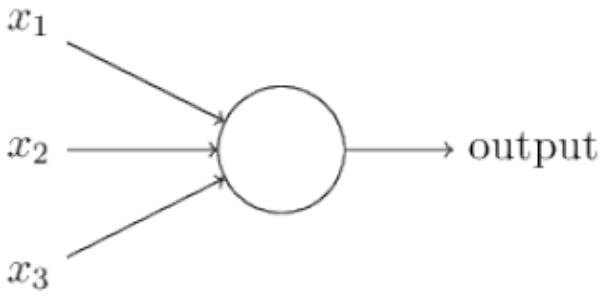


Figure 2.1: Illustration of an Artificial Neuron (Nielsen, 2015 [29]).

In figure 2.1 is illustrated an example of AN with three input values x_1, x_2, x_3, that correspond to three weight w_1, w_2, w_3. In algebraic terms:

$$output = \begin{cases} 0, & \text{if } \sum_i w_i x_i \leq threshold \\ 1, & \text{if } \sum_i w_i x_i > threshold \end{cases}$$

where $\sum_i w_i x_i$ is the dot product $w \cdot x = \sum_i w_i x_i$ where w and x are vectors whose components are the weights and inputs, respectively. For simplify the expressions above we move the threshold to the other side of the inequality and replace it with the so called *bias* (b) where $b \equiv threshold$ that is a paramter of the AN as weights. Therefore the perceptron expressions can be rewritten as:

$$output = \begin{cases} 0, & \text{if } w \cdot x + b \leq 0 \\ 1, & \text{if } w \cdot x + b > 0 \end{cases}$$

Finally we can represent the two possible outputs of a perceptron through the graph of the step function hypothetically represented in figure 2.2 where input and output values are represented respectively on x-axis and y-axis. (Nielsen, 2015 [29]).

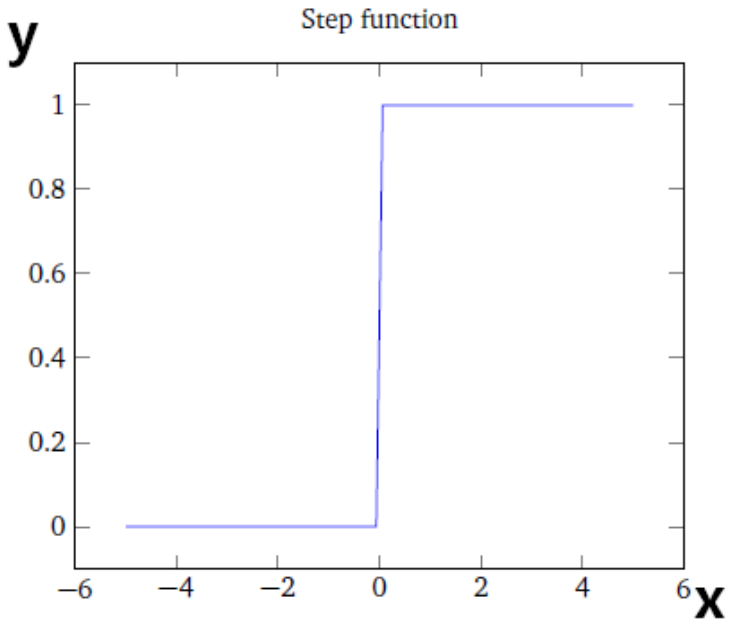


Figure 2.2: Shape of a step function that represents the outputs of a perceptron (Nielsen, 2015 [29]).

2.2 Sigmoid Neuron and the learning principle

The so called sigmoid neuron is the fundamental unit of modern ANNs. The sigmoid neuron works as a perceptron, hence, it takes and weighs inputs and owns a bias. The peculiar difference with the perceptron is that the output is not binary but can take any values between 0 and 1 included. The output of a sigmoid neuron can be defined as a function of inputs, weights and bias, $\sigma(wx + b)$, where sigma represents the sigmoid function:

$$\sigma \equiv \frac{1}{1 + e^{-}z}$$

Precisely considering the inputs $x_1, x_2, ..$, weights $w_1, w_2, ..$ and a bias we can write the sigmoid function as:

$$\sigma(z) \equiv \frac{1}{1 + exp(-\sum_i w_i x_i - b)}$$

where $wx + b = z$. Algebraically when z is a large positive number the output of a sigmoid neuron is approximately 1, on the other hand when z is very negative the output is close to 0. When z is of a modest size the output from a sigmoid neuron can take value between 0 and 1 and as said above in this paragraph this is the peculiar difference with the perceptron model. This is possible because the smoothness of the sigmoid function (see figure 2.3), contrary to the step function as we can see from figure 2.2, allow to change slightly the weights and biases to have a little change in the output while small changes of perceptron paramaters can cause the output to completely flip from 0 to 1 and viceversa. The possibility to apply small changes in weights Δw_i and bias Δb_i for obtaining small changes in the output allows to change parameters values to get as close as possibile to the best output. (Nielsen, 2015 [29]).

$$\Delta output \approx \sum_i \frac{\partial output}{\partial w_i} \Delta w_i + \frac{\partial output}{\partial b} \Delta b$$

The variation of the output changes linearly with the variations of weights and bias (Nielsen, 2015 [29]).

2.3 Learning process of a neural network: gradient descent and backpropagation

The learning process is composed by three sections in sequence: *forward-propagation* of the information, computation of the loss function and *back-propagation*. The forwardpropagation occurs when the network is trained

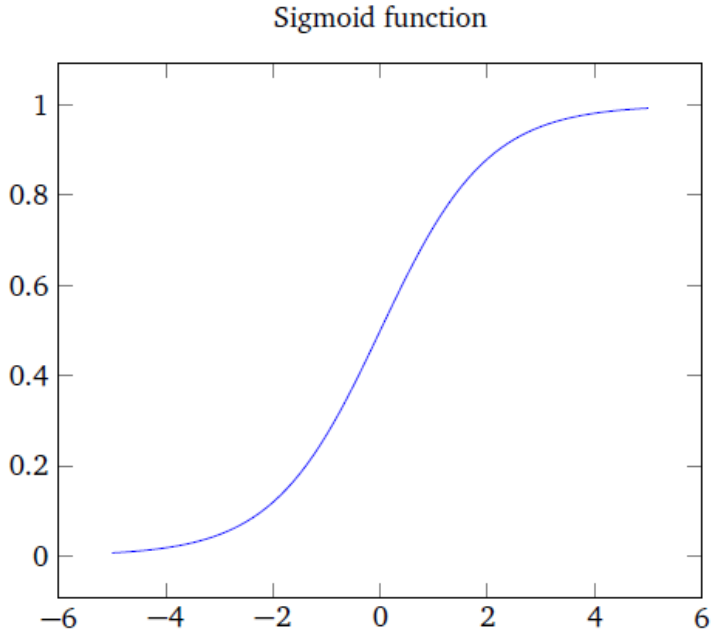


Figure 2.3: Shape of a sigmoid function that represents the outputs of a perceptron (Nielsen, 2015 [29]).

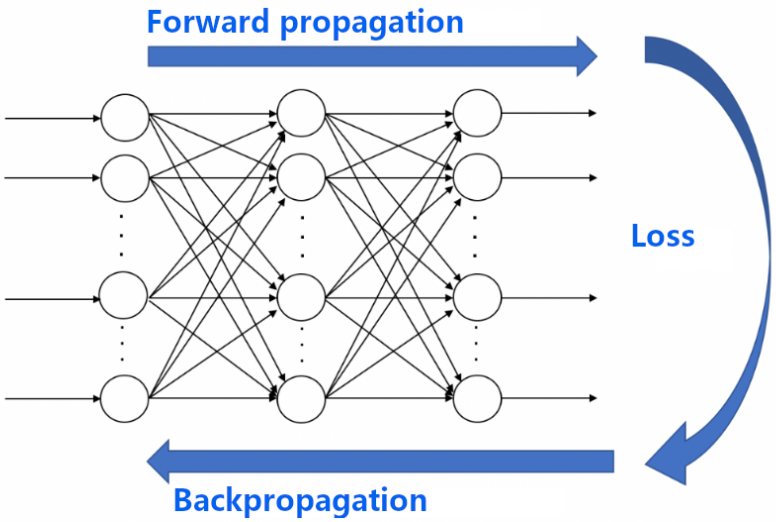


Figure 2.4: Segments of the NN learning process.

with the training dataset, hence, input data enter in the networks and they are processed trought the neurons. Neurons apply their transformaton to the information they receive from the neurons of the previous layer accordingly with their own activation function. Once the data has crossed all the layer, the final layer will give a result that explains how good are the processed input data against a ground truth that is used as a target for prediction. The transformations operated by the neurons and the modifications of parameters accordingly to the loss value is computed through the *gradient descent* techinque. We are indeed on a supervised environment because we have a target and we want to verify how good or bad is the prediction in relation to that target. (Nielsen, 2015 [29]).

We use a *cost function* to do this. The cost function computes the difference between the estimated (prediction) and the expected value (target), hence, we want to minimize the cost function throug the adjustament of the weights and bias of the neurons. Once the difference (loss) has been calculated its value is propagated backward. The backpropagation spread the information through all the neurons that have contribuited to the output. Those neurons will receive only a fraction of the total value of loss in relation to what they have contribuited to the output. The sequence of these three steps makes the network able to learn. (Nielsen, 2015 [29]).

2.3.1 Cost function

The cost function is computed as:

$$C(w,b) \equiv \frac{1}{2n} \sum_{x} \|y(x) - a\|^2$$

$y(x)$ represents the desired output from the network, a is the vector of outputs from the networks, x the input w all the weights in the network, b all the biases. The sum is computed over all the training inputs x. The vector a depends on x, b and w. The cost function considered is the quadratic cost function, also known as MSE Mean Squared Error. We choose that because

we used it as cost function in the NNs of this thesis. The cost function is a function of bias and weights and it represents a tool to quantify how well the network output a approximates the desired output $y(x)$, hence, we need to find an algorithm to adjust weights and biases which let the output to be close to the desired one and so to minimize the cost function $C(w, b) \approx 0$. To do this we use the gradient descent technique. (Nielsen, 2015 [29]).

2.3.2 Gradient descent techinque

The objective is to minimize $C(w, b)$. We can think about the values of bias and weights as the current state of our model. Hence, it?s possibile to define the variation of C as $\Delta C \approx \nabla C \cdot \Delta v$ with v as a substitude of b and w where $\Delta v \equiv (\Delta v_1, \Delta v_2)^T$ is the vector of changes of bias and weights and $\nabla C \equiv (\frac{\partial C}{\partial v_1}, \frac{\partial C}{\partial v_2})^T$ is the gradient vector of C. An interesting property of $\Delta C \approx \nabla C \cdot \Delta v$ is that it provides the possibility to choose Δv to make ΔC negative. In particular we can define $\Delta v = -\eta \Delta C$ where η is a small, positive number called learning rate. With $\Delta C \approx \nabla C \cdot \Delta v$ and $\Delta v = -\eta \Delta C$ we find $\Delta C \approx -\eta \|\nabla C\|^2$, this means that ∇C always decreases, never increases if we change v accordingly with $\Delta v = -\eta \Delta C$. Then, from $\Delta v = -\eta \Delta C$ we can calculate the next state of the system toward the minimization C defining $v \to v' = v - \eta \nabla C$ where v' represents weights and biases values of next state of the system. The gradient descent consists to repeatedly ∇C and then move the system in the opposite direction, indeed, we found that ∇C always decreases. Setting the size of the move $\Delta v = \epsilon$ with ϵ small fixed positive number, from $\Delta v = -\eta \Delta C$ we find $\eta = \frac{\epsilon}{\|\nabla C\|}$ so we can think about the gradient descent as taking small steps toward the best direction to minimize C. The learning rate has to be small enough to prevent ΔC to be positive but not too small because the algorithm would be too slow to reach the minimum of C. Expliciting $v \to v' = v - \eta \nabla C$ for weights and biases we obtain $w_k \to w'_k = w_k - \eta \frac{\partial C}{\partial w_k}$ and $b_l \to b'_l = b_l - \eta \frac{\partial C}{\partial b_l}$ that represent the gradient descent rule in term of components of our system/network. (Nielsen, 2015 [29]).

Stochastic gradient descent

Stochastic gradient descent is a technique used for speed up the learning process. The cost function has a form $C = \frac{1}{n}\sum_x C_x$ so that for a single training example $C_x \equiv \frac{\|y(x)-a\|^2}{2}$. That means that to compute ∇C we have to compute ∇C for each training input and then average them. In the practice this is time consuming and the network would reach the minimum of C slowly because usually the number of training inputs is large. Stochastic gradient descent solves this problem computing ∇C_x for a smaller sample of size m called *batch* of randomly chosen training data. (Nielsen, 2015 [29]). It results that choosing an appropriate size of the batch:

$$\sum_j^m \frac{\nabla C_{Xj}}{m} \approx \sum_x \frac{\nabla C_x}{n} = \nabla C$$

Hence we can explicit ∇C as:

$$\nabla C \approx \frac{1}{m}\sum_j^m \nabla C_{Xj}$$

The training of network is therefore performed with a first batch where the ∇C is computed over the randomly chosen input data of the batch. When it terminates, the algorithm start the same process with another batch and so on until every training data is processed. At this point a training epochs is completed and it's possible to start with another one. The number of batch, epochs and the values of the learning rate and other parameters related to these are called *hyperparameters*. (Nielsen, 2015 [29]).

2.4 Dataset organization

When it comes to organize the data we differentiate it in three segments: training dataset, validation dataset and testing dataset. Training dataset is composed by the elements that the NN sees and from which it learns. Validation data set is composed by data that the NN sees and use for tuning

the hyperparameters as setting the number of hidden neurons or stopping the learning process epochs before the number defined initially. When the network is trained and it reaches the optimal hyperparameters setup the NN has to be tested against a "ground-truth". The testing dataset is composed by elements that the NN doesn't see during the training phase so it allows to assess the performance of the network through the error obtained from the output against the ground-truth. (Nielsen, 2015 [29]).

2.5 Convolutional Neural Network

Convolutional Neural Networks (CNNs) are an example of deep learning technique. Deep Learning involves all those NNs that are composed by many hidden layers of neurons and so a large amount of parameters, contrary to Shallow NNs that are less complex in terms of structure and parameters. CNNs are widely used in image recognition, their architecture allow to consider the spatial structure of an image that wouldn't be possible with a NN of fully connected layers. In image recognition indeed, every pixel of an image is considered as a neuron. It results possible to apply this concept to weather prediction considering the state of the atmosphere as an image with grid points instead of pixel. This not only results to be more accurate when it comes to data-driven weather prediction but makes CNNs faster to train than other NNs. CNNs are built around three concepts: local receptive fields, shared weights and pooling. (Nielsen, 2015 [29]).

2.5.1 Local receptive field

One of the peculiarity of the CNN architecture is the Local Receptive Field (LRF). This idea permit to think differently about the input layer of the network. Neurons aren't disposed vertically as in figure 1.4 but as a square of neurons as in figure 2.5. Input neurons are therefore connected to the neurons in the hidden layers but contrary to those in figure 1.4 every unit in the hidden layer is connected with a small region of the input neurons

square called Local Receptive Field (figure 2.6). A hidden neuron learns about a particular receptive field and the next hidden neuron in the same hidden layer is connected with a window over input neurons shifted to the right of an arbitrarily number of neurons from the previous one (figure 2.7). (Nielsen, 2015 [29]).

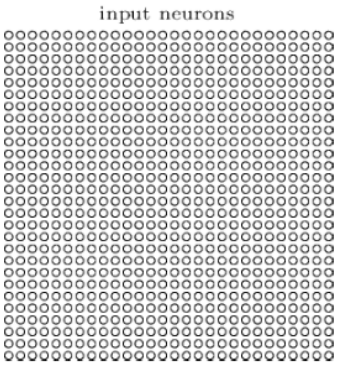


Figure 2.5: Illustration of the disposition of neurons on a concolutional layer. We can think about them as pixels of an image or grid points of a NWP model output (Nielsen, 2015 [29]).

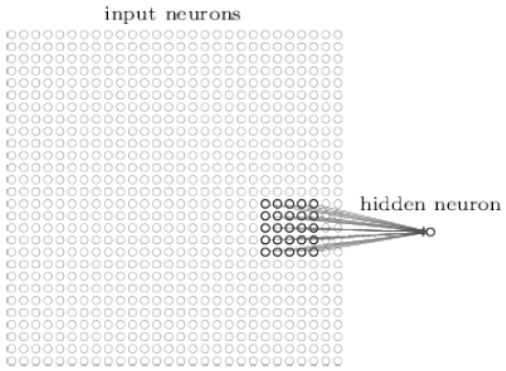


Figure 2.6: Local Receptive Filed of a hidden neuron (Nielsen, 2015 [29]).

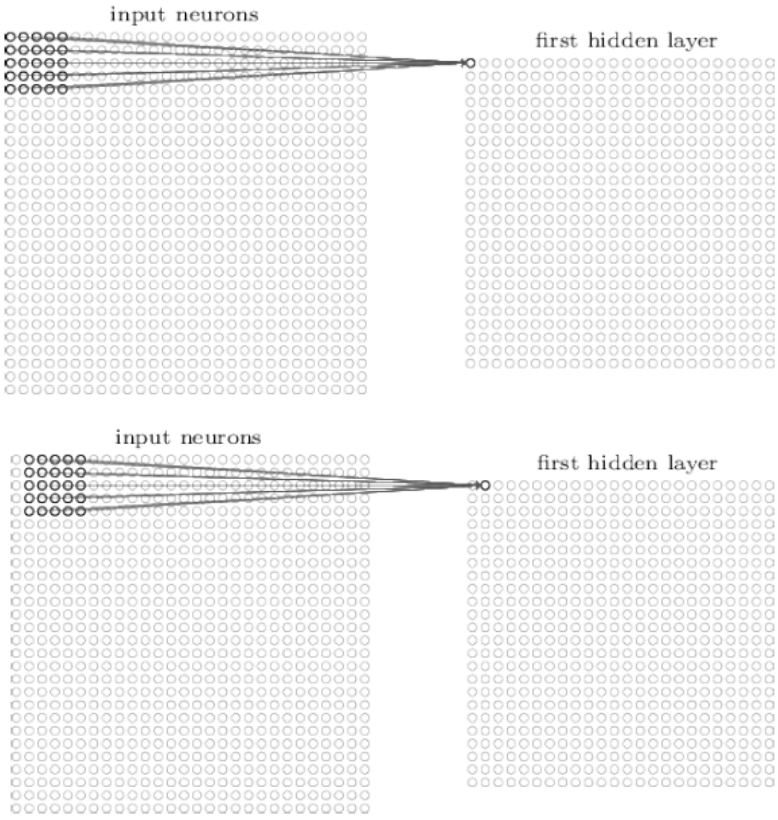


Figure 2.7: The first hidden neuron is connected to the first LRF **(a)** that will be shifted on the right for the second hidden neuron **(b)** and so on until all the hidden neurons have covered the entire input region (Nielsen, 2015 [29]).

2.5.2 Shared weights and biases

Every connection from the Local Receptive Field to the hidden neuron is weighted and each hidden neuron owns a bias value. A peculiarity of CNNs is that each hidden neuron has got the same weights and biases. This concept allows all the hidden neurons of a hidden layer (or convolutional layer) to detect the same feature in the input layer, for this reason, we say that the hidden layer is composed by feature maps. This idea is useful because allows the network to recognize and classify as a particular feature

(i.e. an atmospheric river) elements that are apparently different from each other. At this point of the discussion if we imagine a network able to detect 3 different features, the structure of the CNN would appear as in figure 2.8. A big advantage of shared weights and biases concept, sometimes called as kernels or filters, is that it allows to reduce significantly the number of parameters involved in the network and so improve the speed of training processes. (Nielsen, 2015 [29]).

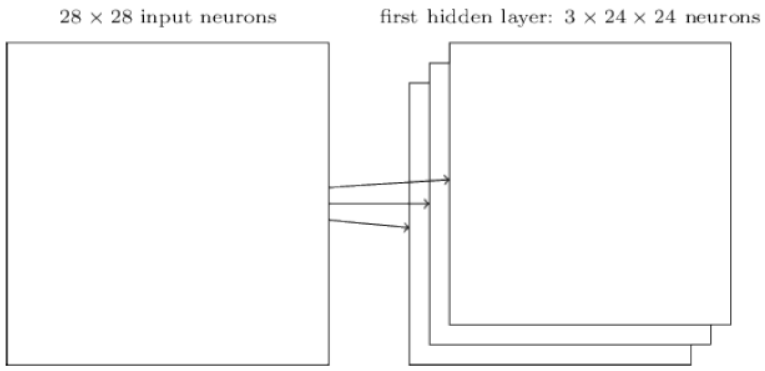


Figure 2.8: Illustration of an input layer composed by 28x28 neurons connected to a hidden layer composed by 3 feature maps with 24x24 neurons each (Nielsen, 2015 [29]).

2.5.3 Pooling layers

Pooling layers are used consecutively to convolutional layers and they contain as many elements as features maps in a hidden layer. A pooling layer is composed through a pooling technique as the max-pooling procedure that provides a simplification of the information contained in the feature maps. Indeed, every unit in the pooling layer is computed from an arbitrarily little window of neurons in the feature map in terms of maximum contributions of the neurons of the window with respect to the feature to which they are

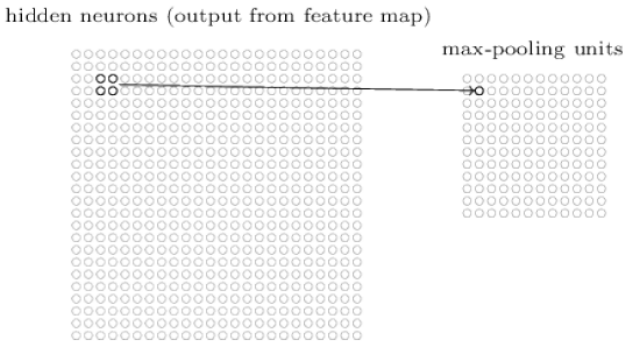


Figure 2.9: Max-pooling technique: the neurons selected from each window are those that give the biggest contribution to the feature they are assigned to (Nielsen, 2015 [29]).

assigned (figure 2.9). A perk of this concept is to reduce the number of parameters. (Nielsen, 2015 [29]).

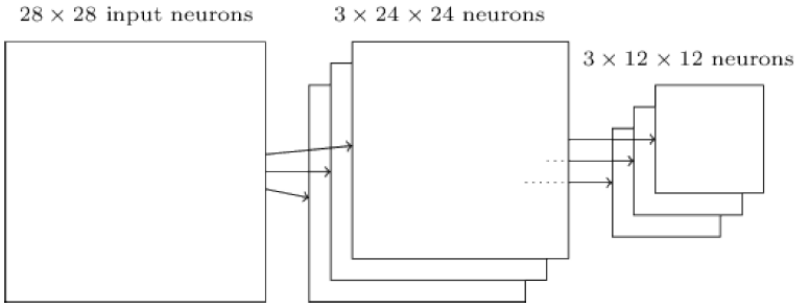


Figure 2.10: Pooling layers are consecutive to the hidden layers (Nielsen, 2015 [29]).

At the end each pooling layer is connected with the output layer whose neurons are disposed vertically as in the 1.4 (figure 2.11). For CNNs the learning processe composed by forward propagation, loss computatin and backpropagation works in the same way that has been described in the sections above. (Nielsen, 2015 [29]).

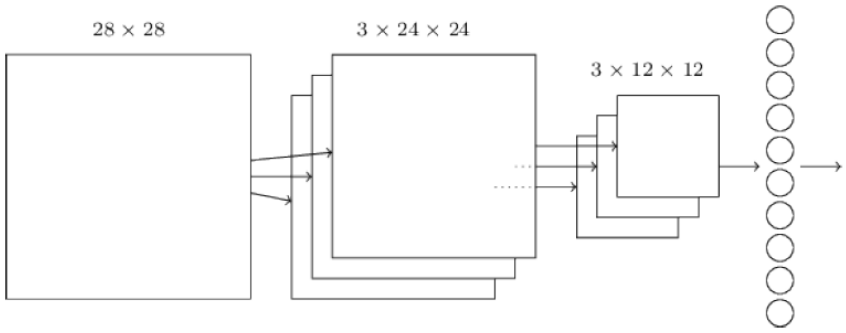


Figure 2.11: Convolutional Neural Network general structure with 3 features maps and 3 pooling layers with 12x12 neurons each (Nielsen, 2015 [29]).

Chapter 3

Atmospheric rivers forecasting

3.1 Identification and detection of Atmospheric Rivers

To understand the critical role of ARs in precipitation events it's important to define methods for identifying and predicting ARs.

An air column of a typical AR contains a quantity of water vapor that if it were condensed it would create a layer of 2 centimeters or more of thickness and the majority of that whitin the first 2.5 km of the column characterized with a strong low-level (>12.5 ms^{-1}) and humid wind (Ralph and Dettinger, 2011 [9]). Therefore the identification and the estimation of the strength of an AR are performed through monitoring and tresholding two key features: water vapor content e wind speed. Accordingly, two quantity are used for detecting ARs: 1) Integrated Water Vapor (IWV) obtained through satellite measurements and reanalyses or models; 2) Computing the vertically Integrated Horizontal Water Vapor Transport (IVT) from reanalysis or model (Ralph et al., 2004 [37]; Ralph and Dettinger, 2011 [9]).

The IWT [kgm^{-2}] (Fionda et al., 2019 [8]) is defined as

$$IWT = \frac{1}{g} \int_{p_s}^{px} q dp$$

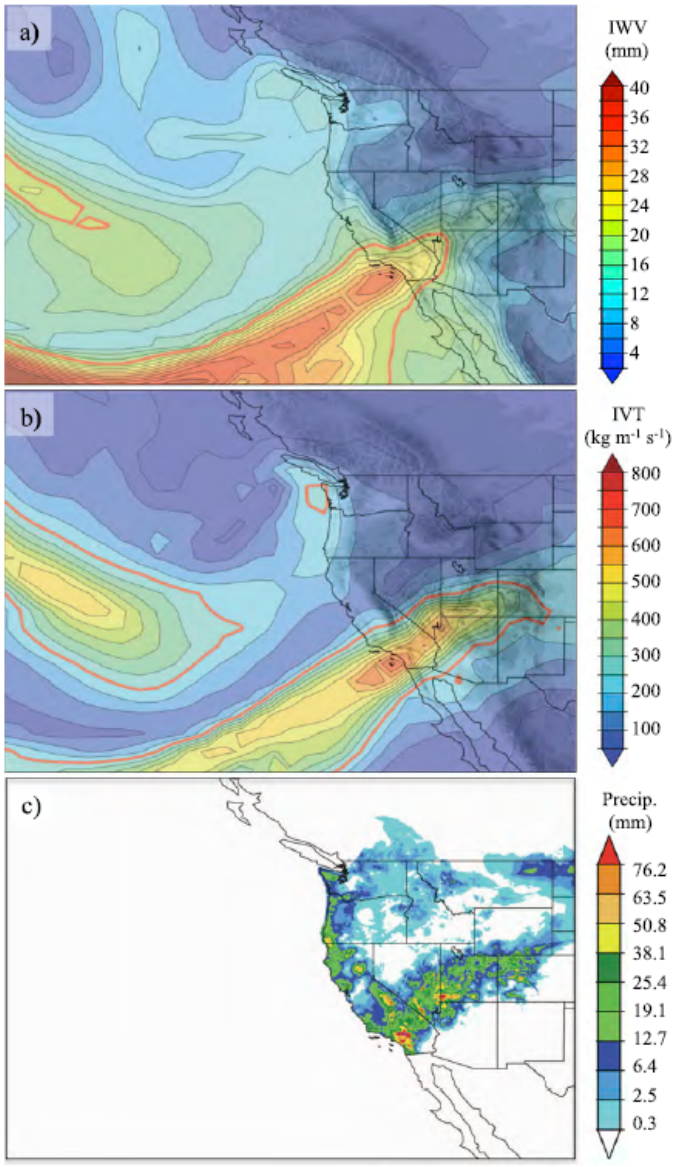


Figure 3.1: IWT and IVT fields are respectively represented in figures (a) and (b) at 0000 UTC 21 December 2010. The red line defines in (a) the threshold value of $20mm$ and in (b) the threshold value of $250kgm^{-1}s^{-1}$. In figure (c) represents the analysis of accumulated precipitation in a day period until 12 UTC 21 December 2010 (Rutz et al., 2014 [17]).

where g is the gravitational acceleration $[ms^{-2}]$, q is the specific humidity (dimensionless), p is the pressure $[kgms^{-2}m^{-2}]$, in particular, p_s is the pressure at the surface and p_x the pressure at a desired level (Rutz et al., 2014 [17]). In Rutz et al., (2014) [17] IWT is computed through an integration from the surface to 100-hPa, with intervals of 50-hPa from the surface to 500-hPa, and 100-hPa from 500 to 100-hPa. An AR is defined, using IWT, as a continuos region $\geq 2000km$ in length and with IWT $\geq 20mm$ (considering mm as if all the water vapor in the column were condensed in liquid water and spread evenly in the column). Adding the wind to the IWT equation, the IVT $[kgm^{-1}s^{-1}]$ is defined as:

$$IVT = \frac{1}{g}\int_{p_s}^{px} qV\,dp$$

where V wind vector $[ms^{-1}]$ (Rutz et al., 2014 [17]). The IVT defines an AR as a contiguous region $\geq 2000km$ in length with $IVT \geq 250kgm^{-1}s^{-1}$. In both cases, IWT and IVT, the length of an AR is defined as the greatest distance between two points within each contiguous feature (Rutz et al., 2014 [17]). The width of an AR it's not considered, indeed, an AR satisfying the required characteristics by the IVT thresholds is not usually wider than 1000 km. Rutz et al., (2014) [17] again, showed how the choice to use IVT, despite some studies use IWT to present results, could be more suitable in the detection of ARs to predict the rain, in particular on the West Coast of United States.

Firstly as it shown in (Junker et al., 2008 [18]; Neiman et al., 2002 [34]; Rutz et al., 2014 [17]) IVT is greatly associated to precipitation over topographical obstacles respect the IWT. Especially on the western United States is more firmly correlated with cool-season precipitation (figure 3.1)(Rutz et al., 2014 [17]). Second, Rutz et al., (2014) [17] shows in figure 3.1 that ARs defined by the treshold IVT$\geq 250kgm^{-1}s^{-1}$ penetrate more in land then when they are described by IWT$\geq 20kgm^{-2}$. Moreover IVT describes in a better way the spatial extension interested by heavy precipitation events and helps to solve the problems related to the decreasing of atmospheric thickness

and IWT values over complex terrain (Rutz et al., 2014 [17]).

3.1.1 Prediction of Atmospheric Rivers

Our overall interest is to predict the presence and intensity of ARs because they produce anomalous precipitation events. So why don't we focus our effort on predicting precipitation directly? Lavers et al., (2016) [23]) found that IVT (and hence ARs) is strongly related to synoptic-scale atmospheric processes such as extratropycal cyclone's warm belts. In the non-AR cases, precipitation is controlled by mesoscale atmospheric processes which can produce highly localized events. This difference of scales makes mesoscale precipitation less predictable than IVT. The IVT, being strongly correlated with precipitation and being associated with large-scale processes turns out to be easier to predict, therefore, is a better way to predict precipitation.

3.2 Skills of NWP models in ARs forecasting

NWP models show large forecasting errors in the landfalling of ARs (Ralph et al., [39]). Ensemble forecast systems provided by five main forecasting centers were analyzed by Wick et al., (2013) [10]. It is found that they reproduce: the IWV content along the AR axis, its water vapor content and the position of the AR once the land is reached and the AR width. The models resulted able to predict the presence of ARs 10 days beforehand, however, they failed in the timing and position of the landfall, especially for longer lead times (Wick et al., 2013 [10]). (Chapman et al., 2019 [50]) tried to improve the prediction and detection of ARs with a Machine Learning approach.

Chapman et al., (2019) [50]) is the study on which this thesis is based. The study evaluated the improvement performed by a Convolutional Neural Network of the IVT prediction from the Global Forecast System (GFS) over a region of interest from 180° W to 110° W longitude, and 10° N to 60° N. The aim of the study was to improve the IVT forecast with respect to the

forecast on the same region provided by a dynamical NWP model using GFS, a climatological forecast (CF) and a persistence forecast (PF) (Chapman et al., 2019 [50]).

The dataset used composed of GFS's IVT forecasts at a 0.5°horizontal spatial resolution on 64 vertical levels with two daily initializations at 0000 and 12000 UTC. A 7 days forecast has been analyzed with 3 hours increment for the firsts 12 hours, 12 hours increment for the next day and 24 hours increment until the 168th hour. The forecasts covered a period of time from October to April from 2006 to 2018. As a ground-truth IVT from MERRA2 reanalysis has been used, with a 0.625°x0.5° grid on 21 pressure levels from 1000 and 300-hPa. GFS forecasts have been regridded and upscaled to MERRA2 resolution for consistency. The IVT forecasts have been divided in training (2008-2016), validation (2016-2017) and testing datasets (2017-2018).

To evaluate the forecast error the study applied four valuation metrics to the forecasts: root mean squared error (RMSE), bias (Bias), centered root-mean-square error (CRMSE) and spatial Pearson Correlation (PC) coefficient.

As shown in figure 3.2 for the first three hours of forecasts the CNN and the NWP model showed better results respect the climatology and persistence forecast, with the CNN improving on BIAS and CMRSE with respect to the NWP model. After five days CNN and the NWP model were almost equal about BIAS while CRMSE was still better for the CNN. At the seventh day CNN performed better than other models in terms of RMSE and showed a higher correlation with the ground-truth through all the forecast period, in particular, at the 7th day CNN's correlation corresponded to that of the NWP model at the 6th day. The study showed that CNNs as postprocessing framework can improve a seven days IVT forecast respect to a GFS NWP model in terms of RMSE improvement at every lead time with a higher correlation between forecast and ground truth after the 12 hours lead time.

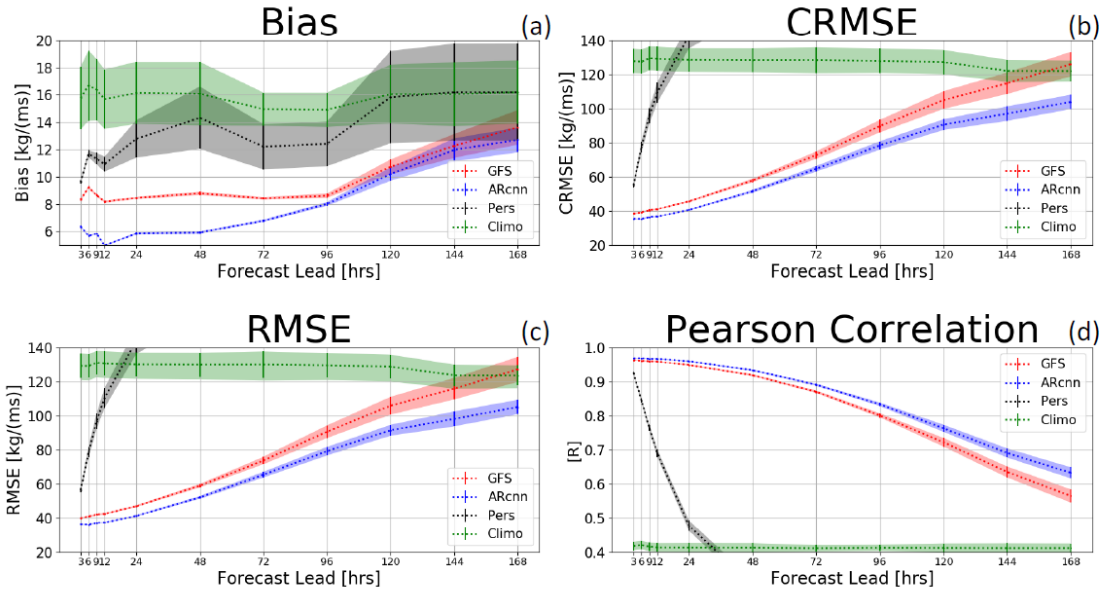


Figure 3.2: Valuation and comparison of the performances of the CNN (ARcnn) in blue, GFS NWP model (GFS) in red, climatological forecast (Climo) in green and persistence forecast (Pers) in grey, in terms of RMSE (c), Bias (a), CRMSE (b) and correlation (d) obtained in Chapman et al., (2019). (Chapman et al., 2019 [50]).

Chapter 4

Improving reforecast with Convolutional Neural Networks

On the basis of the positive results obtained in Chapman et al., (2019) [50] and described in the previous section, we will now use a new deterministic model, so-called Weather Research Forecast (WRF), to re-forecast for a period of 30 years for the same region of Chapman et al., (2019) [50]. We will use two different convolutional networks, CNN1 and CNN2 and we will compare the forecast skill with respect to a reference data set, this time the MERRA2 reanalysis. The reforecasts data are used as training, validation and testing data set for both the CNNs that are evaluated with MERRA2 reanalysis in terms of Root Mean Squared Error (RMSE), Bias and Person Correlation (PC) Coefficient and compared between each other and with West-WRF. The experiment is performed over a region of interest (ROI) from 25°N to 60°N latitude, and 150°W to 115°W longitude.

4.1 Data

4.1.1 West-WRF Reforecast

Reforecasts products are computed from the forecasts of the same model extended for several years in the past. The West-WRF reforecasts have

been created with the goal of: 1) investigating the skill of a high-resolution reforecast; 2) conducting process-based studies of atmospheric rivers; and 3) improving predictive skills by the application of postprocessing techniques and machine learning to reduce errors (Martin et al., 2018 [26]).

West-WRF reforecast dataset involves 30 years of forecasts (1987-2017) with each year composed by the four-months period 1st January-31st March and 1st-31st December. The reforecast is based on WRF version 4.0.1 and the forecasts are initialized at 0000 UTC daily using the 0.5° Global Ensemble Forecast System (GEFS) Version 10. The entire dataset covers the spatial domain showed in figure 4.1 composed by a 9 km horizontal spatial resolution outer domain and a 3 km horizontal spatial resolution inner domain, both having 60 vertical levels until a model top of 10 hPa (Martin et al., 2018 [26]). In this study only the outer domain is used, that it is composed of 618 by 569 grid points. Forecasts are produced for 168 hours lead time (7 days) with time resolution of 1 hour.

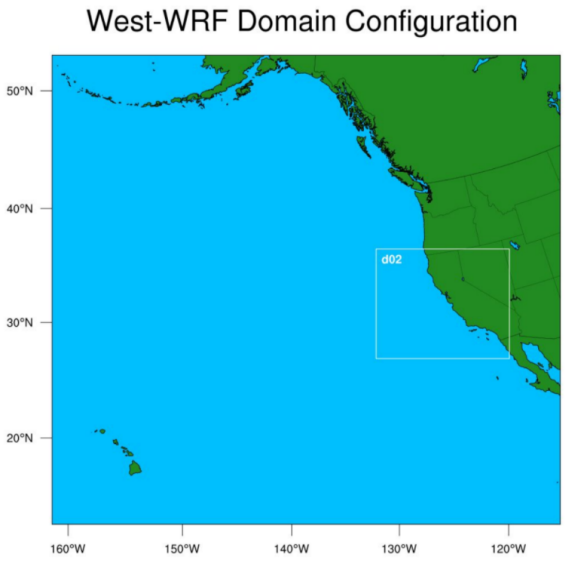


Figure 4.1: The full plot represents the outer domain with 9km horizontal spatial resolution and the smaller domain d02 represents the inner domain with 3km horizontal spatial resolution (source: https://cw3e.ucsd.edu).

4.1.2 MERRA2 reanalysis: ground-truth

Reanalysis data are created by the combination of observations made in the past with a weather model providing the best possible reproduction of the atmosphere at a given time (Dueben and Bauer, 2018 [36]). The benefit to train a network with reanalysis instead of observations is that they provide data for each grid point at each time step, consistently over the entire data set (Dueben and Bauer, 2018 [36]). In this study MERRA2 reanalysis is used as a ground-truth, representing the target for both the CNNs and so to assess their training error and the forecast error of West-WRF. MERRA2 data are structured on 0.625°x0.5° grid interpolated within the interval 1000 and 300 hPa divided in 21 vertical layers (for IVT computation) (Chapman et al., 2019 [50]).

4.2 Methodology and experimental design

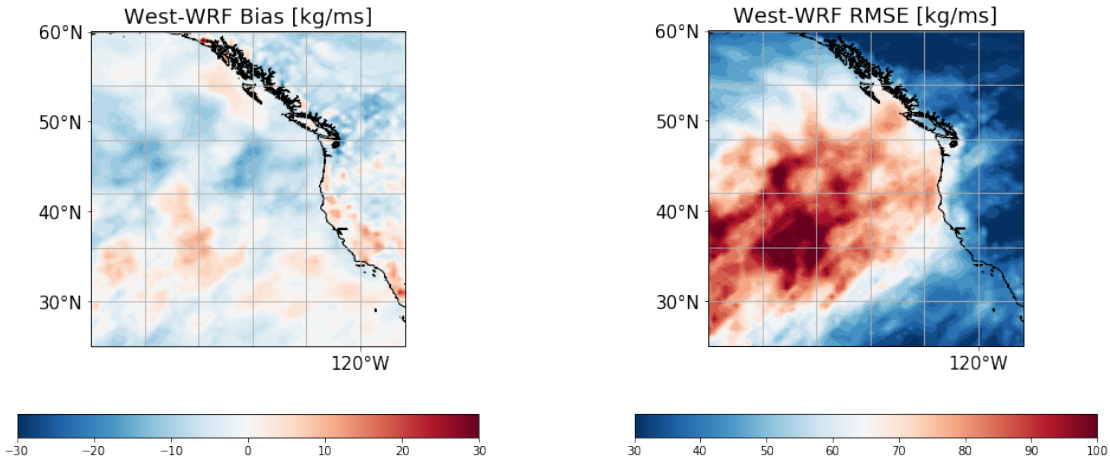
The variable of interest to detect ARs is the amplitude of the IVT vector field. West-WRF provides the variable "IVT"(amplitude of the IVT vector) that represents the forecasted IVT and MERRA2 provides the target variable "IVTm". For consistency West-WRF IVT is regridded and upscaled to MERRA2 resolution. For simplifying the problem and making the training faster we examined a forecast of 48 hours with an increment of 6 hours for a total of 9 forecasts lead times from 0000 (initialization) to 0048 UTC over a time period from 1987 to 2017 for 8 vertical levels from the surface to 200 hPa. The experiments are conducted over a region from 25°N to 60°N latitude, and 150°W to 115°W longitude.

4.2.1 Verification metrics

To evaluate the ability of the two Convolutional Neural Networks we compared the IVT forecast provided by the West-WRF reforecast with the IVT forecast postprocessed by the two CNNs, CNN1 and CNN2. WRF,

CNN1, CNN2 provides a forecasted IVT field (f) whose error $e = f - g$ is computed with respect to the ground-truth (g) IVT field provided by MERRA2. The skill is then quantified in terms of root mean squared error (RMSE), bias (Bias) and Pearson Correlation (PC) coefficient that we will call simply correlation.

4.2.2 Error patterns of West-WRF



(a) (b)

Figure 4.2: Spatial distributions of the West-WRF error patterns defined by the difference between the West-WRF and MERRA2.

In figure 4.2 the error patterns of West-WRF with respect to MERRA2 are shown in terms of Bias and RMSE at 48-hour forecast lead time computed over the testing dataset. The biggest source of error is from the areas where the IVT field is usually more intense (see figure 1.3(1)). This could lead to wrong predictions of positioning and magnitude of the IVT field. West-WRF under-estimates the IVT at high latitudes and over-estimates it at low latitudes. This could be caused by the mesoscale frontal waves associated with ARs characterized by lower predictability and is consistent with the errors pattern of GFS showed in (Chapman et al., 2019 [50]).

4.2.3 Convolutional Neural Nerworks configuration

Both of networks produced the best results with 50 training epochs and batch size 20. CNN1 and CNN2 have been optimized with a mean squared error loss and after the training they provide an estimated IVT field that is corrected from some portion of the forecast error that have been learned from the networks during the training. The CNN1 has performed better with a learning rate of 0.0009, differently from CNN2 that exhibited at 0.0005 learning rate. We tested 4 similar regions and we noticed that reducing the size of it, the evaluation metrics and the speed of training improved for both networks. The activation function used for CNN1 and CNN2 is a linear one. The optimal configuration for both the CNNs was found splitting the training dataset in 24 years (1987-2011), validation and testing dataset in 3 years each one respectively 2012-2014 and 2015-2017, so with different proportions with respect to the standard subdivision 50:25:25 (training,validation,testing). Iteratively, the training data help to optimize the weights of the CNNs while the validation datasets are utilized to calculate the performance metrics during the training. The networks do not see the testing data; indeed, they are independent from the training data and they are used for assessing the postprocessing ability of the networks. For both the networks a new CNN is generated and trained for each forecast lead time. Due to potential similar IVT features across the lead times, the training of each lead time follows a sequential scheme that allows the weights of the network at a particular lead time to be initialized by the network weights of the previous lead time. This reduces the training time with an improvement of the overall error results (Chapman et al., 2019 [50]). The optimal configuration for the architecture and parameters of the networks was reached through a large number of experiments. In the next paragraph results on the performance of the networks accordingly to the variation of parameters and the regions are shown in terms of the valuation metrics.

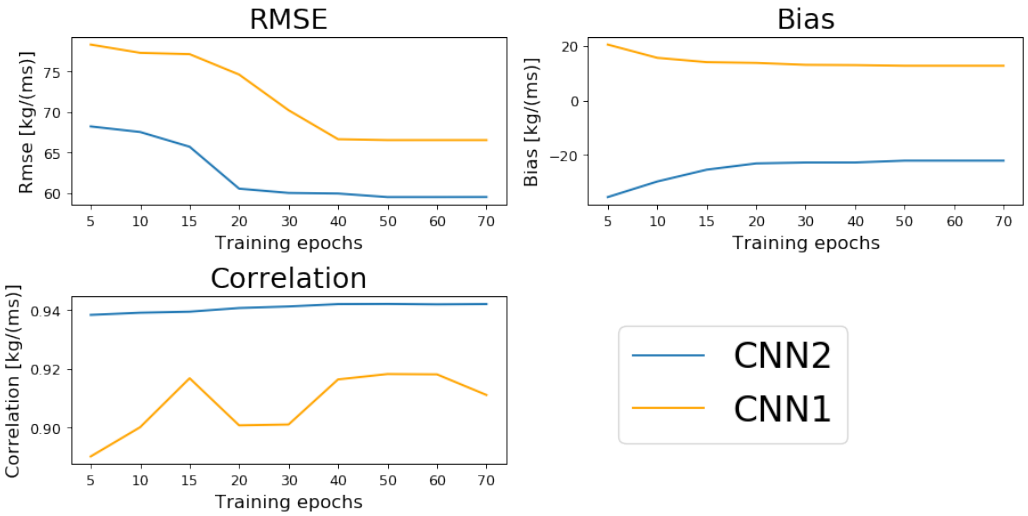


Figure 4.3: Comparison of CNN1 and CNN2 in terms of valuation metrics at 48-hour lead time for each value of training epochs tested.

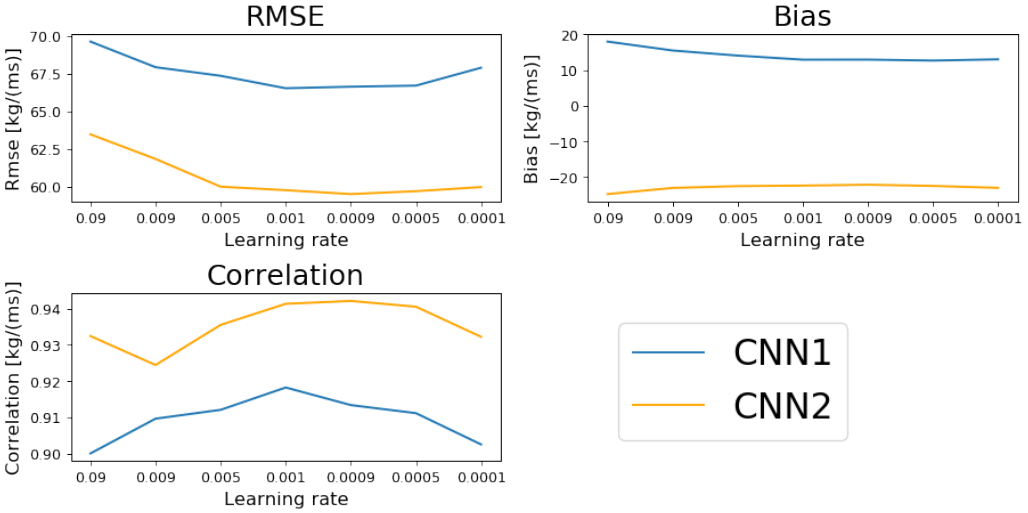


Figure 4.4: Comparison of CNN1 and CNN2 in terms of valuation metrics at 48-hour lead time for each value of learning rate tested.

4.3 CNNs optimization

Convolutional neural networks are regulated by many parameters. In this study for simplicity and consistency we analyzed the best configuration of the

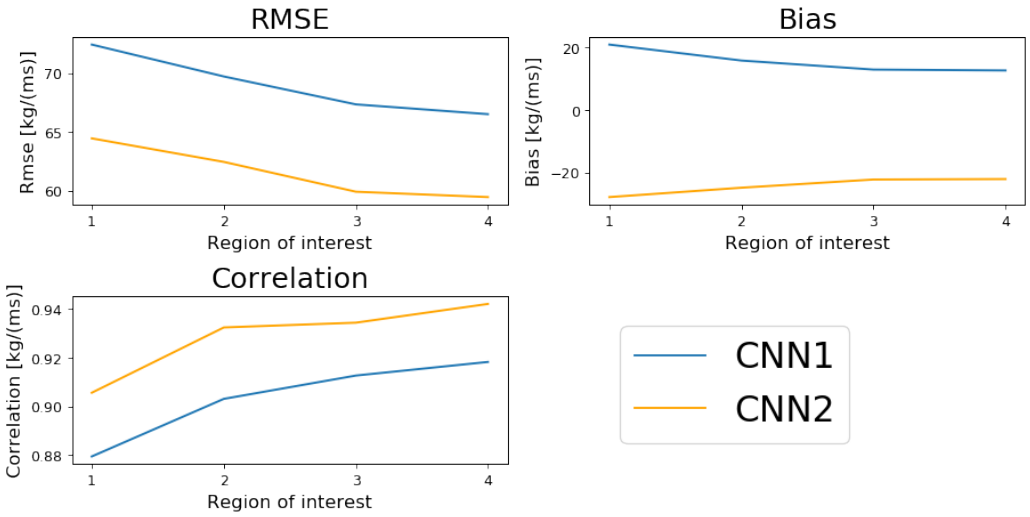


Figure 4.5: Comparison of CNN1 and CNN2 in terms of valuation metrics at 48-hour lead time for each region of interest tested.

networks in terms of epochs training, learning rate and variation of the region of interest size over which we performed the prediction (see figures 4.3,4.4 and 4.5). This of course, does not exclude the possibility that analyzing and changing other parameters would lead to a better configuration of the networks. For testing the best performance in terms of epochs we trained the networks for each forecast lead time with 9 different epochs value (a total of 81 trained lead times) holding the size of the region of interest and the value of the learning rate with their best performance value. In figure 4.3 are represented the three valuation metrics at 48-hour lead time for each value of training epochs we tested. For both the networks we choose 50 epochs as best value. The RMSE of CNN1 and CNN2 decreased significantly after 20 epochs and it is almost stabilized between 40 and 70 epochs. None of the networks reached the 60 and 70 epochs of training, both showed an early stopping of the training within the 50 epochs, except in the training of few lead times, caused by the lack of improvement in the forecast error. Bias and correlation accordingly with the RMSE showed stable value after 40 epochs of training. The best learning rate value was obtained (see figure 4.4)

through the same experiment design we used for obtaining the best epochs value. We noticed that decreasing the learning rate from 0.09 to 0.0001 we obtained a significant improvement of RMSE and correlation. The learning rate reached its best value in terms of RMSE at 0.0009 learning rate for CNN1 and 0.0005 for CNN2 while the best values for the correlation are slightly shifted to the left, 0.001 for CNN1 and 0.0009 for CNN2. We have chosen the RMSE as more representative of the overall performance. About the region of interest (ROI), we tested 4 different regions of interest (see figure 4.5): region 1 from 10°N to 60°N latitude, and 180°W to 110°W longitude; region 2 from 15°N to 60°N latitude, and 170°W to 110°W longitude; region 3 from 20°N to 60°N latitude, and 160°W to 112°W longitude; region 4 from 25°N to 60°N latitude, and 150°W to 115°W longitude that is the region of interest on which is based this study. Initially we did not want to change the ROI because we aimed to compare the performance of the networks trained with West-WRF reforecast data with the results obtained by (Chapman et al., 2019 [50]) with the CNN trained with GFS. However, as first approach to Neural Networks was interesting to see how reducing the size of the ROI, both the networks performed better in terms of valuation metrics and speed of training (not reported). Hence, we have chosen to train CNN1 and CNN2 with region 4.

4.4 Results

4.4.1 CNN1 and CNN2 used as postprocessing tools

Figure 4.6 shows the intensity of the IVT field at 48-hour forecast lead time of the 18th January 2017. The West-WRF IVT field at 48-hour lead time is used as input for the 48-hour lead time CNN1 and CNN2. After the training of the networks CNN1 and CNN2 produced respectively a post-processed forecast in a few milliseconds. The West-WRF IVT field, hence, is processed by the two networks that produced a corrected IVT field compared against the 48-hour MERRA-2 IVT field (figure 4.6(a)). West-WRF

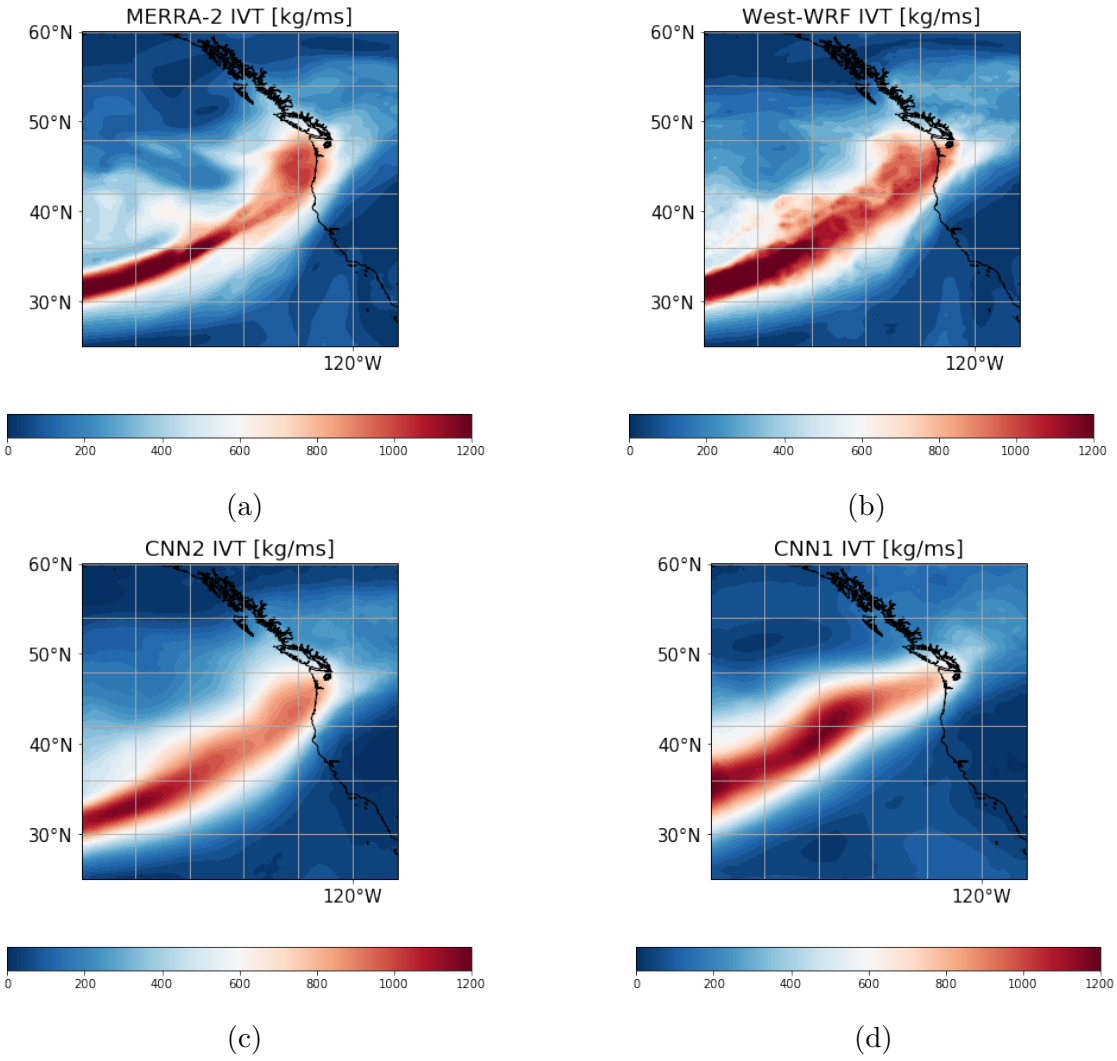
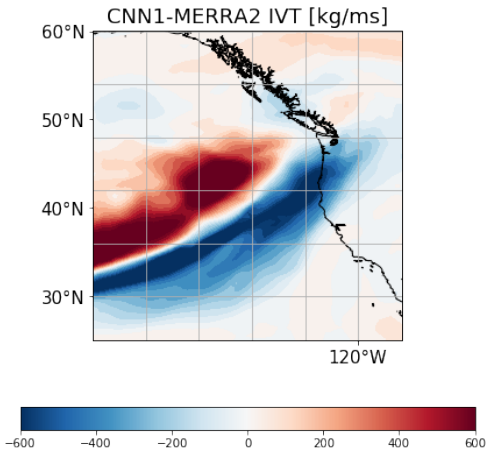
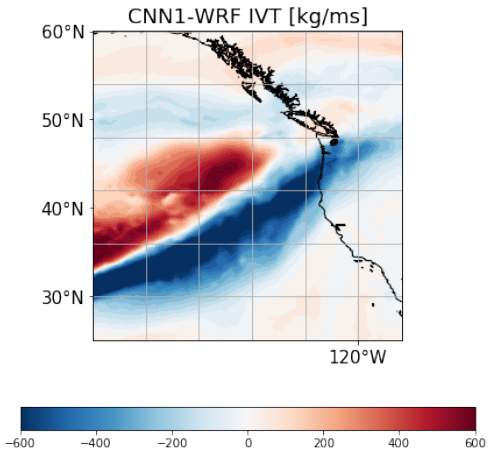


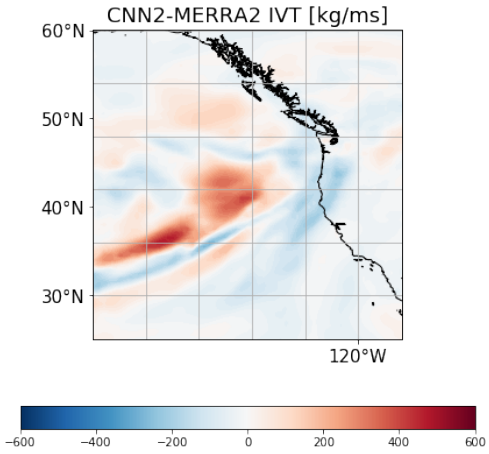
Figure 4.6: IVT field produced by CNN1(d) and CNN2(c) compared with the IVT field provided by MERRA2(a) and West-WRF(b) at 48-hour forecast 18th January 2017.



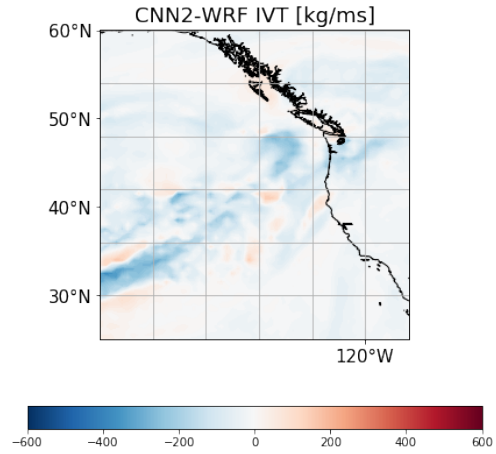
(a)



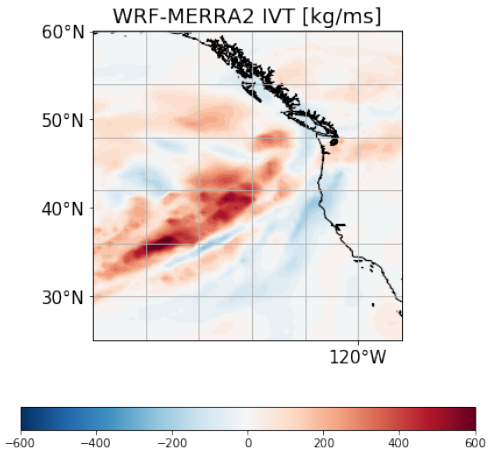
(b)



(c)



(d)



(e)

Figure 4.7: Difference between the IVT fields of CNN1 with MERRA2 and WRF (a),(b); CNN2 with MERRA2 an WRF (c),(d); WRF and MERRA2 (e).

(figure 4.6(b)) and CNN2 (figure 4.6(c)) perform well on the positioning of the AR. West-WRF shows a better result in modelling the intensity of the IVT filed than CNN2 that under-estimates the magnitude of the IVT field. On the contrary CNN2 better predicts the shape and the orientation of the IVT and it correctly reduces the IVT magnitude between the offshore zone of high IVT intensity and the land. CNN1 (figure 4.6(d)) misses the position of the AR and the positioning of the IVT high intensity zones, while seems acceptable the orientation of IVT field.

Figure 4.7 shows the difference of the IVT field between CNN1, CNN2 and West-WRF with MERRA2 (see 4.7(a),(b),(e)) and between CNN1 and CNN2 with West-WRF (see 4.7(b),(c)). From the difference between MERRA2 it is possible to notice how CNN1 fits with MERRA2 only in the center of the AR with a significant difference of intensity above and under the central axis of the AR. This means that CNN1 is significantly missing the position of the IVT field showing the same pattern against West-WRF. This pattern is less pronounced in the difference CNN2-MERRA2. CNN2 can better detect the positioning of the AR with a more pronounced difference in the magnitude of the IVT field in the northern part of the AR. CNN2 and West-WRF show similar pattern against MERRA2 in terms of AR positioning detection and magnitude estimation. CNN2-WRF shows a little difference of IVT field magnitude from the zone of high IVT intensity.

4.5 Overall performance of CNN1 and CNN2 on the testing dataset

The overall performance of CNN1 and CNN2 is shown in figure 4.8 in terms of the valuation metrics, root mean squared error (RMSE), bias, and Pearson Correlation (PC) coefficient from a seasonal point of view considering that both the networks have been tested over the four-months period 1st January-31st March and 1st-31st December for the years 2015,2016 and 2017.

CNN1 and CNN2 have been trained for each forecast lead time with their

best configuration. Considering the variability of the atmospheric fields and, in particular the IVT, it would have been better to produce a forecast with smaller time steps than 6 hours. We did choose that time step to reduce the time necessary for the networks training. Figure 4.6 shows that both CNN1 and CNN2 start with higher RMSE value respect West-WRF. While the CNN1 maintains this trend for the entire prediction, CNN2 significantly reduces the RMSE value with respect to West-WRF. In terms of correlation CNN2 outperforms CNN1 and the prediction made by West-WRF that shows a correlation value at 36-hour lead time that CNN2 reaches at 48-hour lead time. CNN1 shows the best result for bias at the 00-hour but it rapidly deteriorates at 06-hour lead time and remains worse than West-WRF up to the end of the second day. CNN2 although improves the West-WRF forecast for the entire period in terms of RMSE and correlation, shows a negative bias for the entire prediction. The negative bias is connected to the overall under-estimation of CNN2 with respect to the MERRA2 IVT field showed in the study case reported in the previous section (see figure 4.6(c)).

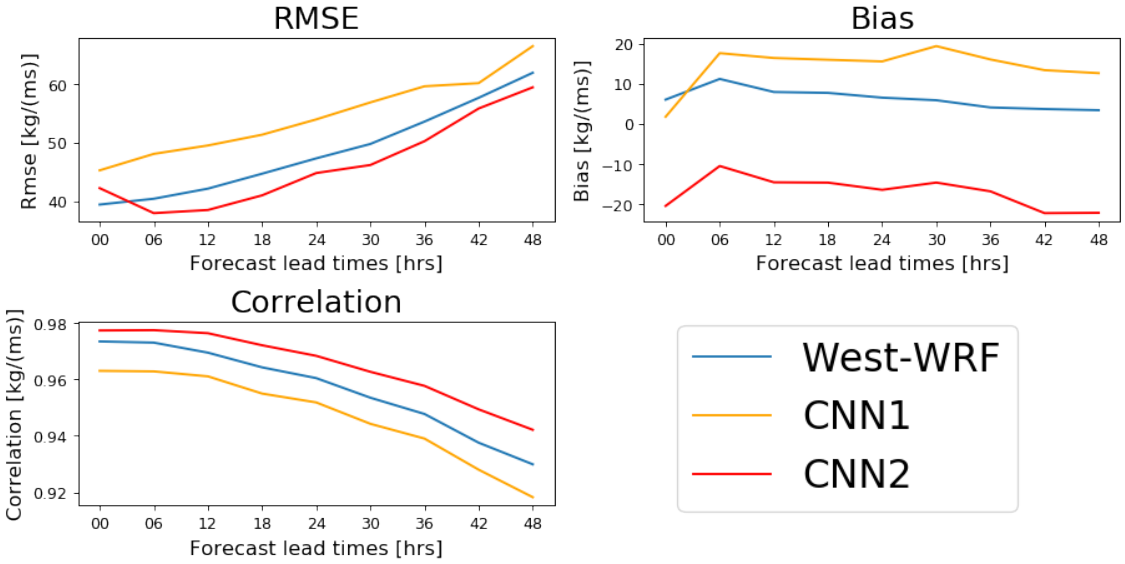


Figure 4.8: Overall performance of CNN1,CNN2 and West-WRF against MERRA2 on the entire testing data for the entire forecast period.

Chapter 5

Conclusions

This thesis was concerned with a special atmospheric phenomena, so called Atmospheric River (AR). It was found that such ARs produce intense precipitation events on land inducing flooding and consequent disasters, if not loss of lives. AR are large scale atmospheric structures, transporting large amount of water vapor and their horizontal scale is several hundred km. Given such large scale, the hypothesis is that ameliorating the AR transport of water vapor would improve the forecast of precipitation events on land. After having defined the key quantity to identify AR and their intensity, the Integrated Vapor Transport (IVT) amplitude, we have studied a high resolution reanalysis field (MERRA2) with re-forecasts from a limited area forecasting model, WRF over the California west coast. We applied two Convolutional Neural Networks (CNN1 and CNN2) to improve a 48 hours forecast of the IVT provided by the West-WRF 30-years reforecast of the CW3E department, UCSD, San Diego. The IVT field is used to detect Atmospheric Rivers over a region of interest from 25°N to 60°N latitude, and 150°W to 115°W longitude. After an optimization phase of the networks, both the CNNs have been trained and validated for each lead time with $\sim$3480 reforecast data, selecting as variable of interest, the amplitude of the IVT vector. Once the networks have been trained and validated, the outputs have been tested over the four-months period 1st Jan-31st Mar and 1st

Dec-31st Dec for the years 2015,2016 and 2017 with ∼300 reforecast data for each network. The corrected IVT fields from the networks have been evaluated against MERRA-2 reanalysis in terms of root mean squared error, bias and correlation. Although the results of CNN1 were outperformed by West-WRF, CNN2 improved slightly the IVT forecast of West-WRF with an improvement of the 0-2% in terms of RMSE. The CNN2 maintained better results with respect to West-WRF in terms of correlation, however, it underestimated the IVT intensity for the entire forecast period. From a general point of view we can consider the CNN2 capable to improve the IVT field of West-WRF model over the region of interest. In conclusion this thesis has shown that CNN is a promising forecasting technique but that still improvements are required, such as a better optimization of hyper-parameters for the neural network and probably a large input data set for the network training.

Bibliography

[1] A.D.Dongare, R.R.Kharde, Amit D.Kachare (2012), Introduction to Artificial Neural Network, *International Journal of Engineering and Innovative Technology (IJEIT), Volume 2, Issue 1*, page 189-193.

[2] Anderson, G. J., & Lucas, D. D. (2018). Machine learning predictions of a multiresolution climate model ensemble. *Geophysical Research Letters, 45*, 4273-4280.

[3] Bao, J.W., S. A. Michelson, P. J. Neiman, F. M. Ralph, and J. M. Wilczak (2006), Interpretation of enhanced integrated water vapor bands associated with extratropical cyclones: Their formation and connection to tropical moisture, *Mon. Weather Rev., 134*(4), 1063-1080.

[4] Bin Guan, Duane E. Waliser (2015). Detection of atmospheric rivers: Evaluation and application of an algorithm for global studies. *Journal of Geophysical Research, 120* (24), 12515-12535.

[5] Carter, G. M., Dallavalle, J. P., Glahn, H. R. (1989). Statistical Forecasts Based on the National Meteorological Center's Numerical Weather Prediction System. *Weather and Forecasting, 4*(3), 401-412.

[6] Delle Monache, L., Nipen, T., Deng, X., Zhou, Y., Stull, R. (2006). Ozone ensemble forecasts: 2. A Kalman filter predictor bias correction. *Journal of Geophysical Research Atmospheres, 111*(5), 1-15.

[7] Delle Monache, L., Nipen, T., Liu, Y., Roux, G., Stull, R. (2011). Kalman Filter and Analog Schemes to Postprocess Numerical Weather Predictions. *Monthly Weather Review, 139*(11), 3554-3570.

[8] Fionda E, Cadeddu M, Mattioli V, Pacione R. Intercomparison of Integrated Water Vapor Measurements at High Latitudes from Co-Located and Near-Located Instruments. *Remote Sensing*. 2019; 11(18):2130.

[9] Fred Martin Ralph, Michael D. Dettinger (2011). Storms, floods, and the science of atmospheric rivers. *Eos Transactions American Geophysical Union, 92* (32), 265-272.

[10] Gary A. Wick, Paul J. Neiman, F. Martin Ralph, Thomas M. Hamill (2013). Evaluation of forecasts of the water vapor signature of atmospheric rivers in operational numerical weather prediction models. *Weather and Forecasting, 28* (6), 1337-1352.

[11] Gershunov, A., T. Shulgina, F. M. Ralph, D. A. Lavers, and J. J. Rutz, 2017: Assessing the climate-scale variability of atmospheric rivers affecting western North America. *Geophys. Res. Lett., 44*, 7900-7908.

[12] Glahn, H. R., Lowry, D. A. (1972). The Use of Model Output Statistics (MOS) in Objective Weather Forecasting. *Journal of Applied Meteorology, 11*(8), 1203-1211.

[13] Hacker, J. P., Rife, D. L. (2008). A Practical Approach to Sequential Estimation of Systematic Error on Near-Surface Mesoscale Grids. *Weather and Forecasting, 22*(6), 1257-1273.

[14] Holden, P. B., Edwards, N. R., Garthwaite, P. H., Wilkinson, R. D. (2015). Emulation and interpretation of high-dimensional climate model outputs. *Journal of Applied Statistics, 42*(9), 2038-2055.

[15] Homleid, M. (1995). Diurnal Corrections of Short-Term Surface Temperature Forecasts Using the Kalman Filter. *Weather and Forecasting, 10*(4), 689-707.

[16] Jonathan A. Weyn, Dale R. Durran, Rich Caruana (2020). Improving data-driven global weather prediction using deep convolutional neural networks on a cubed sphere. *Journal of Advances in Modeling Earth Systems, 12* (9), ee2020MS002109.

[17] Jonhatan J. Rutz, W. James Steenburgh, Fred Martin Ralph (2014). Climatological characteristics of atmospheric rivers and their inland penetration over the Western United States. *Monthly Weather Review, 142* (2), 905-921.

[18] Junker, N. W., R. H. Grumm, R. Hart, L. F. Bosart, K. M. Bell, and F. J. Pereira, 2008: Use of normalized anomaly fields to anticipate extreme rainfall in the mountains of northern California. *Wea. Forecasting, 23*, 336?356.

[19] Knippertz, P., and H. Wernli, 2010: A Lagrangian climatology of tropical moisture exports to the Northern Hemispheric extratropics. *J. Climate, 23*, 987-1003.

[20] Koza, J.R., Bennett, F.H. III, Andre, D., Keane, M.A. (1996 a). Toward evolution of electronic animals using genetic programming. *Artificial Life V: Proc. 5th Int. Workshop on the Synthesis and Simulation of Living Systems* (Langton, C.G., Shimohara, K., Eds.), pp. 327-334.

[21] Lamjiri, M.A., M.D. Dettinger, F.M. Ralph, and B. Guan, (2017). Hourly Storm Characteristics along the U.S. West Coast: Role of Atmospheric Rivers in Extreme Precipitation. *Geophysical Research Letters*, 44.

[22] Larraondo, P. R., Renzullo, L. J., Inza, I., & Lozano, J. A. (2019). A data-driven approach to precipitation parameterizations using convolutional encoder-decoder neural networks. *arXiv preprint arXiv:1903.10274*.

[23] Lavers, D. A., Waliser, D. E., Ralph, F. M., & Dettinger, M. D. (2016). Predictability of horizontal water vapor transport relative to precipitation: Enhancing situational awareness for forecasting western U.S. extreme precipitation and flooding. *Geophysical Research Letters, 43*(5), 2275-2282.

[24] Liu, Y., Racah, E., Correa, J., Khosrowshahi, A., Lavers, D., Kunkel, K., et al. (2016). Application of deep convolutional neural networks for detecting extreme weather in climate datasets. *arXiv preprint arXiv, 1605*, 01156.

[25] Luis Gimeno, Raquel Nieto, Marta Vázquez, David A. Lavers (2014). Atmospheric rivers: a mini-review. *Frontiers in Earth Science, 2* (2), 1-6.

[26] Martin, A., and Coauthors, 2018: Evaluation of atmospheric river predictions by the WRF model using aircraft and regional mesonet observations of orographic precipitation and Its forcing. *J. Hydrometeor.*, 19, 1097-1113.

[27] McCollor, D., Stull, R. (2008). Hydrometeorological Accuracy Enhancement via Postprocessing of Numerical Weather Forecasts in Complex Terrain. *Weather and Forecasting, 23*(1), 131-144.

[28] McGovern, A., Elmore, K. L., Gagne, D. J., Haupt, S. E., Karstens, C. D., Lagerquist, R., et al. (2017). Using artificial intelligence to improve real-time decision-making for high-impact weather. *Bulletin of the American Meteorological Society, 98*(10), 2073-2090.

[29] Michael A. Nielsen (2015). Neural networks and deep learning. *Determination Press.*

[30] Michael Dettinger (2011). Climate change, atmospheric rivers, and floods in California: a multimodel analysis of storm frequency and mag-

nitude changers. *Journal of the American Water Resources Association, 47* (3), 514-523.

[31] Mitchell, T. M. (1997), *Machine Learning* , McGraw-Hill , New York.

[32] Mohamed Elhoseiny, Sheng Huang, Ahmed Elgammal (2015). Weather classification with deep convolutional neural networks. *2015 IEEE International Conference on Image Processing (ICIP)*, Quebec City, QC, Canada, 3349-3353.

[33] Moore, B. J., , P. J. Neiman, , F. M. Ralph, and F. E. Barthold, 2011: Physical processes associated with heavy flooding rainfall in Nashville, Tennessee, and vicinity during 1?2 May 2010: The role of an atmospheric river and mesoscale convective systems. *Mon. Wea. Rev., 140*, 358-378.

[34] Neiman, P. J., F. M. Ralph, A. B. White, D. E. Kingsmill, and P. O. G. Persson, 2002: The statistical relationship between upslope flow and rainfall in California's Coastal Mountains: Observations during CALJET. *Mon. Wea. Rev., 130*, 1468-1492.

[35] Pablo Rozas Larraondo, Iñaki Inza, Jose Antonio Lozano, (2017). Automating weather forecasts based on convolutional networks. *In Workshop on Deep Structured Prediction ICML 17*, Sydney 11 August 2017.

[36] Peter D. Dueben, Peter Bauer (2018). Challenges and design choices for global weather and climate models based on machine learning. *Geoscientific Model Development, 11* (10), 3999-4009.

[37] Ralph, F. M., Neiman, P. J., Wick, G. A. (2004). Satellite and CALJET Aircraft Observations of Atmospheric Rivers over the Eastern North Pacific Ocean during the Winter of 1997/98. *Monthly Weather Review, 132*(7), 1721-1745.

[38] Ralph, F. M., P. J. Neiman, and R. Rotunno (2005), Dropsonde observations in low-level jets over the northeastern Pacific Ocean from CALJET-

1998 and PACJET-2001: Mean vertical-profile and atmospheric-river characteristics, *Mon. Weather Rev., 133*, 889-910.

[39] Ralph, F. M., P. J. Neiman, G. A. Wick, S. I. Gutman, M. D. Dettinger, D. R. Cayan, and A. B. White (2006), Flooding on California's Russian River: Role of atmospheric rivers, *Geophys. Res. Lett., 33*, L13801.

[40] Rasp, S., Pritchard, M. S., & Gentine, P. (2018). Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences, 115*(39) 9684-9689.

[41] Robert James Firth (2016). A novel recurrent convolutional neural network for ocean and weather forecasting. *LSU Doctoral Dissertation.* 2099.

[42] Roeger, C., Stull, R., McClung, D., Hacker, J., Deng, X., Modzelewski, H. (2003). Verification of Mesoscale Numerical Weather Forecasts in Mountainous Terrain for Application to Avalanche Prediction. *Weather and Forecasting, 18*(6), 1140-1160.

[43] Sebastian. Scher (2018). Toward data-driven weather and climate forecasting: Approximating a simple general circulation model with deep learning. *Geophysical Research Letters, 45* (22), 12616-12622.

[44] Scher, S., & Messori, G. (2018). Predicting weather forecast uncertainty with machine learning. *Quarterly Journal of the Royal Meteorological Society.*

[45] Stensrud, D. J., Skindlov, J. A. (2002). Gridpoint Predictions of High Temperature from a Mesoscale Model. *Weather and Forecasting, 11*(1), 103-110.

[46] Stensrud, D. J., Yussouf, N. (2003). Short-Range Ensemble Predictions of 2-m Temperature and Dewpoint Temperature over New England. *Monthly Weather Review, 131*(10), 2510-2524.

[47] Stephan Rasp, Sebastian Lerch (2018). Neural networks for postprocessing ensemble weather forecasts. *Monthly Weather Review, 146* (11), 3885-3900.

[48] Thornes, T., Düben, P., and Palmer, T.: On the use of scaledependent precision in Earth System modelling, Q. J. Roy. Meteor. Soc., 143, 897-908, 2017.

[49] Waliser, D. E., et al. (2012), The Year of Tropical Convection (May 2008?April 2010): Climate variability and weather highlights, *Bull. Am. Meteorol. Soc., 93*, 1189-1218,

[50] W E. Chapman, A. C. Subramanian, L. Delle Monache, S.P. Xie, F.M. Ralph (2019). Improving atmospheric river forecasts with machine learning. *Geophysical Research Letters, 46* (17-18), 10627-10635.

[51] Wilks, D. S., Hamill, T. M. (2007). Comparison of Ensemble-MOS Methods Using GFS Reforecasts. *Monthly Weather Review, 135*(6), 2379-2390.

[52] Yunjie Liu, Evan Racah, Prabhat, Joaquin Correa, Amir Khosrowshahi, David Lavers, Kenneth Kunkel, Michael Wehner, William Collins (2016). Application of deep convolutional neural networks for detecting extreme weather in climate datasets, arXiv:1605.01156.

[53] Zhu, Y., and R. E. Newell (1998), A proposed algorithm for moisture fluxes from atmospheric rivers, *Mon. Weather Rev., 126*, 725-735.