OXFORD BROOKES UNIVERSITY

Faculty of Technology, Design and Environment

Master of Philosophy

Biotic Pump of atmospheric moisture: A conceptual model

Thomas Lafon

August 2012

Directors of studies
Chris Cox
Dr. Robert Beale

Abstract

The Biotic Pump Theory, as described by Drs. Makarieva and Gorshkov (M&G), defines the mechanism by which water vapour is transported from areas of low evaporation to areas of high evaporation, conveniently termed “donor” and “acceptor” regions, respectively, and where only the latter exhibits condensation. The implications of such a theory are critical, especially to moisture regulation of tropical rainforests, yet highly controversial. Unfortunately, most of the theory’s physics cannot be evaluated due to the lack of atmospheric observations over such areas.

This study aims at building a conceptual model of the theory over the Amazon basin as to quantitatively assess the existence and determine the properties of donor and acceptor regions statistically through their respective condensation rates. The model uses the predictive capabilities of Time-Delayed Neural Networks to downscale available atmospheric observations to calculate condensation rates at a scale suited for this analysis. Validation of the downscaling model reveals monthly Mean Absolute Errors to range between 0.022 m s^{-1} and 2.76 m s^{-1} in the predictions of vertical velocity and zonal wind speed, respectively.

Findings quantitatively support the existence of a biotic mechanism regulating the transport of water vapour as these clearly show the presence of donor and acceptor regions. These regions have average spatial distributions of 42% - 58%, respectively, over the whole study

area and correlations are found between wind speeds and condensation rates. Mean annual condensation rate for the Amazon basin is calculated to be of 0.23E-06 mol m^{-3} s^{-1}. Results also show an increase in average condensation rate (0.06E-06 mol m^{-3} s^{-1}) for the last 9 years, which does not strictly adhere to M&G's views on the impacts of deforestation on precipitation. Outcomes hence also suggest a more complex relationship between evaporation and condensation, and therefore highlight the necessity to further refine this novel theory.

Acknowledgments

My interest in studying the Biotic Pump Theory and its impact on rainforests sprung from an unexpected series of events. Thanks to the work placement programmes offered by Oxford Brookes University I have had the chance to work for the Centre of Ecology and Hydrology (CEH) in Wallingford, where I got introduced to climatology and climate modelling. Having finished my Bachelors in Software Engineering at Brookes, I worked in CEH for another year. During this time, I met Angela Maldonado, my beloved partner and founder of the Colombian NGO Fundación Entropika. Entropika's work focusses in community-based research for the conservation of wildlife in the Colombian Amazon. Thanks to her, I was given the opportunity to work with Entropika and the IDEAM (Colombian MetOffice) in the Amazon basin and learn a great deal on local issues related to climate change. Through my work with Entropika, I met Peter Bunyard, Scientific Editor of The Ecologist magazine, who introduced me to the novel theory of Biotic Pump. This is when I decided to combine my skills in computing and my knowledge on climate models to do a Master of Philosophy which aims at quantitatively assessing the reality of such a mechanism as described in the Biotic Pump Theory. Over the course of my thesis, I was lucky to meet and have worked with people from a wide variety of scientific and cultural backgrounds.

Firstly, I would like to deeply thank Jennifer Fowler, coordinator of the physics laboratory at the University of Montana, who has been my external supervisor for this thesis and with whom I have now worked two years consecutively in the search for viable alternatives for

conducting daily soundings in the Amazon basin. Jennifer has been for me a source of motivation and a great help in successfully completing this thesis. I would also like to sincerely thank Peter Bunyard and his enthusiasm, which have been, and remain, a source of inspiration and motivation to me. I wish to thank both of my internal supervisors, Chris Cox and Dr. Robert Beale, as well as Jill Organ, for having been very helpful and understanding of my situation owing to my ties with the Colombian Amazon. They have always been there when I needed them and have guided me with valuable advice, comments and thoughts throughout this process. Furthermore, I would like to convey my outmost gratitude to Drs. Christel Prudhomme and Simon Dadson, my line managers at CEH, who have given me the opportunity to learn about climatology, be actively part of their working team and freely voice my opinion in our line of work. Also, I would like to thank Helen Davies for being so kind to me during the two years I was working there. Additional thanks go to Jaime Tamayo and Fredy Jiménez from the IDEAM in Leticia for their collaboration in conducting soundings in the area, to Antonio Garcia-Mendez from the European Centre for Medium-Range Weather Forecasts for providing the global list of operational sounding station and to Rich Ellis from CEH for his reviews of the first few drafts of this thesis.

I would also like to profoundly thank my dad, Stéphane Lafon, mother, Véronique Lafon, sister, Camille Lafon, family and close friends for their understanding, love and constant support. A very special thanks and much love goes to my beautiful partner, Angela Maldonado, who has always been there for me with her love and guidance.

To conclude, I wish to thank everybody who has supported me throughout the thesis.

Table of Contents

List of acronyms and symbols

%	Per cent
∂	Rate of change
∇	Rate of change along a three-dimensional space
°	Degrees
°C	Degrees Celsius
α	Angle between isothermal and horizontal planes
γ	Molar density of water vapour
air	Air temperature from the NCEP dataset
air.sfc	Surface air temperature from the NCEP dataset
ANN	Artificial Neural Network
BADC	British Atmospheric Data Centre
BioPCM	Biotic Pump Conceptual Model
BPT	Biotic Pump Theory
CGIAR	Consultative Group on International Agricultural Research
COA	Certificate Of Authorization
CRU	Climatological Research Unit
CSI	Consortium for Spatial Information
DS	Downscaled
ESRL	Earth System Research Laboratory
$f_{\uparrow}$	Upward-directed pressure gradient force
$f_{\downarrow}$	Downward-directed weight of water vapour
f_E	Evaporative-condensational force
FAA	Federal Aviation Administration
GCM	Global Climate Model
GCOS	Global Climate Observing System
GPS	Geographical Positioning System
GRUAN	GCOS Reference Upper-Air Network
GUAN	GCOS Upper-Air Network
H	Hidden layer nodes
hgt	Height from the NCEP dataset

hgt.sfc Surface height from the NCEP dataset

IDEAM Instituto de Hidrología, Meteorología y Estudios Ambientales

ITCZ Intertropical Convergence Zone

J Joules

K Kelvins

km Kilometre

m Metre

M&G Makarieva and Gorshkov

MAE Mean Absolute Error

mbar Millibars

MLP Multi-Layer Perceptron

mol Molecules

MSE Mean Squared Error

$\boldsymbol{N}$ Molar density of moist air

$\boldsymbol{N_v}$ Molar density of water vapour

NASA National Aeronautics and Space Administration

NCAR National Centre for Atmospheric Research

NCEP National Centres for Environmental Prediction

NERC Natural Environment Research Council

net_p TDNN for the prediction/downscaling of surface/atmospheric pressure

net_t TDNN for the prediction/downscaling of air temperature

net_u TDNN for the prediction/downscaling of zonal wind speed

net_v TDNN for the prediction/downscaling of water vapour pressure

net_w TDNN for the prediction/downscaling of vertical velocity

NOAA National Oceanic and Atmospheric Administration

OAR Office of Oceanic and Atmospheric Research

omega Vertical velocity from the NCEP dataset

$\boldsymbol{p}$ Atmospheric pressure

$\boldsymbol{p_{sat}}$ Saturation pressure of water vapour

pres.sfc Surface atmospheric pressure from the NCEP dataset

PSD Physical Science Division

$\boldsymbol{p_{stm}}$ Steam-point pressure

$\boldsymbol{p_v}$ Water vapour pressure

$\boldsymbol{R}$	Universal gas constant
RCM	Regional Climate Model
RH	Relative humidity
rhum	Relative humidity from the NCEP dataset
rhum.sfc	Surface relative humidity from the NCEP dataset
s	Seconds
$\boldsymbol{S}$	Molar rate of condensation
SRTM	Shuttle Radar Topographic Mission
t	Month
TD	Time delay memory
TDNN	Time-Delayed Neural Networks
TDNN-A	Ensemble of TDNNs for the downscaling of atmospheric variables
TDNN-S	Ensemble of TDNNs for the downscaling of surface variables
tmp	Air temperature from the CRU dataset
$\boldsymbol{T}$	Air temperature
$\boldsymbol{T_{stm}}$	Steam-point temperature
TS	Timeseries
$\boldsymbol{u}$	Zonal wind speed
UAS	Unmanned Aircraft System
UK	United Kingdom
US	United States
uwnd	Zonal wind speed from the NCEP dataset
vap	Water vapour pressure from the CRU dataset
w	Weighted connections
$\boldsymbol{w}$	Vertical velocity
$\boldsymbol{w_0}$	Velocity vector along the temperature gradient
WMO	World Meteorological Organization
x	Longitude
X	Input nodes
y	Latitude
Y	Output nodes
z	Altitude

Chapter 1

Introduction

1.1 The importance of the Amazon basin

Water is essential for life and its properties enable it to play a critical role in the processes of climate and make this planet habitable. In addition to dissolving minerals, transporting them to the oceans and lubricating the sliding of one tectonic plate over another, water carries energy around the globe, distributing it as part of a planetary conveyor belt system. While in the ocean it is readily available, most of the water found on land comes from large-scale migration of water vapour transported by winds (Gimeno *et al.*, 2010; Penman, 1963). During that passage a water molecule, whether in the form of ice, liquid or vapour may find itself pushed upwards into the atmosphere, during which time it may form clouds (depending on *cloud condensation nuclei*), and later bring about precipitation through means of condensation. Air can hold a certain quantity of water vapour, depending on temperature and therefore on the saturation curve of water vapour. Consequently, the position of air in the atmosphere, which latitude it is in, and how much is exposed to direct sunlight, has profound consequences for local weather and, more generally, climate.

To sustain itself, vegetation requires sufficient soil moisture and subsequently precipitation. Between 50% and 75% of the rainfall over the Amazon rainforest is derived from its own *evapotranspiration* (Lettau *et al.*, 1979; Marengo, 2006; Salati, 1987; Salati and Vose, 1984); this process is known as moisture recycling. Extensive deforestation will inevitably lead to a reduction in evapotranspiration which, according to Hutyra *et al*. (2005), is likely to result in *savannization*. Moreover, deforestation and *edge effects* in the forest surrounding large scale clearings increase the likelihood of destructive fires and forest die-back, as was indeed the case during the 2005 and 2010 Amazon droughts (Cochrane and Laurance, 2002; Hutyra *et al.*, 2005; Laurance and Williamson, 2001; Lewis *et al.*, 2011; Phillips *et al.*, 2009).

Through the use of climate models (i.e. *Global* and *Regional Climate Models*; GCMs and RCMs, respectively) a number of scientists have predicted the effects of tropical rainforest removal on precipitation (e.g. Clark *et al.*, 2001; Henderson-Sellers *et al.*, 1993; Marengo, 2006; Werth and Avissar, 2004). Other than resulting in a predictable reduction in evapotranspiration, an outcome of such research suggests an apparent connection between the deforestation of the Amazon Rainforest and statistically-significant changes in precipitation at other tropical and mid-latitude regions (Avissar and Werth, 2005; Werth and Avissar, 2002). These "teleconnections" have been determined to affect parts of North America, Europe and South Africa (see Figure 1.1). The *South American Low Level Jet Stream* is also known to transport a considerable amount of water vapour from the Amazon Basin to the River Plate Basin, where as much as 50% of the Plate basin's rainfall is derived from the Amazonian teleconnection (Gimeno *et al.*, 2010; Marengo, 2006).

Figure 1.1. *Teleconnections of the Amazon basin. Illustration adapted from Marengo (2006).*

The Amazon Rainforest is hence considered to play a vital role in maintaining the fragile equilibrium of global climate and its deforestation may therefore have a profound impact on local and global precipitation patterns. Figure 1.2 illustrates past and future deforestation extent over the Amazon basin as predicted by Soares-Filho *et al.* (2006).

Makarieva and Gorshkov (2007; 2009a; 2009b; 2010a; 2010b) and Makarieva *et al.* (2009; 2012; 2010) believe that, although GCMs take deforestation of tropical rainforests into account, they do not encapsulate all of the observed dynamics of rainforests and thus fail in producing realistic predictions. Subsequently, they reformulate some of the physics of

atmospheric moisture fluxes in order to fit observations. The “Biotic Pump” opens a new dimension in the understanding of the hydrological cycle and our climate.

Figure 1.2. *Forest cover for scenarios assuming (a) that recent deforestation trends will continue and (b) the Brazilian environmental legislation is implemented through the refinement and multiplication of current experiments in frontier governance. Illustration from Soares-Filho (2006).*

The Biotic Pump Theory (BPT) states, against general belief, that the major physical cause of moisture fluxes is not the non-uniformity of atmospheric and surface heating but the fact that water vapour is invariably upward-directed since its *partial pressure* is not compensated by its weight in the atmospheric column due to condensation. The physical meaning of this behaviour results in densely vegetated forests “sucking-in” atmospheric moisture from the ocean; water vapour travels from regions of low evaporation and no condensation (donor regions, i.e. ocean) to regions of high evaporation and high condensation (acceptor regions, i.e. land). Extensive deforestation will hence lead to a tipping point where the forest cover is too little to exercise this mechanism, engendering changes in local and global water vapour regime.

Although the importance of the Amazon rainforest has been emphasized in an extensive collection of publications, the interaction of its vegetation on precipitation and global water vapour circulation is still poorly understood (e.g. Avissar and Werth, 2005; Makarieva *et al.*, 2009; Meesters *et al.*, 2009; Myneni *et al.*, 2007). Marengo (2006) attributes this to the "lack of basic hydrometeorological observational data systematically collected over time and space". Not only the lack, but also the coarseness of available data over the Amazon basin makes evaluating the existence of the Biotic Pump mechanisms difficult. Available datasets covering the extent of the Amazon basin (mostly interpolated from local stations measurements) have horizontal resolution in the order of several hundred kilometres. At this scale, much of the subgrid-scale features and dynamics, such as topographical features, *convective* processes and hydrological processes, are lost (e.g. Carter *et al.*, 1994; Coulibaly *et al.*, 2005; Hewitson and Crane, 1996; Wigley *et al.*, 1990).

1.2 Overall aim and specific objectives

The overall aim of this research is to assess the existence of the Biotic Pump mechanism as set forward by Makarieva and Gorshkov (2007; 2009a; 2009b; 2010a; 2010b; Makarieva *et al.*, 2009; 2012; 2010) as this theory could potentially provide us with a more profound insight in understanding tropical rainforests and help us with more realistic predictions of deforestation-induced impact on the climate; hence giving us the chance to successfully preserve this fragile ecosystem. It is here assumed that the mechanism of the Biotic Pump, if it exists, will be visible in the observed data of the Amazon basin.

This is here attempted by elaborating a conceptual model of the BPT, the Biotic Pump Conceptual Model (BioPCM). To tackle the issue of data availability, BioPCM makes use of the predictive capabilities of *Time-Delayed Neural Networks* to downscale, from surface conditions, observed *tropospheric* profiles of the variables important to the working of the Biotic Pump. The model then calculates condensation rates from the physics of the BPT. Finally, outputs of the model are analysed in order to identify donor and acceptor regions from their particular properties relating to condensation and, as such, assess the presence of a biotically-regulated mechanism as described by Drs. Makarieva and Gorshkov.

The specific objectives of this research are to:

A. Downscale available observed data of tropospheric profiles of air temperature, *zonal wind* speed, vertical velocity, atmospheric pressure and water vapour pressure using Time-Delayed Neural Networks.

B. Validate the downscaled tropospheric profiles of air temperature, zonal wind speed, vertical velocity, atmospheric pressure and water vapour pressure over the Amazon basin.

C. Calculate and map monthly means of condensation rates in order to identify donor and acceptor regions of the study site and, in this way, assess the presence of a mechanism driven by the Biotic Pump.

D. Formulate conclusions, based on the results, on the possibility of the existence of such a mechanism and propose future developments to ascertain the validity of the proposed conclusions.

BioPCM is conceptual in as such that it is used only for the evaluation of the existence of the Biotic Pump mechanism. If satisfactory, the resulting model will engender the prospect of being implemented in predictive models where simulation of tropical rainforests is required or being assessed.

1.3 Study site and data

The site of interest for this study is the Amazon basin. In order to simplify calculations and data analysis, I focus on a rectangular window located in the centre of the Amazon basin (see Figure 1.3). The latitude of the window spans from 1.25 °N to 12.5 °E and the longitude from 53.75 °W to 73.75 °W, covering a total area of approximately 3709243 km^2. Surface height ranges from -15 m to 5924 m. The highest point is in the Andean mountain range located at the bottom-left corner of the window, and the lowest point is found in the top-right corner close to the mouth of the Amazon River.

Three datasets are used in this study: the NCEP/NCAR reanalysis dataset, the CRU TS-3.1 surface observations and the SRTM 1-km digital elevation data. The NCEP/NCAR dataset is chosen for its global high-resolution coverage of atmospheric variables necessary to conducting this study. The CRU data are selected for their global coverage of basic surface observations at very high-resolution and SRTM 1-km digital elevation data are here used to complement the CRU observations. The main advantages of using these datasets are: a) all three sets are widely-used and accepted by the scientific community, b) NCEP/NCAR and CRU datasets are continuously being updated, and c) these are freely available online.

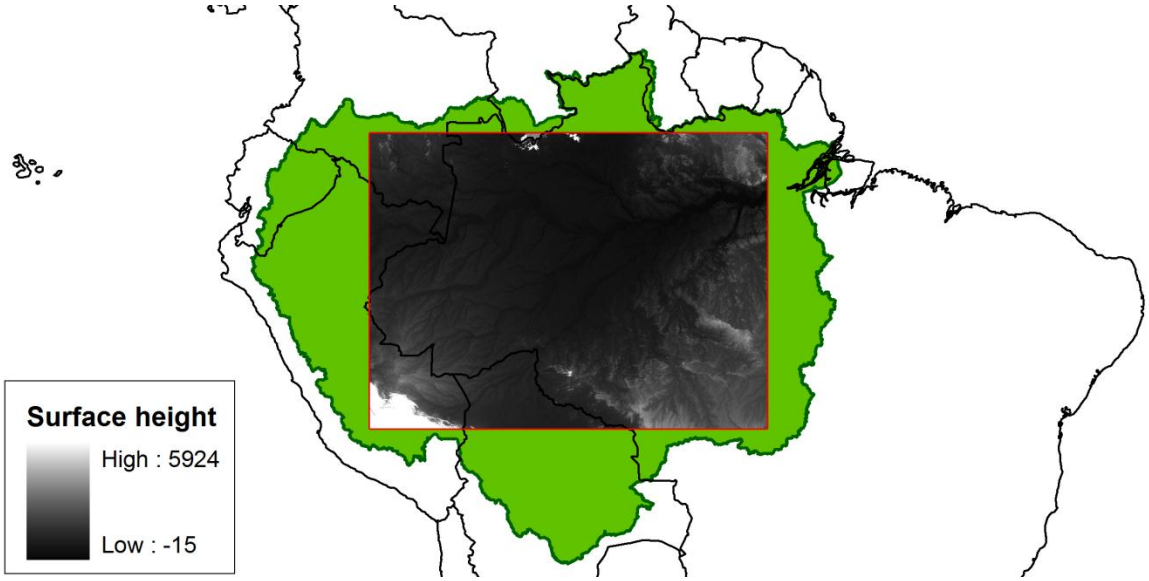


Figure 1.3. *Window delimiting the study area, where the Amazon basin is represented as the green area.*

1.3.1 NCEP/NCAR reanalysis dataset

The NCEP/NCAR reanalysis dataset (Kalnay *et al.*, 1996) was developed by the National Centres for Environmental Prediction and the National Centre for Atmospheric Research, Washington DC, USA. The aim of the reanalysis project is to provide a record of global analyses of atmospheric fields in order to address the needs of the research and climate monitoring community. In doing so, the NCEP and NCAR have created a dataset that covers surface and atmospheric variables globally. The dataset was generated by recovering available land surface, ship, *radiosonde*, aircraft, satellite and other data, controlling the quality of the collected data, and finally, assimilating the collected data. The dataset is composed out of monthly averages of surface and atmospheric variables at a horizontal resolution of 2.5° (approximately 278 km), for 17 vertical pressure levels, ranging from 1000 mbar to 10 mbar, and at Earth's surface. The data covers a time period of 64 years, from 1948 to 2011.

In this study, I use forty years of monthly averages of current climate data (1971 – 2010) from the NCEP/NCAR reanalysis dataset. The first 30 years (1971 – 2000) are used to train the Time-Delayed Neural Networks, which are then validated on the remaining 10 years (2001 – 2010). In total, five atmospheric variables and three surface variables from the NCEP/NCAR reanalysis dataset are here used (see Table 1.1).

Level	Variable	NCEP/NCAR acronym	Unit
Atmosphere	Air temperature	air	°C
	Height	hgt	m
	Vertical velocity	omega	m s^{-1}
	Relative humidity	rhum	%
	Zonal wind speed	uwnd	m s^{-1}
	Atmospheric pressure*	N/A	mbar
	Water vapour pressure**	N/A	mbar
Surface	Air temperature	air.sfc	°C
	Surface height	hgt.sfc	m
	Relative humidity	rhum.sfc	%
	Atmospheric pressure	pres.sfc	mbar

Table 1.1. *NCEP/NCAR reanalysis data variables used in the study.*

** Calculated from height.*

*** Calculated from relative humidity and temperature (see Appendix I).*

Note that atmospheric variables are converted from pressure levels to altitude using the height data and a simple interpolation method. Pressure for a specific altitude is then determined from the height data and water vapour pressure is calculated from relative humidity and air temperature data (see Appendix I for a detailed description of the calculation of water vapour pressure). The set of NCEP/NCAR reanalysis data variables listed in Table 1.1 are hereafter referred to as the "NCEP dataset". The NCEP/NCAR reanalysis dataset is provided by the NOAA/OAR/ESRL PSD, Colorado, USA, from their Web site at http://www.esrl.noaa.gov/psd/.

1.3.2 CRU TS-3.1 surface observations

The CRU TS-3.1 time series (Mitchell and Jones, 2005) was developed by the Climate Research Unit of the University of East Anglia, UK, in an attempt to construct a global high-resolution dataset of homogenised monthly surface observations that will address the needs of climatologists. The dataset was created by collecting available observations from surface stations, correcting for station data inhomogeneity, and interpolating the data onto a 0.5° (approximately 56 km) grid at a global level. The period covered by the data is of 109 years, from 1901 to 2009.

I here use 39 years of CRU TS-3.1 surface observations (1971 – 2009) as input to the NCEP-trained Time-Delayed Neural Networks. By doing so, the neural networks predict the atmospheric variables from the CRU TS-3.1 surface data and, as such, downscale the data. Four surface variables of the CRU TS-3.1 time series are used (see Table 1.2).

Level	Variable	CRU TS-3.1 acronym	Unit
Surface	Air temperature	tmp	°C
	Water vapour pressure	vap	mbar
	Relative humidity*	N/A	%
	Atmospheric pressure**	N/A	mbar
	Height***	N/A	m

Table 1.2. *CRU TS-3.1 data variables used in the study.*

** Calculated from vapour pressure and temperature (see Appendix I).*

*** Calculated using Time-Delayed Neural Networks (see Chapter 3).*

**** Taken from the SRTM 1-km elevation data, upscaled to fit the CRU resolution (see Heading 1.3.3).*

Note that *relative humidity* is here calculated from water vapour pressure and air temperature data (see Appendix I for a detailed description of this calculation). Pressure has

to be calculated using the Time-Delayed Neural Networks (procedure similar to that of the downscaling of the atmospheric variables from surface variables; see Chapter 3), and height is taken from the SRTM 1-km elevation data, upscaled to fit the CRU TS-3.1 resolution (see Chapter 1, Heading 1.3.3). Likewise to that of NCEP, it is necessary to complete the CRU TS-3.1 dataset as these variables are necessary for the downscaling procedure. The set of CRU TS-3.1 data variables listed in Table 1.2 are hereafter referred to as the "CRU dataset". The CRU TS-3.1 surface observations are provided by the BADC/NERC, Didcot, UK, from their Web site at http://badc.nerc.ac.uk/.

1.3.3 SRTM 1-km elevation dataset

The SRTM 1-km digital elevation data (Jarvis *et al.*, 2008) was developed by NASA's (National Aeronautics and Space Administration, USA) Shuttle Radar Topographic Mission. The objective of the mission is to provide high-resolution surface elevation data in order to promote the use of geospatial science and applications for sustainable development and resource conservation in the developing world. The 1-km data is a resampled version of the SRTM 90 m digital elevation database version 4.1, released in 2008. For this study, I upscale the 1-km data further to 0.5°, as to fit the CRU dataset resolution, using a simple gridbox-average algorithm. Note that all further references to the CRU dataset include the SRTM 1-km elevation data upscaled to fit the CRU dataset resolution. The SRTM 1-km elevation dataset is provided by the CGIAR/CSI from their Web site at http://srtm.csi.cgiar.org/.

1.4 Structure of thesis

The thesis is divided as follows. Chapter 2 describes the theory of the Biotic Pump as well as the physics used for calculating condensation rates; it is the main literature review of this work. Chapter 3 outlines the elaboration and application of the ensembles of Time-Delayed Neural Networks used for the downscaling of the variables necessary for the calculation of condensation rates. This is followed in Chapter 4 by the calculation and mapping of condensation rates, and the identification of donor and acceptor regions. The thesis is finalized by a discussion and conclusion in Chapter 5.

By quantitatively assessing the existence of a biotically-driven mechanism as described by Drs. Makarieva and Gorshkov, I here intend to contribute to the body of knowledge on atmospheric physics with a particular focus on condensation-induced fluxes. The results and conclusions presented in Chapters 4 and 5, respectively, constitute my contribution to these topics. These results and conclusion have to be considered with care due to the controversial nature of the proposed theory.

Chapter 2

Theory of the Biotic Pump

In this chapter I give a brief overview of the history of the research on the Biotic Pump, as well as of the physical mechanism around which the theory revolves and the main arguments of the opposition to the existence of such a mechanism. This is then followed by a detailed explanation of the condensation-induced mechanism that causes winds, the transport of water vapour, and the presence of so-called “donor” and “acceptor” regions. I also discuss the local and global implications of such a mechanism with regards to deforestation and water regulation, and its possible impact on the wider scientific community. Finally, I conclude the chapter by outlining the physical formulas of condensation rate, which are then later used in Chapter 4 as part of the assessment of the Biotic Pump Theory (BPT) with regards to its condensation-induced dynamics.

2.1 Brief overview

The Biotic Pump Theory was developed by Drs. Anastassia Makarieva and Victor Gorshkov (hereafter referred to as “M&G”), scientists of the Petersburg Nuclear Physics Institute in St.

Petersburg, Russia. The theory was first published in 2007 (Makarieva and Gorshkov, 2007) and subsequently refined in their 2009, 2010 and 2012 publications (Makarieva and Gorshkov, 2009a; 2009b; 2010a; 2010b; Makarieva *et al.*, 2009; 2012; 2010). The scientific community has been increasingly divided about the findings of M&G, the main criticism being that the theory is not supported by basic physical principles (Meesters *et al.*, 2009). However, M&G have been able to defend their research and have taken critics and comments on board, as to leave no doubt that the Biotic Pump is indeed a reality. This on-going controversy has asked for an objective evaluation of the BPT, part of which will be attempted in this thesis.

In short, the theory describes, through the use of fundamental physics, the mechanism of densely vegetated closed-canopy forests as biotic pumps that attract moisture from the oceans to the land. In this way, the forest keeps the supply of moisture to land constant as to sustain its vegetation and, indirectly, regulate global circulation of water vapour. The novelty of the Biotic Pump mechanism lies in the fact that M&G physically derive vertical and horizontal pressure changes induced by condensation to be the primary factor driving winds. These winds transport water vapour from regions of high pressure (donor – low evaporation, no condensation) to low pressure (acceptor – high evaporation, high condensation). This differs from conventional knowledge that states that the major cause of pressure gradients (and hence winds) is due to uneven heating of the surface and atmosphere. In their publications, M&G claim that the scientific community has overlooked this mechanism. This point of view was first expressed by Lorenz (1967; 1983) and is supported by, for example, Schneider (2006), Sheil and Murdiyarso (2009).

2.2 Condensation-induced circulation of moisture

The atmosphere is composed out of two types of gases: condensable gases, such as water vapour, and non-condensable gases, such as Nitrogen, Oxygen, Argon, etc. The air contained in the atmosphere is said to be “dry” when no water vapour is present in its composition, and “wet” when it holds water vapour. In a static atmosphere, water on land will evaporate owing to the incoming energy of solar radiation and mix in with the air of the lower altitudes of the troposphere (i.e. lower part of the atmosphere where most hydrological atmospheric processes take place). Any upward displacement of this wet air will cause the water vapour to condense (when reaching saturation point) as a result of the decrease in temperature with altitude. The condensation of the water will remove the excess water vapour from the atmospheric column (i.e. through precipitation) and hence reduce the total pressure of the column.

Two forces act upon water vapour: an upward-directed force, resulting from the vertical compression of the water vapour, and a downward-directed force, generated by the weight of the water vapour in the air column. In removing water vapour through condensation, the force resulting from the vertical compression of water vapour in the atmospheric column will be greater than the weight of the remaining water vapour. This will result in an invariably upward-directed force that acts equally on air parcels with positive and negative *buoyancy* (see Figure 2.1). The force is a result of both processes of evaporation and condensation and is subsequently termed “evaporative-condensational force” by M&G (2007; 2010a).

The evaporative-condensational force is dependent on the local concentration of water vapour which, by definition, makes it dependent on evaporation rates; the higher the evaporation rate, the more water vapour is present in the atmosphere, the more condensation will take place, and hence the stronger the resulting force (assuming that enough soil moisture is available to maintain evaporation). In order to exist, the evaporative-condensational force hence requires continuous evaporation that will compensate for the loss of moisture from condensation of the ascending water vapour molecules.

Figure 2.1. *The upward-directed evaporative-condensational force (f_E) equal to the difference between the upward-directed pressure gradient force ($f_\uparrow$) and downward-directed weight ($f_\downarrow$) of water vapour. Illustration from M&G (2007).*

Owing to the removal of water vapour through condensation, inhomogeneities in the rates of condensation between two regions will engender differences in the respective atmospheric pressures of these areas. Under Newton's second law of motion, the resulting pressure gradient will force air to flow from areas of high pressure to areas of low pressure,

conveniently termed “donor” and “acceptor” regions by M&G, respectively. This defines the condensation that occurs during the ascension of water vapour as being responsible for, both, vertical and horizontal pressure gradients, and hence its circulation.

According to the BPT, evaporation occurs in both, donor and acceptor region, while condensation takes place only in the acceptor region. This implies that the amount of precipitation in the acceptor region is always higher than local evaporation and that, both, acceptor and donor regions can be identified through their respective rates of condensation. With increasing evaporation, the condensation rate grows more rapidly than that of evaporation as these are not directly related (condensation rate is dependent on vertical velocity). Therefore, the acceptor region has always higher evaporation rates than the donor region. Following this logic, one can see that evapotranspiration of the Amazon basin, with its high *Leaf Area Index*, exceeds evaporation over the ocean with the net result that the forest draws in moist air from the ocean to the land (Fig. 2.2).

Figure 2.2. *The physical principle that the low-level air moves from areas with weak evaporation to areas with more intensive evaporation, where black arrows represent evaporation flux (width schematically indicates the magnitude of this flux) and empty arrows are the horizontal and ascending fluxes of moisture-laden air in the lower atmosphere. Illustration from M&G (2007).*

It is important to note that, whether induced by temperature changes or the evaporative-condensational force, other conditions can also bring about the rising and condensation of air parcels. For example, moving air will be automatically lifted along a mountain side and forced to condense as it ascends. Such mechanism dictated by topography is currently not included in the conceptual model and will hence not be reflected in the results over the relatively little Andean mountain region of the study site.

2.3 Implications for Earth's climate and research

M&G (2007) state that "in the absence of biotic control, air fluxes transporting ocean-evaporated moisture to the continents weaken exponentially as they propagate inland" (see Figure 2.3). This has been proven using empirically-established data, showing that the dampening of such fluxes is of the order of several hundred kilometres; hence much less than the linear dimension of the continents. Conventional models predict up to a 30% of reduction in local precipitation associated with continental-scale deforestation (Bonan, 2008), while M&G predict that even smaller-scale deforestation will lead to more than a 95% reduction in local precipitation over the deep interior of the continents (Makarieva and Gorshkov, 2007).

The logic behind this statement is that deforestation will lead the forest to a tipping point at which the vegetative cover will be too little to attract enough water vapour to sustain the basin's vegetation in the deep interior (i.e. > 2000 km from the ocean-source). This will lead to a rapid decrease in forest cover, further reducing forest evapotranspiration and acting as

a positive feedback to the dying of the forest vegetation. Ultimately, this will result in the savannization or *desertification* of the Amazon basin and engender dramatic changes in local and global precipitation patterns. The main point to retain from this statement is that, if the BPT is found to be true, total loss of the Amazonian rainforest due to deforestation will happen much quicker than currently believed by the scientific community. This has been another source of criticism by the opposition qualifying this theory as "appealing to conservationists" (Meesters *et al.*, 2009).

Figure 2.3. *Impact of deforestation predicted by conventional climate models against the BPT. Data from M&G (2007).*

Other than underlining the lack of knowledge on atmospheric dynamics with regards to water vapour fluxes, the theory opens new lines of investigations for a variety of fields ranging from the management of water resources, risk assessments of forest fires to

paleoclimatology. For example, if found to be true, the theory urges the re-assessment of current deforestation impact scenarios in order to enable decision-makers to efficiently preserve tropical rainforests. Furthermore, the BPT could potentially provide new information and explanations to old controversies such as the birth and evolution of continental vegetation and, as such, of human existence. All of which reinforces the argument for the necessity to evaluate objectively this emerging theory.

2.4 Physics of condensation

In hydrology, condensation is defined as the process by which water changes from its gaseous state to its liquid form. Water vapour contained in an air parcel will expand and cool as it ascends in the atmosphere; it will remain unsaturated as long as its temperature does not reach the temperature at which water starts condensing, known as the *dew-point* temperature. When the dew-point is reached, the air parcel's relative humidity, defined as the amount of water found inside a mixture of air relative to the maximum amount that mixture can hold, becomes 100% and the water vapour starts condensing (assuming that the conditions of the saturated water vapour remain constant and that sufficient cloud condensation nuclei are available). As a consequence, a cloud forms until condensation is brought to an end.

The process of condensation constitute an important part of the hydrological cycle as all water evaporated from Earth's surface precipitates back to the surface by means of condensation. For the most part, evaporation comes from extensive water bodies (i.e.

oceans and seas), having a direct impact on the large-scale migration of moisture from the oceans to the continents, and hence playing a critical role in the global distribution of water vapour. Note that there are no major differences in the physics of evaporation from different surface types, only differs the way in which the surface controls the amount of moisture that is being evaporated (Shuttleworth, 1993).

M&G (2010a) state that "the process of condensation should be reflected in the changes of […] relative molar density of water vapour". Consequently, molar rate S (mol m^{-3} s^{-1}) of condensation per unit volume can be calculated using:

$$S = w_0 \nabla N_v - \gamma w \nabla N + \gamma \left(\frac{N}{T}\right) u \nabla T \quad [1]$$

, with

$$p = NRT,\ p_v = N_v RT \quad [2]$$

,

$$\gamma = \frac{p_v}{p} \quad [3]$$

,

$$\nabla = \frac{\partial}{\partial x} i + \frac{\partial}{\partial y} j + \frac{\partial}{\partial z} k \quad [4]$$

, and

$$w_0 = w \cos \alpha - u \sin \alpha \quad [5]$$

where w_0 (m s^{-1}) is the velocity vector directed along the temperature gradient towards the lower temperature, w (m s^{-1}) is vertical velocity along the z-axis and u (m s^{-1}) is horizontal velocity along the x-axis (i.e. zonal wind speed; see Figure 2.4). N and N_v (mol m^{-3}) are molar density of moist air as a whole and of water vapour, respectively, p and p_v (mbar) are

atmospheric and water vapour pressure, respectively, γ is relative molar density of water vapour, T (K) is air temperature, R (J K^{-1}) is the universal gas constant (8.31447215 J K^{-1}), α is the angle between the *isothermal* and the horizontal planes, and i, j, k are the unit orthogonal vectors specifying the directions of coordinate axes x, y and z, respectively. Terms $\gamma w \nabla N$ and $-\gamma(N/T)u\nabla T$ in Equation 1 are both subtracted from the total density change ($w_0 \nabla N_v$) as these represent the equilibrium gravitational expansion and density change of water vapour, respectively, which are unrelated to condensation. Equation 2 is the equation of state of moist air as a whole and of water vapour, respectively, and Equation 5 relates the magnitudes of velocities u, w and w_0.

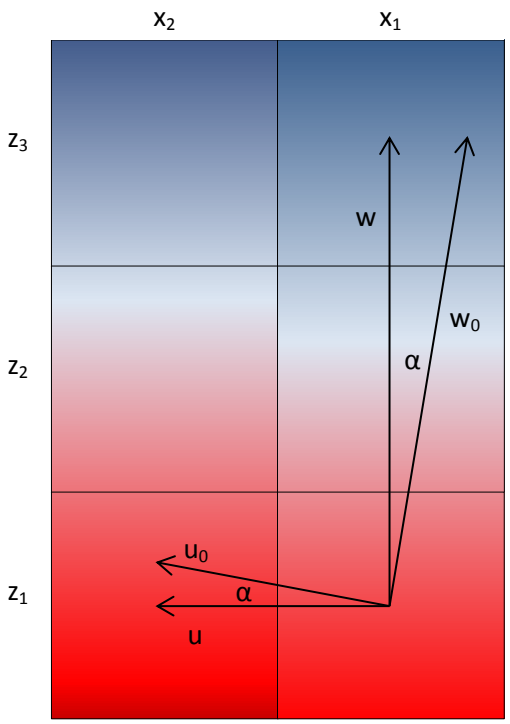


Figure 2.4. *Schematic representations of zonal winds (*u *and* u_0*) and vertical velocities (w and* w_0*), where* u_0 *is the horizontal velocity in the isothermal plane, perpendicular to the temperature gradient. Warmer temperatures are represented in red and colder temperatures in blue.*

In this thesis, Equations 1 to 5 are used to calculate condensation rates in order to distinguish between donor regions and acceptor regions through their respective condensative properties (recall that condensation is found only in the acceptor region). By doing this, one can assess whether this aspect of the BPT is reflected in the observed data, as well as analyse the spatial and temporal properties of the regions, and hence support or not the existence of such a mechanism. However, to use Equations 1 to 5 at a scale suitable for this study, it is necessary to downscale the atmospheric profiles of the five input variables using Time-Delayed Neural Networks:

- Air temperature - T
- Atmospheric pressure - p
- Water vapour pressure - p_v
- Zonal wind speed - u
- Vertical velocity - w

The construction of the neural network ensembles used in the downscaling of these five atmospheric variables and the implementation of the procedure are outlined in the next chapter.

Chapter 3

Downscaling observations

I start this chapter off by exposing the issue of data availability and the restrictions set on research by such limitations. This is then followed by the description of the workings and applications of a procedure that makes use of *Artificial Neural Networks* (ANNs) to downscale available observations in order to tackle, to a significant extend, the issue of data inconsistency and coarseness. It is important to note that the downscaling of the atmospheric variables is achieved using only widely-available surface variables. The specific type of neural network used in this study is an ANN dependent on its previous state, also known as a Time-Delayed Neural Network (TDNN). I also describe and justify the choice of architecture, and detail the training and validation of two TDNN ensembles. The first ensemble, "TDNN-A" (where "A" stands for "atmosphere"), is used to downscale the atmospheric profiles of the five input variables to the equations for calculating condensation rate. The second neural network ensemble, "TDNN-S" (where "S" stands for "surface"), is used to predict and downscale surface relative humidity only, as this variable is not available from the observed surface dataset but necessary to the downscaling procedure of TDNN-A. The chapter is concluded by the explanation and implementation of the downscaling procedure.

3.1 The issue of data availability

It is here assumed that donor and acceptor regions can be identified through their respective properties based on condensation rates. In order to calculate this variable with a certain degree of confidence, it is imperative to have both ground data and upper-air observations. Unfortunately, data available over the Amazon basin is scarce and even more so for atmospheric data (see Appendix II for a detailed review of current state of upper-air observations over the Amazon basin).

While accuracy and coverage of satellite imagery have improved markedly in recent years, radiosonde-collected data still provides us with the most detailed measurements of the troposphere (i.e. lower part of the atmosphere where most hydrological atmospheric processes take place) on account of their fine vertical resolution (Gettleman *et al.*, 2011; Kivi *et al.*, 2009). However, biases in data (temperature, relative humidity, pressure, wind direction and wind speed) are compounded in the tropics by relying on such sparse, historic and continued, sampling of important variables (Randel and Wu, 2006; Sherwood *et al.*, 2005). Figure 3.1 shows the locations of operational sounding stations that form the upper-air network (GUAN) of the Global Climate Observing System (GCOS) as recorded by the World Meteorological Organization (WMO) in February 2011.

As of 2011, only 50% of the South American GUAN stations transmitted regular reports (CIIFEN, 2012). This lack of data is further highlighted by the GCOS goal for establishing a reference network for upper-air climate observations (GRUAN) as an extension to GUAN. GRUAN standards are more rigorous than that of GUAN and aim at addressing historic

biases in the data. Of the 15 initial GRUAN sites, none were in South America. There is a current plan for expansion, as the importance of monitoring the Amazon is recognized, however the budgets to support this research have not yet materialized.

Figure 3.1. *Global positions of operational sounding stations (▲) as recorded by the WMO in February 2011. The Amazon basin is represented by the green area.*

The primary factor in the lack of sounding data availability is, quite simply, cost. It is important to recognize that this issue is recurrent in developing countries of Africa and South America, where budgets allocated to soundings is often limited or a minute part of a bigger budget (e.g. Cheruiyot, 2006; Douglas, 2008). This limitation results in the inability of a vast majority of developing countries to comply with the standards set by the WMO in their 2007 revision of the WMO Convention originally published on the 11th October 1947.

I here propose to use the predictive properties of ANNs for the downscaling of available surface and atmospheric data of the NCEP and CRU datasets, in order to calculate

condensation rates at a resolution suited for the evaluation of the existence of the Biotic Pump mechanism.

3.2 Time-Delayed Neural Networks

ANNs are powerful intelligent networks that mimic the natural neural network of the human brain. ANNs are intelligent in the sense that they learn to relate input to output through a trial-and-error process; the networks are not programmed to solve a problem, they are trained. The strength of ANNs is their ability to assimilate highly non-linear problems (Hayati and Mohebi, 2007). Research on ANNs started in the 1940's (McCulloch and Pitts, 1943) and has since then been applied successfully to a wide variety of fields where prediction is needed (e.g. finance, medicine, engineering, robotics, physics, etc.; ASCE, 2000a). More recently, ANNs have shown promising results in several studies that aimed at predicting or downscaling climate variables using neural networks (ASCE, 2000b; Baboo and Shereef, 2010; Cannon and Whitfield, 2002; Haylock *et al.*, 2006). For example, El-Shafie *et al.* (2011) compared the performance of an ANN to that of a multi-regression model in the prediction of precipitation and found that the former is more accurate. Another example is the work of Dibike and Coulibaly (2006) that successfully uses neural networks in the downscaling of precipitation and temperature and quantitatively assesses their superiority over a regression-based statistical downscaling model. Overall, ANNs are generally believed to be more powerful downscaling tools than other regression techniques usually used for the downscaling of climatological data since ANNs do not assume any set architecture; they let the data define it (von Storch *et al.*, 2000).

The most commonly used ANN architecture is probably that of a *Multi-Layer Perceptron* (MLP). MLPs are composed out of an input layer, one or more hidden layers and an output layer. The number of nodes (or neurons) in the input and output layers will represent the number of inputs and outputs given, while the hidden layers consist of an arbitrary number of nodes. Each node is connected by weighted vertices (or synapses) to each and every node of the following layer (see Figure 3.2). By means of forwarding the inputs through the network architecture (*feed-forwarding*), ANNs relate inputs to outputs by adapting the weights between nodes through a process known as *error backpropagation* (Haykin, 1998). Error backpropagation is an iterative adaptive learning function that works back from the calculated output, finding each node's contribution to the output error and updating the weight of the node in the direction that shows the smallest error. While being forwarded to the next layer, nodes' outputs are summed up and fed to the *transfer function* of the receiving node. As such, transfer functions determine the input-output behaviour of individual nodes.

Although various studies have used MLPs successfully in predicting or downscaling climate variables (Bustami *et al.*, 2007; Hayati and Mohebi, 2007), some claim that the common MLP architecture is not suited for predicting variables that vary over time (e.g. Coulibaly *et al.*, 2001). For this particular study I will hence use a temporal neural network termed "Time-Delayed Neural Network" (TDNN). TDNNs are MLPs that keep the previous states of the input layer in memory. Such networks have shown to be efficient in the prediction of time series (e.g. Coulibaly *et al.*, 2005; Dibike and Coulibaly, 2006; Gautam and Holz, 2000), the hypothesis being that a climate variable depends/relates to the previous state of that same variable (i.e. that climate variables are temporally dependent). The exact architecture

of the network (i.e. number of hidden layers, number of nodes in the hidden layers, *adaptive learning function*, transfer function and number of previous input states kept in memory) has to be arbitrarily determined by testing the performance of various structures.

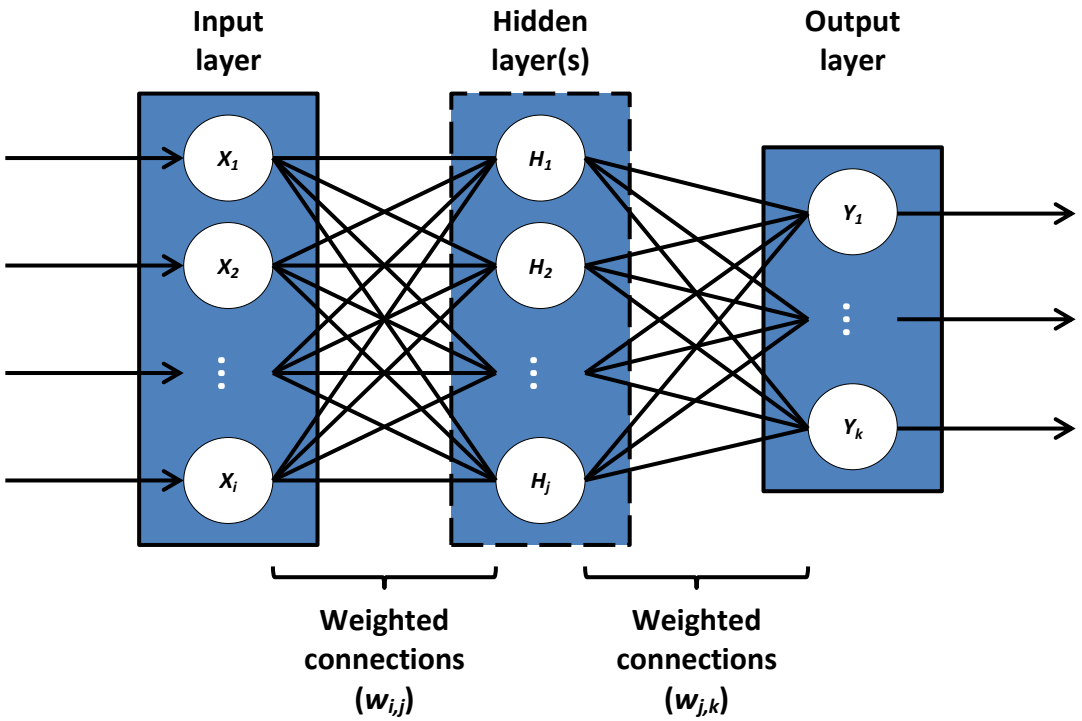


Figure 3.2. *Architecture of a multi-layer perceptron. Input nodes are represented as X_i, output nodes as Y_k, and the nodes from the hidden layer(s) are represented as H_j. The number of weighted connections w is the number of nodes from the previous layer times the number of nodes from the next layer.*

3.3 Methodology and model validation

This chapter aims at elaborating an ensemble of TDNNs that will be used to downscale monthly mean values of the variables necessary for the calculation of condensation rates. As was concluded in Chapter 2, five atmospheric variables need to be downscaled in order to

achieve this. The general idea is to have TDNNs predict these five variables from widely-available surface variables (i.e. surface temperature, surface relative humidity and surface pressure). Each TDNN is trained on the NCEP dataset at a horizontal resolution of 2.5° x 2.5° and then fed with the CRU surface data (at a resolution of 0.5° x 0.5°) to predict atmospheric and surface data at a finer resolution and, as such, downscale the data. Unfortunately, the original CRU dataset does not have surface pressure measurements, meaning that surface pressure too needs to be predicted using neural networks.

To do this, two TDNN ensembles are elaborated using the neural network toolbox of MATLAB R2010a (version 7.10.0.499). The first TDNN ensemble (hereafter referred to as "TDNN-A", where "A" stands for "atmosphere") is used for downscaling the five atmospheric variables: air temperature, pressure, water vapour pressure, zonal wind speed and vertical velocity. All five variables are necessary for the calculation of condensation rate. The second ensemble (hereafter referred to as "TDNN-S", where "S" stands for "surface") is used for downscaling surface pressure, which is a necessary input for the downscaling of the atmospheric variables. Both TDNN ensembles are trained over 30 years (1971 – 2000) and their predictions validated on the remaining 10 years (2001 – 2010) using the NCEP reanalysis data at a horizontal resolution of 2.5° x 2.5°. The trained networks of TDNN-A are then fed with 39 years (1971 – 2009) of CRU surface observations together with the surface pressure outputs of TDNN-S, at a horizontal resolution of 0.5° x 0.5°, to predict the tropospheric profiles of relevant atmospheric variables at a finer resolution (i.e. downscale atmospheric data to 0.5° x 0.5°).

In this study, accuracy of the simulations of the neural network ensembles on their training periods is not described in great detail for the main reason that the function of the trained networks is to downscale climatological data by predicting values given a set of unseen input data. The importance therefore does not lie as much in the accuracy of the model's simulations in predicting the training data, but rather in its performance in predicting the target values of the unseen data (i.e. the 10 years of data the networks have not been trained on). The accuracy of the predictions given the unseen dataset is hence quantified during the validation of the networks. Here, validation is defined as the quantification of the uncertainties that lie within the model's predictions. For example, looking at the validation results, one can say that the TDNN ensemble TDNN-A (used for the downscaling of atmospheric variables) can predict tropospheric profiles of air temperature, using unseen input data, with an average accuracy of plus or minus 0.41 °C. In contrast, saying that TDNN-A can simulate its own training dataset (i.e. its final performance after having converged) for air temperature with an average Mean Squared Error (MSE) of 1E-20 °C does not tell how reliable the ensemble is when fed with data it has not been trained to recognize.

It is important to note that this study is following the methodology used by Coulibaly *et al.* in his 2005 article on the downscaling of precipitation and surface temperature using temporal neural networks. The study of Coulibaly *et al.* makes use of 30 years of the NCEP dataset to train its networks and then validates the networks (i.e. by quantifying the accuracy of its predictions) using 10 years of unseen NCEP data. Training and validation of the network ensembles are found in Headers 3.4.2 and 3.5.2 ("Training and validation") and the discussion of the results is found in Header 5.2.1 ("Downscaling observations using Time-Delayed Neural Networks").

3.4 Temporal neural network ensemble: TDNN-A

3.4.1 Network ensemble and architecture

Ideally, one would want to construct only one TDNN per target variable for the whole of the study site, as to take spatial variation into account and diminish the user's workload. However, a network that covers the whole extent of the study site will be extremely time-consuming to train. From experience with this data, one can assume an exponential correlation between training time and the amount of training data. The study site is hence divided into eight equal parts as to reduce the strain on the computer's processor (see Figure 3.3).

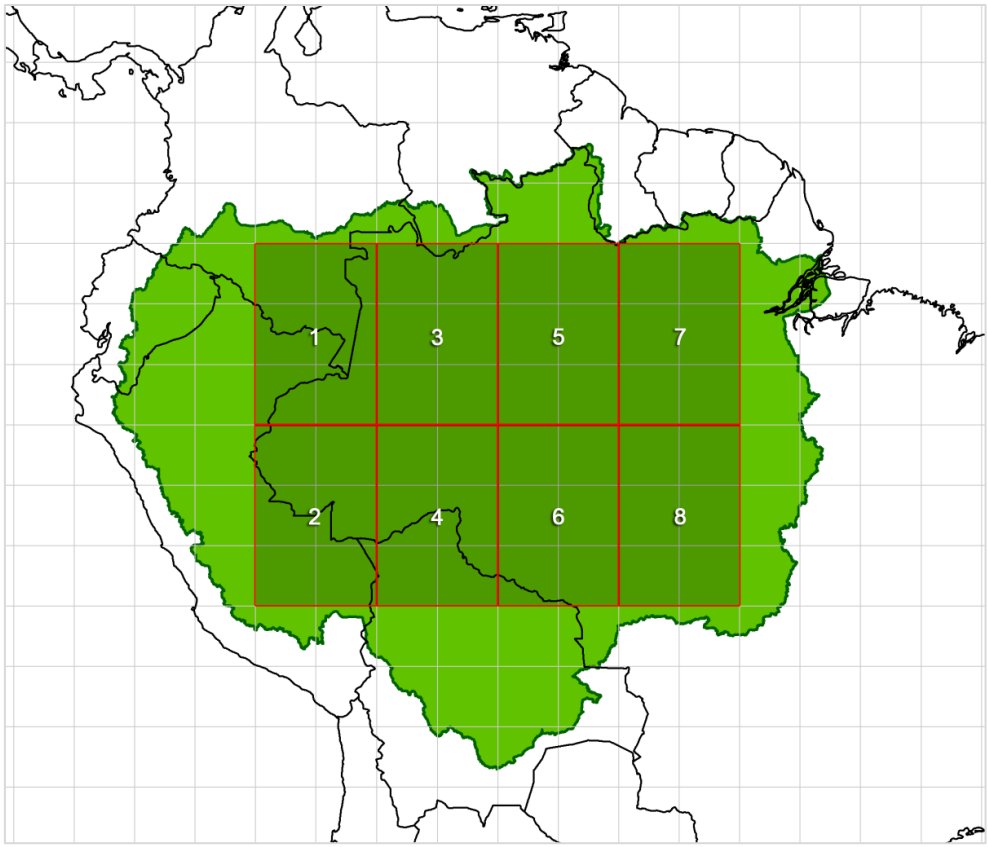


Figure 3.3. *Eight divisions of the study site, where the Amazon basin is represented as the green area. Grid boxes correspond to the resolution of the NCEP dataset.*

TDNN-A is composed out of eight TDNNs for each of the five target variable (one per study site part, per target variable), giving us a total of 40 neural networks. Each network has eight inputs (month, longitude, latitude, surface height, altitude of vertical layer, surface air temperature, surface relative humidity and surface pressure) and only one corresponding output (either air temperature, pressure, water vapour pressure, zonal wind speed or vertical velocity), depending on which target variable it has been trained to predict (see Table 3.1 and Figure 3.4). Using an ensemble of networks over a single neural network has shown to mitigate the problem of *data overfitting* (i.e. when the network is too finely-tuned to the training data; Cannon and Whitfield, 2002).

Inputs		
Variable	***Acronym***	***Unit***
Month	t	1 - 12
Longitude	x	°
Latitude	y	°
Surface height	z(0)	m
Altitude of vertical layer	z	m
Surface air temperature	T(0)	°C
Surface relative humidity	RH(0)	%
Surface pressure	p(0)	mbar

Outputs (at a given altitude)		
Variable	***Acronym***	***Unit***
Air temperature	T(z)	°C
Atmospheric pressure	p(z)	mbar
Water vapour pressure	$p_v(z)$	mbar
Zonal wind speed	u(z)	m s^{-1}
Vertical velocity	w(z)	m s^{-1}

Table 3.1. *Description of the input and output variables of the TDNN-A ensemble.*

Note that all inputs are used, even if these do not show direct correlations with the targeted output; there may or may not exist an indirect relationship for which the TDNN will automatically adapt its weights accordingly. Furthermore, location, altitude of the vertical layer, and month of the year are included in the input array in order for the network to

account for the target's spatial and temporal variations. Including the time period of the targeted variable as well as its horizontal and vertical location has shown to improve the network's prediction (e.g. Liu *et al.*, 2001; Smith *et al.*, 2005). For example, given that the month is January, longitude is 287.5°, latitude is -7.5°, the height of Earth's surface is at -27 m, the altitude of the vertical layer of interest is 100 m, surface air temperature is 24.99 °C, surface relative humidity is 93.30%, and surface pressure is 1013.5 mbar, the network should predict water vapour pressure at an altitude of 100 m to be 28.74 mbar.

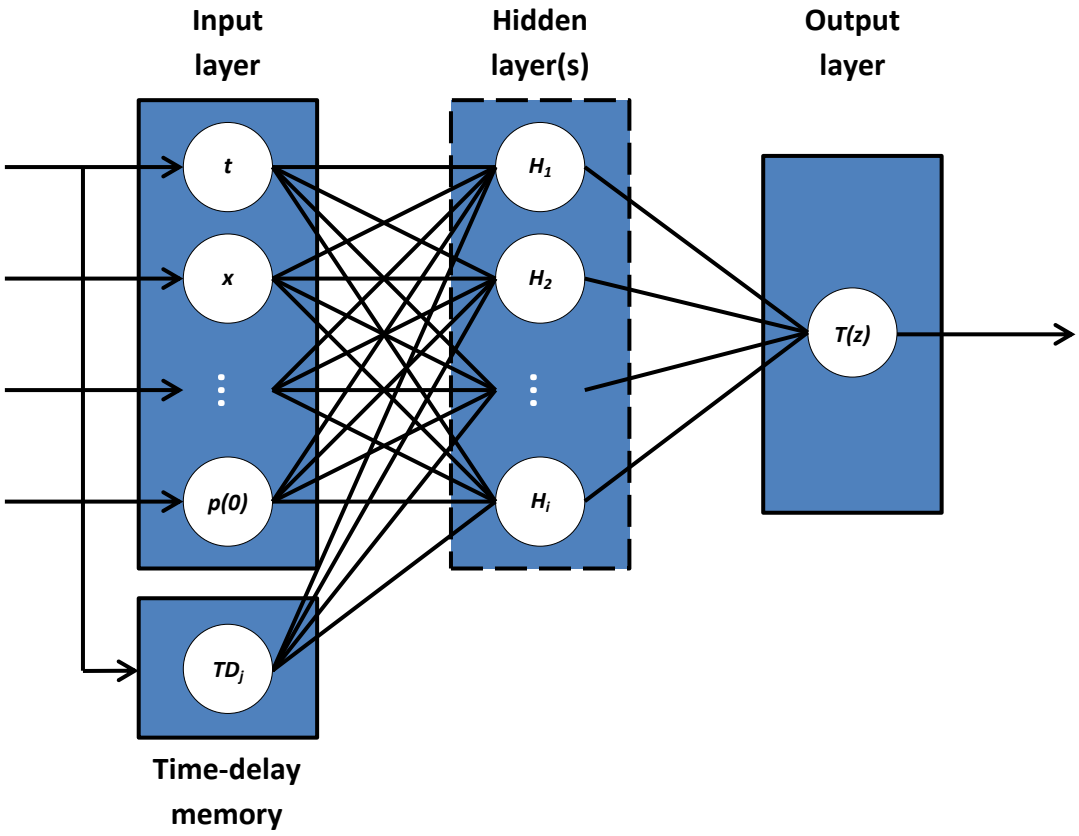


Figure 3.4. *Example architecture of the Time-Delayed Neural Networks used in this study. The TDNN in this example has eight inputs (t, x, y, z(0), z, T(0), RH(0) and p(0); see Table 3.1) and one output, air temperature at a given height (T(z); see Table 3.1).* H_i *represents the number of hidden nodes and* TD_j *is the time-delay which holds j-steps of the input layer (eight inputs * j-steps) in memory. The time-delay memory is forwarded together with the input layer to the hidden layer(s) and can hence be seen as part of the network's input.*

The best network architecture for this study was determined to be a feed-forward network with two hidden layers, each consisting of 70 nodes. It should be noted that, in this particular application of neural networks, the number of nodes used in each layer is not considered to be dependent on the intrinsic relationships between the input and output variables, but rather on the amount of training data used; the more nodes are being used, the closer the predictions of the neural network get to actual observations. This view is supported by Geman *et al.* (1992) who also state that the greater the number of hidden nodes used, the stronger it is prone to overfitting. However, data overfitting can be overcome by using large training sets at least 30 times greater than the number of hidden nodes (Sarle, 2002). In this particular study, the training dataset is some 160 times greater. The problem of data overfitting is further reduced by using an ensemble of neural networks as opposed to a single network. However, more nodes require greater processing capacity in order to train the network. The computer used in this study is able to handle a maximum of 70 nodes per layer, which justifies the amount of nodes used in this particular case.

It is common practice to use only one hidden layer in the construction of any MLP, as it has been proven that such architecture is able to approximate any complex relationship (Hornik *et al.*, 1989). MLPs are subsequently termed “universal approximators”. However, in this specific study, an architecture that encompasses two hidden layers performs much better than that of one hidden layer, both in training time and in the accuracy of its predictions. Figure 3.5 plots network performance against time to illustrate this point. Here, four networks that differ in number of hidden layers and nodes are being trained with a subset of the entire pressure training dataset for testing purposes. Training is stopped after 100 iterations or sooner if the minimum performance gradient is smaller than 1E-10. Networks 1

and 2 are composed out of one hidden layers with 70 and 140 nodes, respectively, while networks 3 and 4 are composed out of two hidden layers, where each layer counts 35 and 70 nodes, respectively; networks 1 and 3 hence each have 70 hidden nodes, and networks 2 and 4 each have 140 hidden nodes. Note that this example run is one of many that were undertaken to ascertain the recurrence of this behaviour.

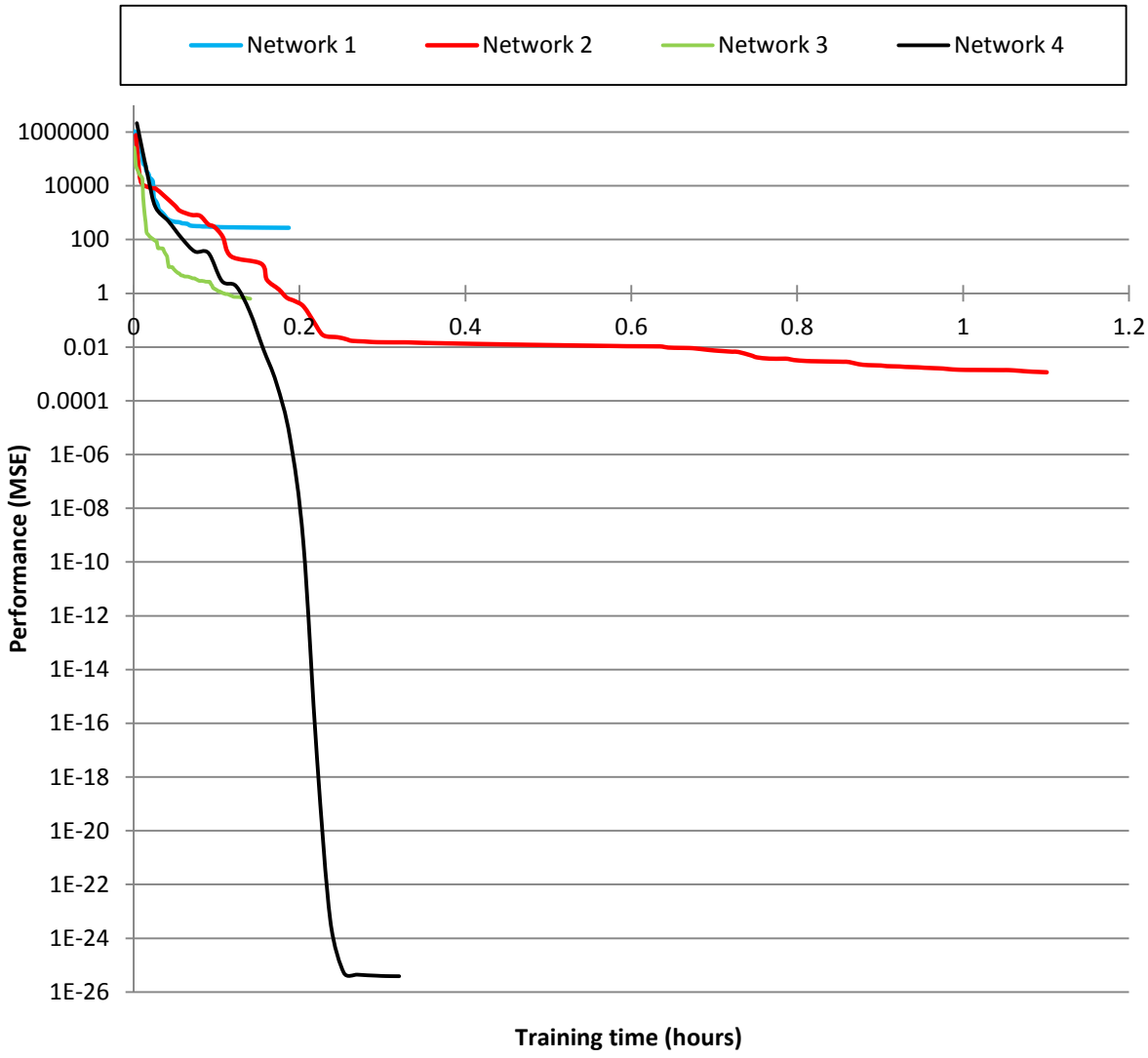


Figure 3.5. *Performance of Networks 1, 2, 3 and 4 against training time, where performance is defined as the Mean Squared Error (MSE) between training targets and training outputs. Networks 1 and 2 are composed out of one hidden layer counting 70 and 140 hidden nodes, respectively, and networks 3 and 4 are composed out of two hidden layers with each 35 and 70 hidden nodes, respectively.*

After 100 iterations, network 1 reaches a performance error of 274 mbar in 11 minutes and 13 seconds while network 2, with twice as many nodes, reaches an error of 0.0115 mbar in 1 hour 6 minutes and 3 seconds. This suggests that more nodes diminish training error but increase training time. Network 3 reaches 100 iterations after 8 minutes and 26 seconds with a final performance error of 0.628 mbar, which suggests that two layers perform faster that one layer with the same total number of hidden nodes. Network 4 shows a very small performance error of 3.86E-26 mbar after 19 minutes and 12 seconds of training. It hence offers the best combination of performance against time. It should also be taken into account that network 4 converges only after 20 iterations.

The number of time steps kept in memory (the time-delay) ranges from 9 to 12 steps. Since input and output values are monthly means, one can say that the network keeps 9 to 12 months in memory. The exact number of memory spaces allocated in each network depends on the temporal variance of the target variable and is detailed in Table 3.2.

Network	Target variable	Time-delay	Inputs	Output
net_t[1:8]	Air temperature	12	[t,x,y,z(0),z,T(0),RH(0),p(0)]	T(z)
net_p[1:8]	Atmospheric pressure	09	[t,x,y,z(0),z,T(0),RH(0),p(0)]	p(z)
net_v[1:8]	Water vapour pressure	09	[t,x,y,z(0),z,T(0),RH(0),p(0)]	$p_v(z)$
net_u[1:8]	Zonal wind speed	12	[t,x,y,z(0),z,T(0),RH(0),p(0)]	u(z)
net_w[1:8]	Vertical velocity	12	[t,x,y,z(0),z,T(0),RH(0),p(0)]	w(z)

Table 3.2. *Description of each network of the TDNN-A ensemble with regards to its target variable, the number of memory spaces allocated to its time-delay, the array of inputs and corresponding output (see Table 3.1. for inputs and outputs acronyms).*

The training function in use is the Levenberg-Marquardt backpropagation algorithm (LMA; Bishop, 1996). As opposed to the standard gradient descent backpropagation algorithm, the Widrow-Hoff algorithm (Haykin, 1998), the LMA provides a solution to the problem of

minimizing a function over a parameter space; meaning that in many cases it finds a solution even if it starts very far off the final minimum error. The adaptive function of the network is a gradient descent with momentum. The momentum is an optimization of the regular gradient descent function, which proves to be more efficient on complex error surfaces. The activation functions in the hidden layers are tan-sigmoid transfer functions and the output layer has a linear transfer function. The linear function is useful when calculating accurate errors during the training process.

3.4.2 Training and validation

Each TDNN is trained using a set of 30 years of the NCEP monthly means data (1971 – 2000). During the training, both inputs and targeted outputs are provided to the network. The network adapts its internal weights in order to relate inputs to targeted outputs. Each training set is composed of 12 input-target pairs per year, per vertical layer, per gridbox; which amounts to a maximum of 17280 pairs. Input-target pairs are left out when surface height is higher than the altitude of the target data. The network is considered to have *converged* when the training has reached 2 hours or 100 iterations; MSE between actual output and predicted output are then usually in the order of 1E-20. Note that during the first hour of training MSE descends rapidly, and thereafter diminishes marginally. The training is stopped after 2 hours or 100 iterations mainly due the restrictions imposed by the processing capability of the platform used for this study. Better network performance could be expected from longer training periods; this may however also increase bias in the network's predictions caused by data overfitting.

The trained network is then validated against unseen NCEP data using the remaining 10 years (2001 – 2010) in order to assess how well the network performs outside its training period; here, the process of validation defines the bias in the network's predictions. Unlike during the training of the network, targets are not provided and the network's weights remain unchanged; the network now simulates output from given inputs. Inputs are taken from the 10 years of the NCEP data the network was not trained on. Accuracy of these simulations is quantitatively determined for each TDNN by calculating Mean Absolute Errors (MAE) between the network's outputs and the NCEP data corresponding to the given inputs. Results are assessed monthly to check for temporal variation in the network's performance, and per vertical layer to check for vertical spatial variation (Tables 3.3 and 3.4, respectively).

Horizontal spatial variation of MAE is not included as no significant variations were found in the results. Results also show that all networks perform well in the prediction of their respective output variables. From network averages, one can note that MAE for the prediction of zonal wind speed is higher than others. This reflects the chaotic behaviour and inhomogeneity observed in horizontal wind speeds, both over time and altitude. It could also be argued that relating profiles of zonal winds to surface conditions is not sufficient for predicting this variable accurately. Contrarily, vertical velocity shows a very small MAE due to the data's small variability (difference between greatest and smallest values is on average 0.17 m s^{-1}). Note that the average of the averages (bottom right corner of each table) can be considered as a reference index of the overall performance of TDNN-A which could be compared to the performance of similar ANN ensembles.

Mean Absolute Error per month						
	net_t[1:8] (°C)	net_p[1:8] (mbar)	net_v[1:8] (mbar)	net_u[1:8] (m s^{-1})	net_w[1:8] (m s^{-1})	Average
Jan	0.44	0.46	0.65	2.70	0.024	**0.85**
Feb	0.56	0.48	0.70	2.84	0.028	**0.92**
Mar	0.50	0.48	0.67	2.73	0.022	**0.88**
Apr	0.46	0.44	0.62	2.72	0.020	**0.85**
May	0.51	0.47	0.58	2.92	0.022	**0.90**
Jun	0.50	0.40	0.56	2.83	0.020	**0.86**
Jul	0.47	0.42	0.54	2.66	0.020	**0.82**
Aug	0.52	0.38	0.54	3.26	0.021	**0.94**
Sep	0.52	0.41	0.61	2.56	0.020	**0.82**
Oct	0.44	0.33	0.59	2.73	0.020	**0.82**
Nov	0.43	0.31	0.58	2.35	0.022	**0.74**
Dec	0.43	0.34	0.65	2.80	0.025	**0.85**
Average	**0.48**	**0.41**	**0.61**	**2.76**	**0.022**	**0.86**

Table 3.3. *Monthly Mean Absolute Errors between outputs of the TDNN-A ensemble and NCEP data.*

Mean Absolute Error per vertical layer						
	net_t[1:8] (°C)	net_p[1:8] (mbar)	net_v[1:8] (mbar)	net_u[1:8] (m s^{-1})	net_w[1:8] (m s^{-1})	Average
100m	0.47	0.54	1.11	2.85	0.029	**1.00**
750m	0.37	0.32	0.77	2.20	0.021	**0.74**
1500m	0.49	0.19	0.90	2.25	0.024	**0.77**
3000m	0.45	0.40	0.79	2.58	0.022	**0.85**
4250m	0.41	0.38	0.59	2.65	0.021	**0.81**
5500m	0.47	0.44	0.45	2.72	0.021	**0.82**
7500m	0.55	0.51	0.37	3.11	0.023	**0.91**
9500m	0.64	0.59	0.30	3.73	0.021	**1.06**
Average	**0.48**	**0.42**	**0.66**	**2.76**	**0.023**	**0.87**

Table 3.4. *Mean Absolute Errors per vertical layer between outputs of the TDNN-A ensemble and NCEP data.*

Little temporal and vertical dependence is shown by the results, suggesting that the ensemble is robust to such variations and is more dependent on the variance of the target variable and/or its relation to input variables. The only noticeable correlations are found in the MAE of the predictions of water vapour pressure, where errors decrease with altitude, and for zonal wind speeds, where errors increase with altitude. This again could be explained through the decreasing variance between the minimum and maximum values of water vapour pressure with altitude and vice versa for zonal wind speeds.

3.5 Temporal neural network ensemble: TDNN-S

3.5.1 Network ensemble and architecture

The structure of each TDNN of TDNN-S remains largely the same as that used for the downscaling of atmospheric variables: two hidden layers of 70 nodes each, an LMA training function, a gradient descent with momentum as adaptive function, tan-sigmoid transfer functions for the hidden layers and a linear transfer function for the output layer. Three properties change in the structure of the TDNNs and TDNN-S: 1) there are only six input variables, surface pressure and altitude of the vertical layer are left out, 2) the output variable is only found at the surface, and 3) the study site window is divided into four equal parts (see Figure 3.6), not eight, since the amount of data being processed is less because of having only one vertical layer (surface). TDNN-S hence encompasses only four networks; one per study site part for one target variables (see Table 3.5).

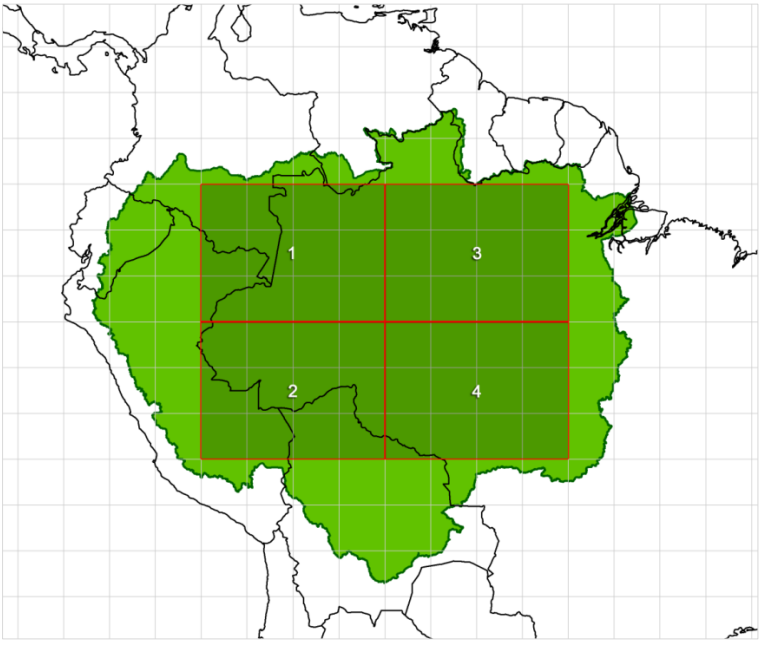


Figure 3.6. *Four divisions of the study site, where the Amazon basin is represented as the green area. Grid boxes correspond to the resolution of the NCEP dataset.*

Network	Target variable	Time-delay	Inputs	Output
net_p[1:4]	Pressure	11	[t,x,y,z(0),T(0),RH(0)]	p(0)

Table 3.5. *Description of each network of the TDNN-S ensemble with regards to its target variable, the number of memory spaces allocated to its time-delay, the array of inputs and corresponding output (see Table 3.1. for inputs and outputs acronyms).*

3.4.2 Training and validation

Likewise to the training of TDNN-A, TDNN-S is trained using the NCEP dataset over the period 1971 – 2000 and then validated on the remaining 10 years. Each training set is composed of 12 input-target pairs per year, per gridbox; which amounts to a maximum of 2160 pairs. Again, training is stopped after 2 hours or after 100 iterations due to the restrictions imposed by the computer's processing capability. The accuracy of the simulations is determined using MAE and is illustrated in Table 3.6.

Mean Absolute Error per month	
	net_p[0] (mbar)
Jan	0.44
Feb	0.56
Mar	0.50
Apr	0.46
May	0.51
Jun	0.50
Jul	0.47
Aug	0.52
Sep	0.52
Oct	0.44
Nov	0.43
Dec	0.43
Average	**0.48**

Table 3.6. *Monthly Mean Absolute Errors between outputs of the TDNN-S ensemble and NCEP data.*

The results show that the network performs well in the prediction of surface pressure with an average MAE of 0.48 mbar for an area coverage greater than 3500000 km^2. Again, no significant temporal dependence is shown by the results, suggesting that the ensemble is robust to such variations.

3.6 Downscaling

The trained and validated ensembles are now used to predict atmospheric variables at a smaller horizontal resolution. This is achieved by using the CRU surface data and the surface pressure outputs of TDNN-S as inputs to the TDNN-A ensemble (see Figure 3.7). Moreover, the CRU surface variables are at a smaller resolution than that of the NCEP dataset; some five times smaller. As such, TDNN-A downscales the data from 2.5° cells to 0.5° cells (approximately 278 km to 55 km).

Note that the first year of the TDNN-A output is not taken into account in the comparison as it is considered to be the time necessary for the calibration of the time-delay memory. Furthermore, the year 2010 is also not taken into account as this data is not available from the CRU dataset. This gives us a total usable period of 38 years (1972 – 2009). Furthermore, it should be born in mind that additional uncertainties could have been added to the output due to the inherent differences between the NCEP and CRU surface data. The data that results from the downscaling procedure is hereafter referred to as “CRU-DS”, where “DS” stands for “downscaled” (see Table 3.7).

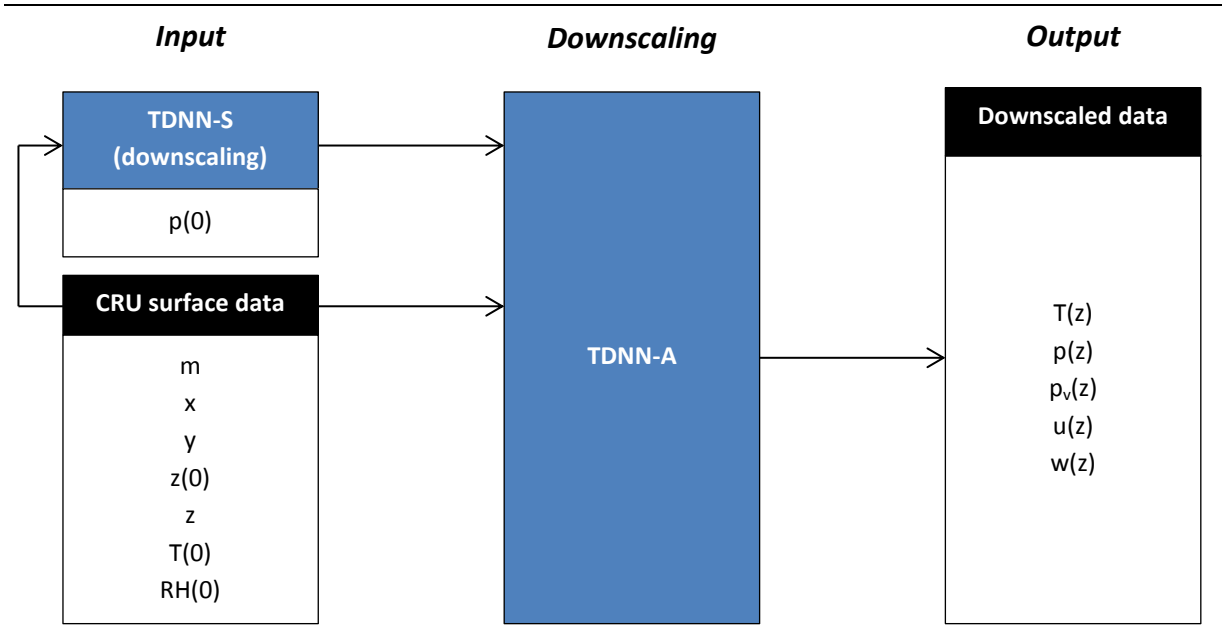


Figure 3.7. *Diagram of the structure of the downscaling procedure using the CRU surface data and both TDNN ensembles, TDNN-S and TDNN-A. Ensembles are represented in blue and datasets in black.*

Level	Variable	CRU-DS acronym	Unit
Atmosphere	Air temperature	T	°C
	Vertical velocity	w	m s^{-1}
	Zonal wind speed	u	m s^{-1}
	Atmospheric pressure	p	mbar
	Water vapour pressure	p_v	mbar
Surface	Air temperature*	T	°C
	Vertical velocity	w	m s^{-1}
	Water vapour pressure*	p_v	mbar

Table 3.7. *Downscaled data variables used in the study (CRU-DS).*

** Available from the original CRU dataset without TDNN downscaling*

The validation of the predictions of the trained ensembles has shown that temporal neural networks are suited for the downscaling of atmospheric and surface variables, with the lowest error found where the variance between minimum and maximum value of the target variable is lowest. The tropospheric profiles of the five variables resulting from the downscaling procedure, referred to as the CRU-DS dataset, are used in the next chapter as

input to Equations 1 to 5 (See Heading 2.4) in order to calculate condensation rates. The condensations rates are then mapped as to study the properties of regions with condensation versus regions where no condensation takes place, enabling the formulation of conclusions with regards to the quantitative assessment of this particular aspect of the Biotic Pump Theory.

Chapter 4

Results

The results of calculating condensation rates over the study area are presented in this chapter. Firstly, I explain the procedure followed in this study for determining gridbox values of condensation rates and then analyse the mapping of donor and acceptor regions over space and time. This is followed by a study of the changes in average condensation rate over the period 2001 – 2009. The chapter is finalized by an examination of the relationship between vertical velocities, zonal wind speeds and condensation rates for the whole of the study area over the period 1972 - 2009. All of these analyses are done for the purpose of quantitatively assessing the Biotic Pump Theory with regards to the existence and working of donor and acceptor regions.

4.1 Dynamically allocated donor and acceptor regions

Calculating the condensation rate of each 0.5° gridbox using Equations 1-5 (see Heading 2.4) gives us the possibility to assess whether a gridbox is considered to be a donor region or an acceptor region. Recall that both can be identified based upon their respective

condensation rates (i.e. condensation is found in the acceptor region while none is found in the donor region). Furthermore, if the region undergoes condensation, one can determine how strong the condensation rate is at that particular point. The condensation rate is calculated in between gridboxes and vertical layers as the equation accounts for the differences in temperature, and molar density of dry and water vapour, in the x-, y- and z-direction (see Figure 4.1). This gives us a total of 7448 condensation rate values (38 x 28 gridboxes in the x- and y-direction, respectively, for seven vertical layers) per month for each of the 38 years (1972 – 2009).

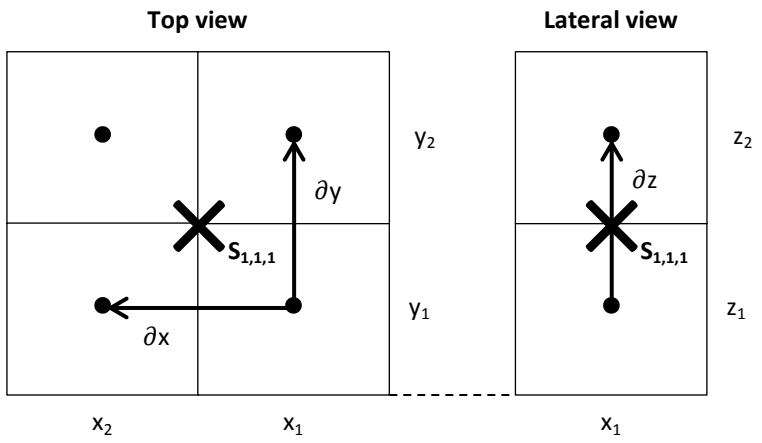


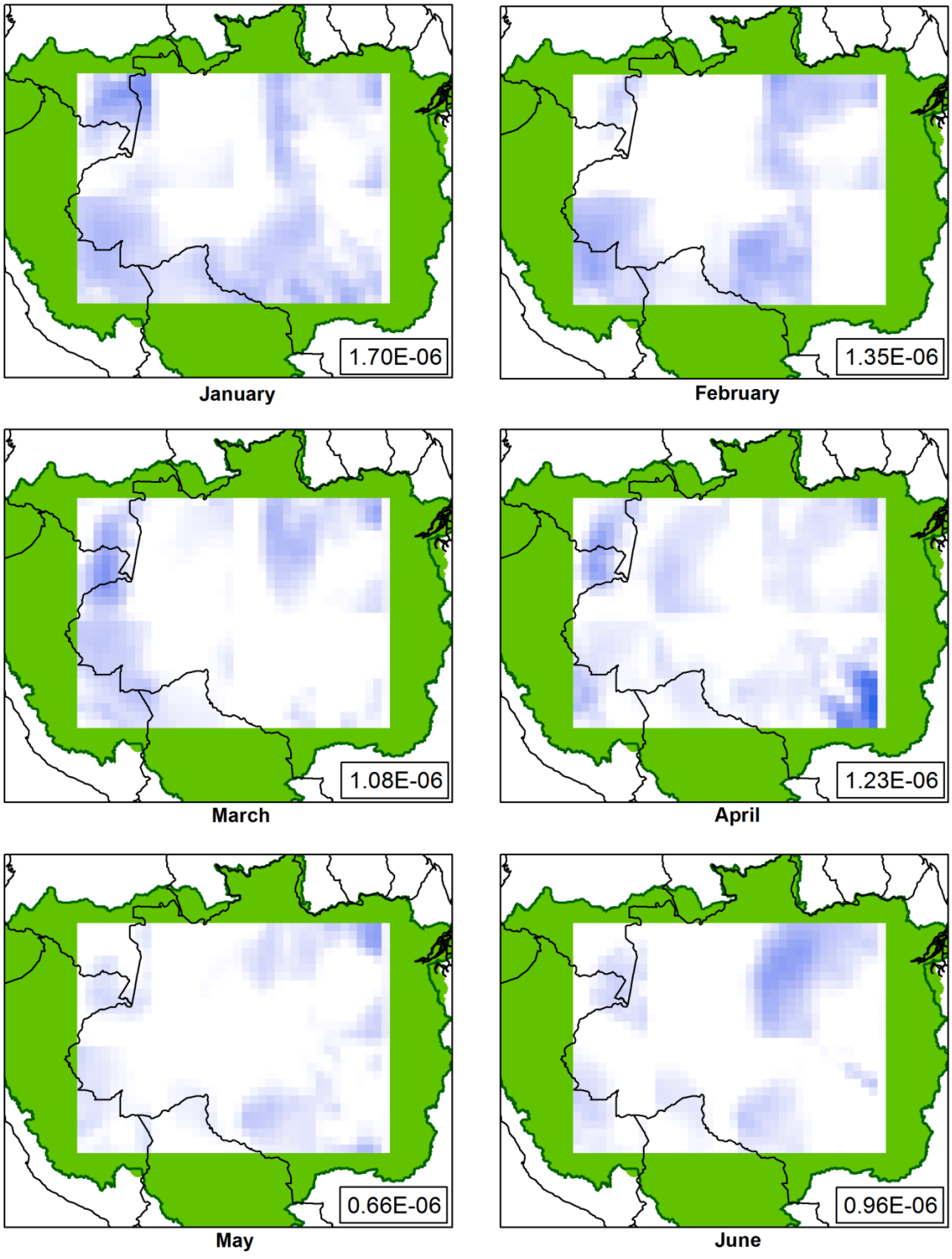
Figure 4.1. *Schematic representation of the gridbox calculation of condensation rate* $S_{x,y,z}$ *in between gridboxes and vertical layers, where ∂x, ∂y and ∂z are changes in a specific variable in the x-, y- and z-direction, respectively.*

The equation for condensation rate (Equation 1; Makarieva and Gorshkov, 2010a) however does not return zero values when no condensation is taking place. Instead, it returns negative condensation rate values which show gradients ranging between -1.38E-05 mol m^{-3} s^{-1} and -1.49E-09 mol m^{-3} s^{-1}. These values could be interpreted as being the degree to which atmospheric conditions are prone to produce condensation (e.g. the lower the negative

value, the less likely it is for condensation to take place). Negative values are hence substituted for zero, meaning that there is no condensation. Moreover, one should bear in mind that having a zero value for a gridbox does not mean that absolutely no condensation takes place over these 56 km^2; what it does say is that on average no condensation takes place (i.e. the negative values are on average more important than the positive condensation rate values).

Monthly means of condensation rate values, averaged over all seven vertical layers, are mapped in Figure 4.2; the value in the bottom-right corner of each individual map represents the catchment average of the condensation rate (i.e. sum of all condensation rate values divided by the total number of gridboxes). The set of maps clearly illustrates the existence of donor and acceptor regions at macro-scale. The distribution and size of these regions vary over time showing seasonality; however these do not show any spatial dependence. The highest rate of condensation is found during the month of October with 1.71E-06 mol m^{-3} s^{-1}, while the lowest rate is found in May with 0.66E-06 mol m^{-3} s^{-1}. Mean annual condensation rate for the Amazon basin is of 1.23E-06 mol m^{-3} s^{-1}. The wet and dry seasons (December to May and June to November, respectively) show on average similar rates of condensation, which are of 1.24E-06 mol m^{-3} s^{-1} and 1.23E-06 mol m^{-3} s^{-1}, respectively. Furthermore, from comparing seasonal variations between the calculated condensation rates and the climatologic averages (1960 – 1990) of monthly precipitation in Manaus, Brazil and Leticia, Colombia, one can notice a one- to two-month lag between the high and low points of condensation rates and precipitation occurrences (see Figure 4.3). Widths of the donor regions range between 56 km and 1680 km, and that of the acceptor regions range between 56 km and 2128 km. The ratio of donor versus acceptor region is of

41.88% - 58.12%, respectively. Note that the eight divisions of the study site pertaining to the downscaling of the atmospheric variables using the Time-Delayed Neural Networks (TDNNs) are clearly visible in the maps.



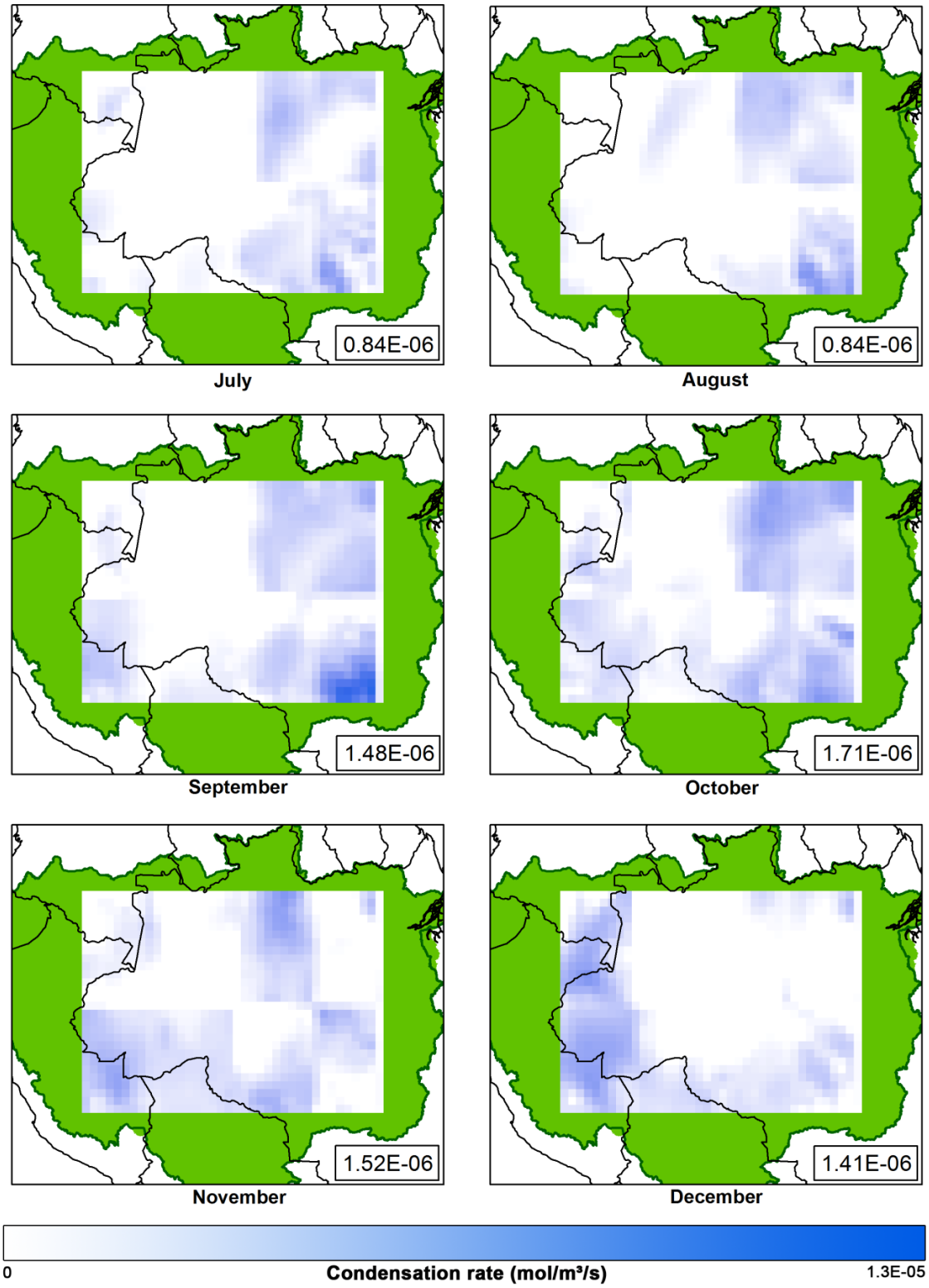


Figure 4.2. *Monthly means of condensation rate (mol m^{-3} s^{-1}) averaged over 1972 – 2009 and seven vertical layers, where the Amazon basin is represented as the green area. Catchment average of condensation rate is shown in the bottom-right corner of each map.*

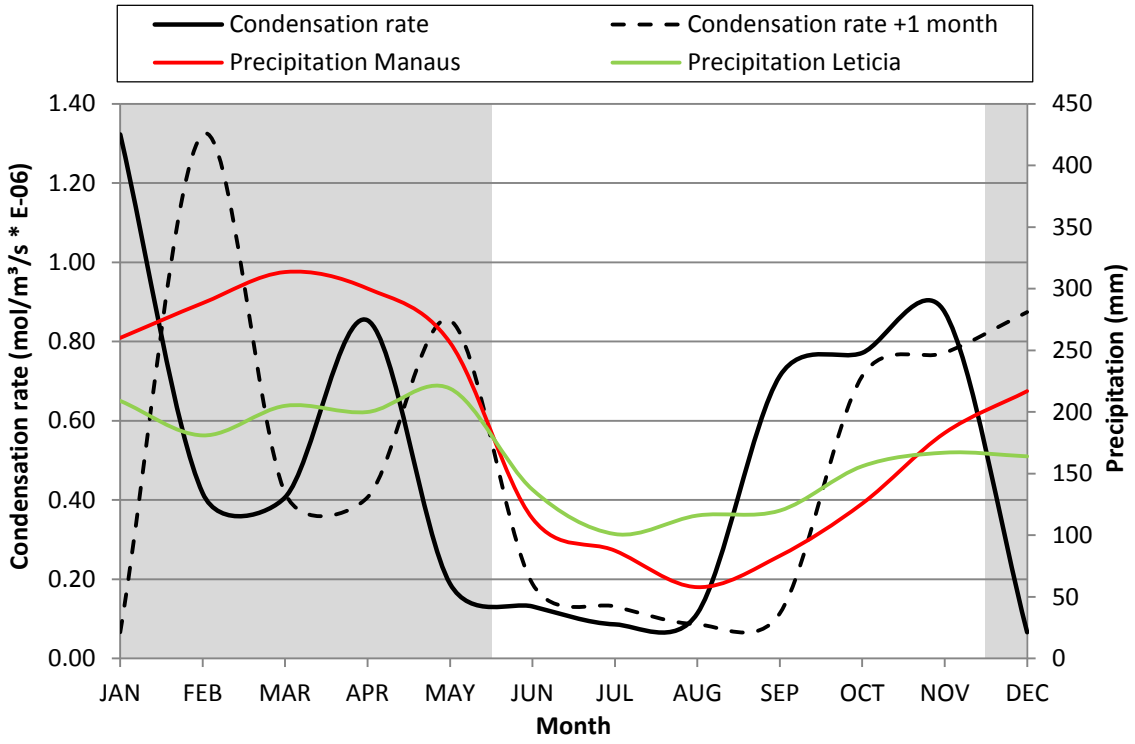


Figure 4.3. *Monthly catchment means of condensation rate, averaged over 1972 – 2009 and seven vertical layers, and monthly climatological means of precipitation in Manaus, Brazil, and Leticia, Amazonas, averaged over 1960 – 1990. The shaded areas represent the wet season. Climatological means are taken from the Hong Kong Observatory website (http://www.weather.gov.hk/).*

4.2 Changes in condensation rate

Monthly changes in condensation rate over the last 9 years of the data (2001 – 2009) are compared against the preceding 29 years (1972 – 2000) as to assess the effects of the ever-increasing clearing of the Amazon basin vegetation (e.g. Davidson E. A. *et al.*, 2012; Moore *et al.*, 2007; Soares-Filho *et al.*, 2006; Werth and Avissar, 2002) on the process of condensation. We here only choose 29 years to define the baseline climatology as the total period of the data is of 38 years. Ideally, the period 1971 – 2010 would have been chosen, enabling the calculation of changes over the last 10 years of data against the preceding 30

years. However, the year 1971 of data is part of the calibration period of the downscaling procedure and 2010 is not available from the CRU dataset. It is important to bear in mind that the World Meteorological Organization standards recommend a period of 30 years to determine climatological means (WMO, 1983).

Condensation rate values are averaged over the whole catchment for the seven vertical layers for each of the selected periods. The changes in average condensation rate, illustrated in Figure 4.4, suggest that the last 9 years have undergone a significant increase in condensation rate during the months of October and November, with an average increase of 0.38E-06 mol m^{-3} s^{-1} and 0.55E-06 mol m^{-3} s^{-1}, respectively. A slight decrease of 0.17E-06 mol m^{-3} s^{-1} is found during the month of June. On average, an increase in condensation rate is found compared to that of the baseline climatology defined over the period 1972 – 2000. This increase is on average of 0.06E-06 mol m^{-3} s^{-1} over the whole of the study site. Changes in condensation rate are found over the dry season, while no significant changes are recorded during the wet season. Figure 4.5 shows the annual progress of mean condensation rate over the entire study period (1972 – 2009). The trend of the annual mean condensation rate indicates a yearly increase of 0.0009E-06 mol m^{-3} s^{-1}.

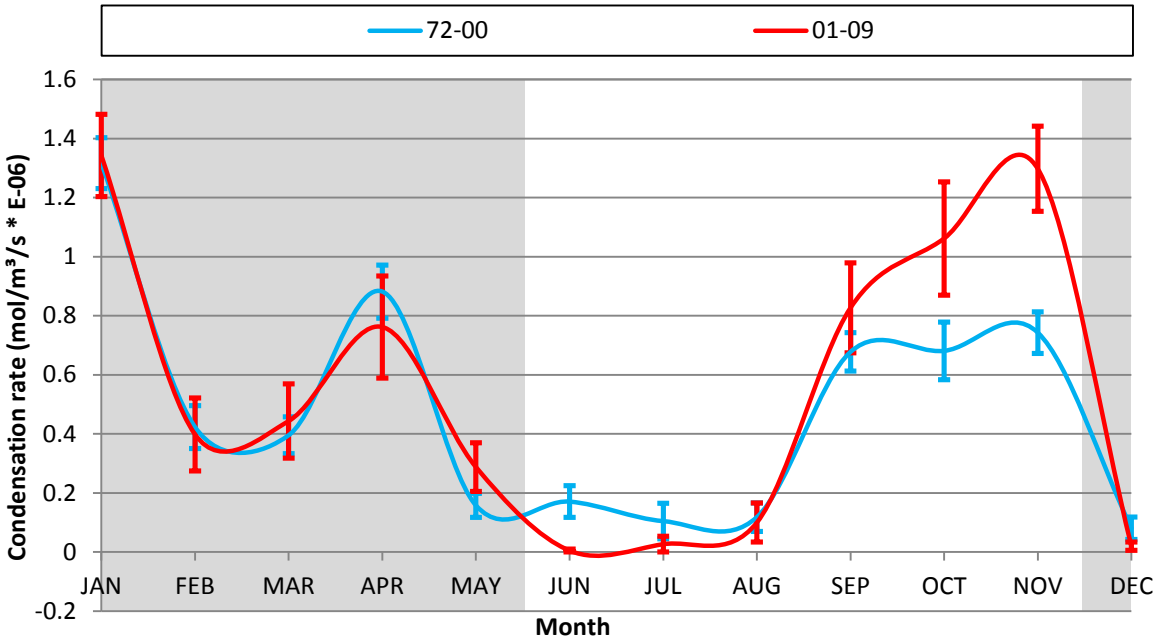


Figure 4.4. *Monthly means of condensation rate for the last 9 years of data (2001 – 2009) averaged over the study site catchment and the seven vertical layers. Baseline climatology is calculated over the preceding 29 years (1972 – 2000). The error bars represent the catchment monthly standard errors between yearly condensation rates and average condensation rate for the periods 1972 – 2000 and 2001 – 2009. The shaded areas represent the wet season.*

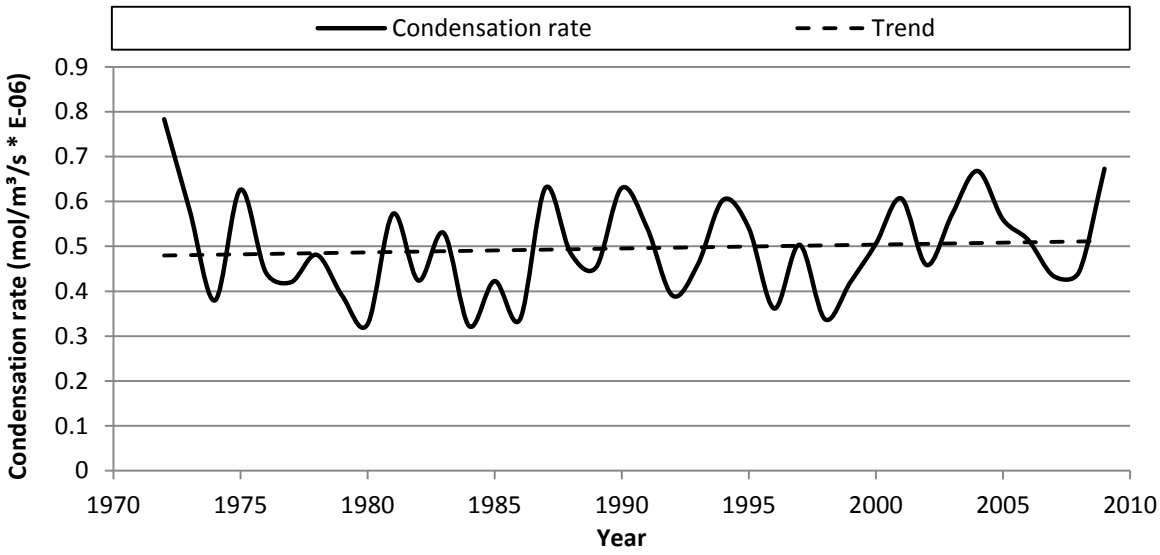


Figure 4.5. *Yearly means of condensation rate for the period 1972 – 2009 averaged over the study site catchment and the seven vertical layers.*

4.3 Condensation rate versus zonal wind speed

According to M&G, the major force driving winds is induced by condensation. Catchment annual means of condensation rate, zonal wind speed and upward velocity are plotted against altitude in Figure 4.6 as to assess whether such relationship is reflected in the form of correlations between these three variables. Annual means are averaged over 38 years, from January 1972 to December 2009.

Figure 4.6 suggests that the great majority of water vapour condensates between the surface layer up to about 1250 m. Between 1250 m and 1750 m the condensation rate becomes negative. It then increases slightly and becomes positive again. From an altitude of 2250 m condensation rate decreases marginally and steadily until reaching 6250 m where its increase is more pronounced. Zonal wind speed and upward velocity follow similar curves to one another. Both decrease together with condensation rate, are then approximately flat where condensation rates decrease slightly and steadily, and finally increases significantly together with condensation rate. Furthermore, results suggest that from the surface layer up until an altitude of approximately 625 m zonal winds are predominantly blowing from East to West (where wind speed is positive), while these are travelling from West to East at higher altitudes (where wind speed is negative). Bear in mind that these statements are valid at catchment level only; this will not apply where surface height is higher, as for example in the Andean region of the study site (bottom-left corner) where maximum surface height is of 5925 m.

Now that the results of this study have been analysed it is possible to formulate conclusions with regards to the existence and properties of acceptor and donor regions, and hence determine whether the overall results support or not to the existence of a biotically-regulated mechanism. The final chapter of the thesis encompasses a summary of the Biotic Pump theory and the approach used to quantitatively assess the validity of its hypothesis with regards to condensation-induced dynamics of water vapour and winds. This is then followed by a discussion on the results of the downscaling of observations using TDNNs and on the results of the analysis of condensation rates over the study area. Finally, conclusions and recommendations for future research on the topic are formulated.

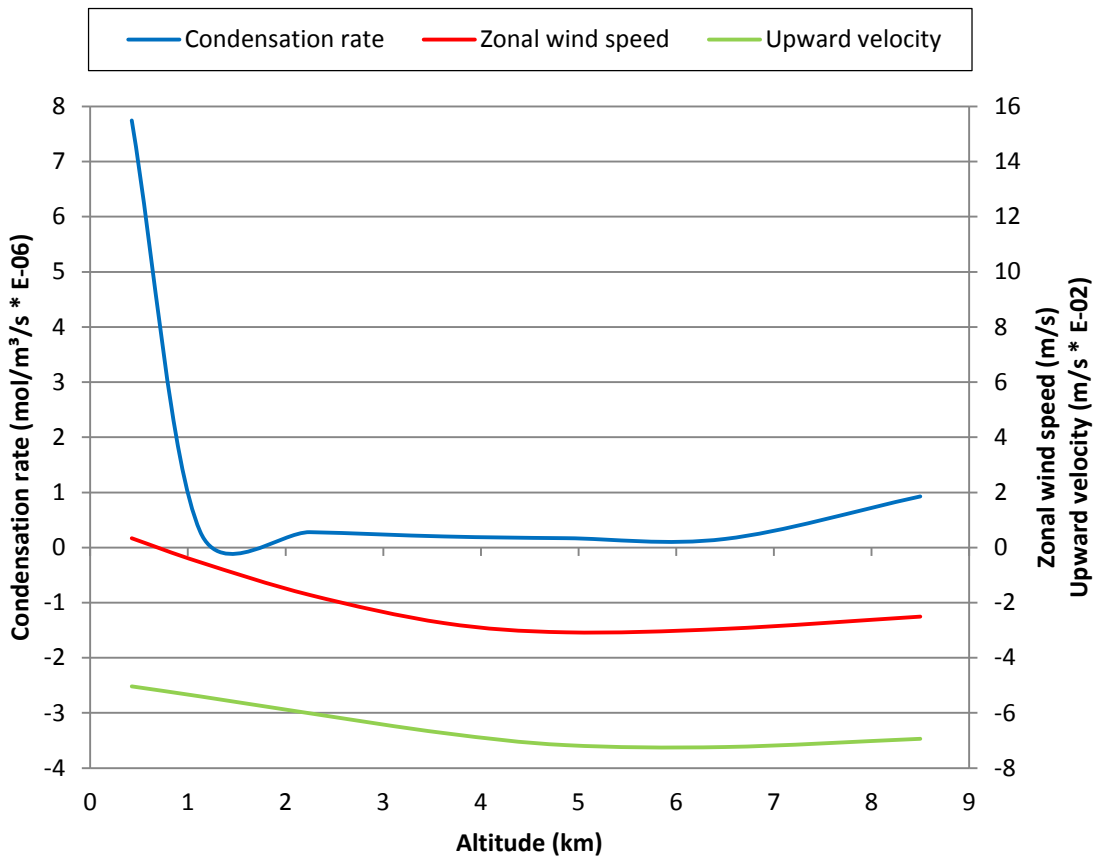


Figure 4.6. *Catchment annual means of condensation rate, zonal wind speed and upward velocity against altitude.*

Chapter 5

Discussion and conclusions

5.1 Summary

The Biotic Pump Theory (BPT), as set forward by Drs. Makarieva and Gorshkov (M&G; 2007; 2009a; 2010a; 2010b; Makarieva *et al.*, 2009; 2012; 2010), states that the major force driving winds, and hence the movement of water vapour, is induced by the changes in atmospheric pressure caused by condensation while, traditionally, winds are believed to be products of the differences of Earth's surface heating. The implications of such a theory are highly relevant to a wide variety of scientific fields as it involves the global regulation and distribution of water – also known as "the blue gold". Other than emphasizing the fragile equilibrium of tropical rainforests with regards to human-induced deforestation, the theory could potentially provide answers to questions as old as human existence; for example, that of the evolution of vegetation on land. Although built on fundamental physics, the BPT is highly controversial as it defies elementary and longstanding knowledge of atmospheric processes. This study therefore aims at investigating the validity of the physics of the BPT over the Amazon basin.

The BPT describes the mechanism by which water vapour is transported from areas of low evaporation to areas of high evaporation, conveniently termed “donor” and “acceptor” regions, respectively. According to M&G, only acceptor regions exhibit condensation because of the high evaporation rates found in those regions. This condensation of water vapour engenders a decrease in atmospheric pressure over acceptor regions, which subsequently creates a flux (i.e. winds) from high pressure to low pressure systems. Unfortunately, due to the lack of funding allocated towards collecting atmospheric observations over such isolated areas as the Amazon basin, it is near impossible to test a great majority of the physical processes described by M&G as available observations are simply too coarse and inconsistent. Current atmospheric observation datasets over the Amazon basin are interpolated onto a 2.5° (approximately 278 km) grid from satellite observations and very few sounding stations. However, much of the atmospheric processes need to be determined at higher resolutions (e.g. Carter *et al.*, 1994; Coulibaly *et al.*, 2005; Hewitson and Crane, 1996; Wigley *et al.*, 1990) to provide for an acceptable degree of confidence.

In view of the complications and the nature of the theory, I here focus on implementing a conceptual model (Biotic Pump Conceptual Model; BioPCM) that downscales available data, calculates condensation rates and maps these rates in order to differentiate donor regions from acceptor regions using their respective condensative properties. As such, I hope to quantitatively assess the existence of these regions and, by doing so, contribute to research on the BPT and atmospheric dynamics. The model uses the predictive capabilities of Time-Delayed Neural Networks (TDNNs) to downscale atmospheric observations from higher-resolution surface observations, enabling us to account for sub-grid processes. Artificial

neural networks are here chosen as these have shown promising results in the prediction and downscaling of meteorological variables (e.g. ASCE, 2000b; Baboo and Shereef, 2010; Cannon and Whitfield, 2002; El-Shafie *et al.*, 2011; Haylock *et al.*, 2006). Condensation rate for each gridbox is then calculated using the physics detailed in the BPT. Finally, donor and acceptor regions are mapped and identified through their respective condensative properties. The construction and running of BioPCM constitutes the study's specific objectives A to C as described in Chapter 1.

Two TDNN ensembles are created for the downscaling procedure. The first ensemble, TDNN-S, is used for predicting the missing surface variable necessary to the downscaling of atmospheric observations from higher-resolution surface observations. The second ensemble, TDNN-A, is then used together with the outputs of TDNN-S for the downscaling of the five atmospheric variables essential to the calculation of condensation rate. Before being used for the downscaling procedure, both ensembles are trained over 30 years (1971 – 2000) and their predictions are then validated against the next 10 years (2001 – 2010). Overall, results of the validation suggest that TDNNs are an effective method for predicting tropospheric profiles of air temperature, pressure, water vapour pressure, zonal wind speed and vertical velocity, using only widely-available surface variables (temperature, relative humidity and pressure). The largest errors are found in the predictions of zonal wind speeds, where mean absolute error is on average 2.76 m s^{-1}. One can hypothesize that this results from the erratic behaviour of winds, both in space and time. Alternatively, it is possible to conclude that predicting zonal wind speed from only three surface variables (temperature, relative humidity and pressure) is not sufficient for achieving an accurate estimate. Lowest mean absolute errors between observations and simulations are of 0.022 m s^{-1} for vertical

velocity. In general, predictions of the TDNNs get better as the gap between maximum and minimum value of the target variables gets smaller.

Once validated, the TDNN ensembles are used for the scaling of atmospheric observations from 2.5° down to 0.5° (approximately 278 km and 56 km, respectively) for the period 1972 – 2009; timespan differs from the training and validation periods as the first year is used for the calibration of the neural network's time-delay and because the year 2010 is not available in the surface observations dataset. It is important to note that due to the constraints set by the equipment's processing capabilities, it was necessary to divide the study site into eight equal parts. This limitation produced inhomogeneities between the divisions of the study site in the form of clearly visible lines in the resulting condensation rate maps (see Figure 4.2). One could overcome this by implementing an averaging procedure in between study site divisions or by using a more powerful computer that is able to handle the whole of the study site. An alternative approach would be that of having a "sliding window" where only one TDNN per target variable is being trained successively on each division of the study site. As such, the neural network will take spatial variation into account, not only inside each study site division, but also in between divisions. This approach hence reduces the user's workload (i.e. only one neural network is trained per target variable as opposed to eight), and makes smooth transitions between divisions possible. However, it is likely that performance of the TDNN will decline due to an increase in area, and hence in the range of climatic conditions, to be covered by only one network; TDNNs specific to one study site division are likely to be more accurate.

5.2 Discussion

Monthly condensation rates are calculated from the output of the downscaling procedure together with the equations of M&G (2010a). These rates are then mapped. Donor and acceptor regions are clearly visible in the maps and have on average a ratio of 42% - 58%, respectively. This coincides with M&G's theory on condensation-induced dynamics with regards to two points: 1) there exists regions where, on average, no condensation occurs and others where condensation does occur, termed donor and acceptor regions, respectively, and 2) the ratio of the distribution of donor and acceptor region is approximately 50:50 (Makarieva and Gorshkov, 2010a). However, one would need to have a look at the global distribution of these dynamically allocated regions in order to verify the second statement fully.

Donor and acceptor regions show seasonal variation, with a minimum condensation rate in May and a maximum in October. However, no significant differences are found between the wet and dry seasons experienced by the Amazon basin over the course of the year. This coincides with the findings of Myneni *et al.* (2007) that state that evapotranspiration over the Amazon basin increases slightly during the dry season due to an increase in the vegetation's Leaf Area Index, which in turn would maintain condensation rates over the area roughly constant. The width of the regions extends from one gridbox to the whole of the study site (56 km and 1680 km, respectively). These properties of acceptor and donor regions are not mentioned in the BPT making comparison of these results with existing ones impossible. This further underlines the necessity for extending research in this particular branch of atmospheric physics.

Results suggest that most water vapour condensates between the surface up to an altitude of approximately 1250 m. Zonal winds are directed from East to West up to an altitude of 625 m and then switch to the opposite direction. Moreover, the curves of zonal wind speed and vertical velocity both decrease and increase together with that of condensation rate, showing correlation. It remains unclear however whether the increase in average condensation rate calculated over the dry season for the last 9 years could be directly related to deforestation and, as such, act as a significant factor in the increasing frequency of severe floods and droughts in the Amazon. Following the BPT, an increase in condensation would be linked to an increase in evaporation, which does not agree with M&G's view on the impacts of deforestation. This suggests that the relationship between condensation and evaporation is more complex.

Deforestation will ultimately lead to a decrease of both evaporation and condensation over the deforested area but this will first experience a period of increased evaporation and condensation. Densely-vegetated forests keep moisture below the canopy, and as such "protect" water vapour from evaporation. The only way for soil moisture to evaporate back to the atmosphere in these conditions is through plant transpiration; vegetation thus regulates the amount of moisture held in the soil. When deforested, this biotic control on soil moisture-atmosphere exchange will cease to exist and the area with the now bare soil evaporates more than before. Assuming a biotic regulation of the incoming water vapour through condensation-induced pressure changes, increased evaporation and condensation will increase the land-ocean pressure gradient and lead to greater attraction of water vapour from the oceans and consequently further increase evaporation and condensation. This positive feedback will remain functioning until the forest cover is too little to attract

sufficient water vapour from the ocean, at which point evaporation, and subsequently condensation, decreases. In the case that biotic regulation of water vapour does not occur, extensive deforestation will still lead to a decrease in evaporation and condensation since the inland of the continent is greatly dependent on the process of moisture recycling. The difference between both cases is the degree to which evaporation and condensation decrease.

Interestingly, this reasoning could explain the recent increase in the frequency of droughts and floods in the Amazon basin. Assuming that a) there exists biotic regulation of water vapour, b) temperature has increased because of deforestation (i.e. because of an increase in surface albedo from bear soil and an increase in greenhouse gasses such as water vapour and carbon dioxide), and c) the process is still in its first stage (i.e. increased evaporation and condensation), it is possible to envisage that higher temperatures and the increased evaporation result in higher atmospheric concentrations of water vapour due to an increase in the capacity for air to hold moisture. Consequently, water vapour will be kept in the atmosphere for longer periods, hence reducing soil moisture stores and increasing drought incidence. The water vapour will finally precipitate when reaching the highest altitudes of the troposphere, causing displaced and more intense rainfalls, which in turn may create floods. Following this reasoning, it is possible to conclude that deforestation will, in its first stage, be responsible for decreasing the frequency of rainfall but increasing its intensity, subsequently engendering droughts and floods.

5.3 Conclusions

Results have shown the presence and delimitations of donor and acceptor regions through their respective condensative properties, as well as clear correlations between horizontal and vertical winds, and condensation. However, results also show an increase in condensation rates over the dry season for the last 9 years. This, according to the BPT, would results from an increase in evaporation, which does not strictly adhere to M&G's views on the impacts of deforestation on precipitation. The findings hence support the existence of a biotic mechanism regulating the transport of water vapour as proposed by M&G. However, these outcomes also suggest a more complex relationship between evaporation and condensation, and hence highlight the need to further refine this novel theory. Other than assessing the existence of the BPT, results also provide for a refined insight into the interactions between vegetation and atmospheric processes such as evaporation and condensation, and further underline the necessity to promote research in this area.

5.4 Recommendations for future research

Although this study has shown promising results, it is important to continue the quantitative and impartial assessment of the BPT since many aspects of the theory remain yet to be evaluated. Future research should see the application of this methodology globally in order to identify acceptor and donor regions and determine their distribution at a global scale. One should aim at following the same approach but at an even higher resolution since much of the atmospheric process over tropical rainforest are very localized.

Efficiency and accuracy of artificial neural network prediction for the downscaling of atmospheric data should also be further improved. Research should focus on exploring approaches that will diminish or completely remove inhomogeneities between study site divisions and on comparing different neural network types and architectures. For example, such research could assess how a network with an architecture comprised out of eight outputs performs against that comprised of only one output such as those used in this study.

Another aspect to look at is the correlation between the evaporative-condensational force, condensation nuclei, vegetation density and condensation rate. This could provide for further explanations to the complex nature of the interactions between evaporation and condensation. The effects of latent heat released by condensation on air temperature and the uplifting of air due to topography should also be taken into account in a more complex model. Furthermore, changes in condensation rates should not only be examined over time (see Figure 4.4) but also spatially. Such results could for example give us more insight on the dependence of the *Intertropical Convergence Zone* (ITCZ) to condensation-induced pressure changes since variations in the ITCZ can drastically affect climatic conditions in equatorial regions engendering droughts and/or floods.

According to M&G (2010a), even reforesting areas affected by deforestation might not be enough to restore the fragile equilibrium developed by tropical rainforest over thousands of years as much of the physical properties of both vegetation and atmospheric dynamics remain unknown. In order to fully understand the interaction between vegetation and

water vapour, it remains imperative to tackle the issue of data availability and build high-resolution global datasets of atmospheric variables. As long as our observational data remains uncertain, so will our model predictions.

References

ASCE (2000a). Artificial neural networks in hydrology. I: Preliminary concepts. *Journal of Hydrologic Engineering* 5 (2), pp.115-123.

ASCE (2000b). Artificial neural networks in hydrology. II: Hydrologic applications. *Journal of Hydrologic Engineering* 5 (2), pp.124-137.

Avissar, R. and Werth, D. (2005). Global hydroclimatological teleconnections resulting from tropical deforestation. *Journal of Hydrometeorology* 6, pp.134-145.

Baboo, S. S. and Shereef, I. K. (2010). An efficient weather forecasting system using artificial neural networks. *International Journal of Environmental Science and Development* 1 (4), pp.321-326.

Bishop, C. M. (1996). *Neural networks for pattern recognition.* 1st ed. New York, USA: Oxford University Press.

Bonan, G. B. (2008). Forest and climate change: Forcing feedbacks and the climate benefits of forests. *Science* 320, pp.1444-1449.

Bustami, R., Bessaih, N., Bong, C. and Suhaili, S. (2007). Artificial neural network for precipitation and water level predictions of Bedup River. *International Journal of Computer Science* 34 (2).

Cannon, A. J. and Whitfield, P. H. (2002). Downscaling recent streamflow conditions in British Columbia, Canada using ensemble neural network models. *Journal of Hydrology* 259, pp.136-151.

Carter, T. R., Parry, M. L., Harasawa, H. and Nishioka, S. (1994). *IPCC technical guidelines for assessing climate change impacts and adaptations.* London, UK: University College London.

Cheruiyot, J. (2006). Performance of two types of meteorological balloons in upper air soundings in a tropical region. In: *WMO Technical Conference on Meteorological and Environmental Instruments and Methods of Observations*. Geneva, Switzerland.

CIIFEN (2012). *GCOS assessment of the status and needs for climate observations in South America 2003-2011.* Geneva, Switzerland: Centro Internacional para la Investigación del Fenómeno de El Niño.

Clark, D. B., Xue, Y., Harding, R. J. and Valdes, P. J. (2001). Modeling the impact of land surface degradation on the climate of tropical North Africa. *Journal of Climate* 14, pp.1809-1822.

Cochrane, M. A. and Laurance, W. F. (2002). Fire as a large-scale edge effect in Amazonian forests. *Journal of Tropical Ecology* 18, pp.311-325.

Coulibaly, P., Anctil, F., Aravena, R. and Bobée, B. (2001). ANN modeling of water table depth fluctuation. *Water Resources Research* 37, pp.885-896.

Coulibaly, P., Dibike, Y. and Anctil, F. (2005). Downscaling precipitation and temperature with temporal neural networks. *Journal of Hydrometeorology* 6, pp.483-496.

Davidson E. A., de Araújo, Artaxo P., Balch J. K, F., B. I., Bustamante M. M. C., Coe M. T., DeFries R. S., Keller M., Longo M., Munger J. W., Schroeder W., Soares-Filho B. S., M., S. J. C. and Wofsy, S. C. (2012). The Amazon basin in transition. *Nature* 481, pp.321-328.

Dibike, Y. and Coulibaly, P. (2006). Temporal neural networks for downscaling climate variability and extremes. *Neural Networks* 19, pp.135-144.

Douglas, M. W. (2008). Progress towards the development of the glidersonde: A recoverable radiosonde system. In: *WMO Technical Conference on Meteorological and Environmental Instruments and Methods of Observation*. St. Petersbourg, Russia.

El-Shafie, A. H., El-Shafie, A., El Mazoghi, H. G., Shehata, A. and Taha, M. R. (2011). Artificial neural network technique for rainfall forecasting applied to Alexandrai, Egypt. *International Journal of the Physical Sciences* 6 (6), pp.1306-1316.

Gautam, D. K. and Holz, K. (2000). Neural network based system identification approach for the modelling of water resources and environmental systems. In: *Proceedings of the Second Joint Workshop on Artifical Intelligence Methods in Civil Engineering Applications*.

Geman, S., Bienenstock, E. and Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural Computation* 4, pp.1-58.

Gettleman, A., Selkirk, H., Fujiwara, M., Shiotani, M., Hasebe, F., Alexander, J., Vömel, H., Hurst, D., Peter, T. and Morris, G. (2011). *Balloonsonde Tropical Tropopause Experiment (BATTREX) Science Plan.* Washington DC, USA: NASA.

Gimeno, L., Drumond, A., Nieto, R., Trigo, R. M. and Stohl, A. (2010). On the origin of continental precipitation. *Geophysical Research Letters* 37, pp.1-7.

Hayati, M. and Mohebi, Z. (2007). Application of artificial neural networks for temperature forecasting. *World Academy of Science, Engineering and Technology* 28, pp.275-279.

Haykin, S. (1998). *Neural networks: A comprehensive foundation.* 2nd ed. New Jersey, USA: Prentice Hall.

Haylock, M. R., Cawley, G. C., Harpham, C., Wilby, R. L. and Goodess, C. M. (2006). Downscaling heavy precipitation over the United Kingdom: A comparison of dynamical and statistical methods and their future scenarios. *International Journal of Climatology* 26 (10), pp.1397-1415.

Henderson-Sellers, A., Dickinson, R. E., Durbidge, T. B., Kennedy, P. J., McGuffie, K. and Pitman, A. J. (1993). Tropical deforestation: Modeling local- to regional-scale climate change. *Journal of Geophysical Research* 98 (D4), pp.7289-7315.

Hewitson, B. and Crane, R. G. (1996). Climate downscaling: Techniques and application. *Climate Research* 7, pp.85-95.

Hornik, K., Stinchcombe, M. and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks* 2, pp.359-366.

Hutyra, L. R., Munger, J. W., Nobre, C. A., Saleska, S. R., Vieira, S. A. and Wofsy, S. C. (2005). Climatic variability and vegetation vulnerability in Amazônia. *Geophysical Research Letters* 32 (L24712), pp.1-4.

Jarvis, A., Reuter, H. I., Nelson, A. and Guevara, E. (2008). *Hole-filled SRTM for the globe Version 4.* Cali, Colombia: International Centre for Tropical Agriculture.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetma, A., Reynolds, R., Jenne, R. and Joseph, D. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society* 77, pp.437-471.

Kivi, R., Kujanpää, J., Aulamo, O., Hassinen, S., Heikkinen, P., Calbet, X., Montagner, F. and Vömel, H. (2009). Observations of water vapor profiles over northern Finland by satellite and balloon borne instruments. In: *EUMETSAT Meteorological Satellite Conference*. Bath, UK.

Laurance, W. F. and Williamson, G. B. (2001). Positive feedbacks among forest fragmentation, drought, and climate change in the Amazon. *Conservation Biology* 15 (6), pp.1529-1535.

Lettau, H., Lettau, K. and Carlos, L. (1979). Amazonia's hydrologic cycle and the role of atmospheric recycling in assessing deforestation effects. *Monthly Weather Review* 107 (3), pp.227-238.

Lewis, S. L., Brando, P. M., Phillips, O. L., van der Heijden, G. M. F. and Nepstad, D. (2011). The 2010 Amazon drought. *Science* 331 (6017), p.554.

Liu, H., Chandrasekar, V. and Xu, G. (2001). An adaptive neural network scheme for radar rainfall estimation from WSR-88D observations. *Journal of Applied Meteorology* 40, pp.2038-2050.

Lorenz, E. N. (1967). *The Nature and Theory of the General Circulation of the Atmosphere.* Geneva, Switzerland: World Meteorlogical Organization.

Lorenz, E. N. (1983). A history of prevailing ideas about the general circulation of the atmosphere. *Bulletin of the American Meteorological Society* 64, pp.730-734.

Makarieva, A. M. and Gorshkov, V. G. (2007). Biotic pump of atmospheric moisture as driver of the hydrological cycle on land. *Hydrology and Earth System Sciences* 11, pp.1013-1033.

Makarieva, A. M. and Gorshkov, V. G. (2009a). Condensation-induced dynamic gas fluxes in a mixture of condensable and non-condensable gases. *Physics Letters A* 373, pp.2801-2804.

Makarieva, A. M. and Gorshkov, V. G. (2009b). Reply to A. G. C. A. Meesters et al.'s comment on "Biotic pump of atmopsheric moisture as driver of the hydrological cycle on land". *Hydrology and Earth System Sciences* 13, pp.1307-1311.

Makarieva, A. M. and Gorshkov, V. G. (2010a). The Biotic Pump: Condensation, atmospheric dynamics and climate. *International Journal of Water* 5 (4), pp.365-385.

Makarieva, A. M. and Gorshkov, V. G. (2010b). Potential energy of atmospheric water vapor and the air motions induced by water vapor condensation on different spatial scales. *arXiv:1003.5466v1 [physics.gen-ph]*, pp.1-37.

Makarieva, A. M., Gorshkov, V. G. and Li, B. (2009). Precipitation on land versus distance from the ocean: Evidence for a forest pump of atmospheric moisture. *Ecological Complexity* 6, pp.302-307.

Makarieva, A. M., Gorshkov, V. G. and Li, B. (2012). Revisiting forest impact on atmospheric water vapor transport and precipitation. *Theoretical and Applied Climatology*.

Makarieva, A. M., Gorshkov, V. G., Sheil, D., Nobre, A. D. and Li, B. (2010). Where do winds come from? A new theory on how water vapor condensation influence atmospheric pressure and dynamics. *Atmospheric Chemistry and Physics Discussions* 10, pp.24015-24052.

Marengo, J. A. (2006). On the hydrological cycle of the Amazon basin: A historical review and current state-of-the-art. *Revista Brasileira de Meteorologia* 21 (3), pp.1-19.

McCulloch, W. S. and Pitts, W. (1943). A logic calculus of the ideas immanent in nervous activity. *Bulletin of Mathemtical Biophysics* 5, pp.115-133.

Meesters, A. G. C. A., Dolman, A. J. and Bruijnzeel, L. A. (2009). Comment on "Biotic pump of atmospheric moisture as driver of the hydrological cycle on land" by A. M. Makarieva and V. G. Gorshkov, Hydrol. Earth Syst. Sci., 11, 1013-1033, 2007. *Hydrology and Earth System Sciences* 13, pp.1299-1305.

Mitchell, T. D. and Jones, P. D. (2005). An improved method of constructing a database of monthly climate observations and associated high-resolution grids. *International Journal of Climatology* 25 (6), pp.693-712.

Moore, N., Arima, E., Walker, R. and da Silva, R. R. (2007). Uncertainty and the changing hydroclimatology of the Amazon. *Geophysical Research Letters* 34 (L14707).

Myneni, R. B., Yang, W., Nemani, R. R., Huete, A. R., Dickinson, R. E., Knyazikhin, Y., Didan, K., Fu, R., Negrón Juárez, R. I., Saatchi, S. S., Hashimoto, H., Ichii, K., Shabanov, N. V., Tan, B., Ratana, P., Privette, J. L., Morisette, J. T., Vermote, E. F., Roy, D. P., Wolfe, R. E., Friedl, M. A., Running, S. W., Votava, P., El-Saleous, N., Devadiga, S., Su, Y. and Salomonson, V. V. (2007). Large seasonal swings in leaf area of Amazon rainforests. *Proceedings of the National Academy of Science* 104 (12), pp.4820-4823.

Penman, H. L. (1963). *Vegetation and Hydrology.* Farnham Royal, UK: Commonwealth Agricultural Bureau.

Phillips, O. L., Aragão, L. E. O. C., Lewis, S. L., Fisher, J. B., Lloyd, J., López-González, G., Malhi, Y., Monteagudo, A., Peacock, J., Quesada, C. A., van der Heijden, G., Almeida, S., Amaral, I., Arroyo, L., Aymard, G., Baker, T., Bánki, O., Blanc, L., Bonal, D., Brando, P., Chave, J., Cristina, Á., de Oliveira, A., Dávila Cardozo, N., Czimczik, C. L., Feldpausch, T. R., Aparecida Freitas, M., Gloor, E., Higuchi, N., Jiménez, E., Lloyd, G., Meir, P., Mendoza, C., Morel, A., Neill, D. A., Nepstad, D., Patiño, S., Cristina Peñuela, M., Prieto, A., Ramírez, F., Schwarz, M., Silva, J., Silveira, M., Sota Thomas, A., ter Steege, H., Stropp, J., Vásquez, R., Zelazowski, P., Alvarez Dávila, E., Andelman, S., Andrade, A., Chao, K., Erwin, T., Di Fiore, A., Honorio, E., Keeling, H., Killeen, T. J., Laurance, W. F., Peña Cruz, A., Pitman, N. C. A., Núñez Vargas, P., Ramírez-Angulo, H., Rudas, A., Salamão, R., Silva, N., Terborgh, J. and Torres-Lezama, A. (2009). Drought sensitivity of the Amazon rainforest. *Science* 323 (5919), pp.1344-1347.

Randel, W. J. and Wu, F. (2006). Biases in stratospheric and tropospheric temperature trends derived from historical radiosonde data. *Journal of Climate* 19, pp.2094-2104.

Salati, E. (1987). The forest and the hydrological cycle. In: Dickinson, R. E. (ed.) *The Geophysiology of Amazonia: Vegetation and Climate Interactions*. New York, USA: John Wiley & Sons, pp.273-296.

Salati, E. and Vose, P. B. (1984). Amazon basin: A system in equilibrium. *Science* 225 (4658), pp.129-138.

Sarle, W. S. (2002). *Neural Network FAQ, part 3 of 7: Generalization*. Available at: ftp://ftp.sas.com/pub/neural/FAQ3.html#A_over (Accessed: 27 September 2012).

Schneider, T. (2006). The general circulation of the atmosphere. *Annual Review of Earth and Planetary Science* 34, pp.655-688.

Sheil, D. and Murdiyarso, D. (2009). How forests attract their rain: An examination of a new hypothesis. *BioScience* 59 (4), pp.341-347.

Sherwood, S. C., Lanzante, J. R. and Meyer, C. L. (2005). Radiosonde daytime biases and late-20th century warming. *Science* 309 (5740), pp.1556-1559.

Shuttleworth, W. J. (1993). Evaporation. In: Maidment, D. R. (ed.) *Handbook of Hydrology*. New York, USA: McGraw-Hill, pp.4.1-4.53.

Smith, B. A., McClendon, R. W. and Hoogenboom, G. (2005). An enhanced artificial neural network for air temperature prediction. *World Academy of Science, Engineering and Technology* 7, pp.7-12.

Soares-Filho, B. S., Nepstad, D. C., Curran, L. M., Cerqueira, G. C., Garcia, R. A., Ramos, C. A., Voll, E., McDonald, A., Lefebvre, P. and Schlesinger, P. (2006). Modelling conservation in the Amazon basin. *Nature* 440, pp.520-523.

von Storch, H., Hewitson, B. and Mearns, L. (2000). Review of empirical downscaling techniques. In: *Regional Climate Development under Global Warming: Conference Proceedings RegClim Spring Meeting*. Jevnaker, Torbjørnrud, Norway.

Werth, D. and Avissar, R. (2002). The local and global effects of Amazon deforestation. *Journal of Geophysical Research* 107 (D20), pp.8087-8094.

Werth, D. and Avissar, R. (2004). The regional evapotranspiration of the Amazon. *Journal of Hydrometeorology* 5, pp.100-109.

Wigley, T. M. L., Jones, P. D., Briffa, K. R. and Smith, G. (1990). Obtaining subgrid scale information from coarse-resolution general circulation model output. *Journal of Geophysical Research* 95, pp.1943-1953.

WMO (1983). *Guide to Climatological Practices.* Geneva, Switzerland: World Meteorological Organzation.

Glossary

Adaptive learning function: Learning where a system programs itself by adjusting weights or strengths until it produces the desired output.

Artificial Neural Networks: A computational model based on the structure and functions of biological neural networks.

Buoyancy: The upward force exerted upon an air parcel (or any object) by virtue of the density (mainly temperature) different between the parcel and that of the surrounding air.

Cloud condensation nuclei: Tiny particles upon whose surfaces condensation of water vapour begins in the atmosphere.

Convection: Motions in a fluid that result in the transport mixing of the fluid's properties. In meteorology, convection usually refers to atmospheric motions that are predominantly vertical, such as rising air currents due to surface heating. The rising of heated surface air and the sinking of cooler air aloft is often called free convection.

Convergence: The notion that some functions and sequences approach a limit under certain conditions; in this case, the conditions set to limit the training of the artificial neural networks.

Data overfitting: This occurs when a statistical model describes random error or noise instead of the underlying relationship. A model which has been overfit will generally have poor predictive performance, as it can exaggerate minor fluctuations in the data.

Desertification: A general increase in the desert conditions of a region.

Dew-point: The temperature to which air must be cooled (at constant pressure and constant water vapour content) for saturation to occur.

Edge effects: Effect of the juxtaposition or placing side by side of contrasting environments on an ecosystem.

Error backpropagation: A method of training in that it uses gradient information to modify the network weights to decrease the value of the error function on subsequent tests of the inputs.

Evapotranspiration: Sum of evaporation and plant transpiration from the Earth's land surface to atmosphere.

Feed-forwarding: A multi-layer perceptron network in which the outputs from all neurons go to the owing but not preceding layers, so there are no feedback loops.

Global Climate Models: A mathematical model of the general circulation of a planetary atmosphere or ocean and based on the Navier-Stokes equations on a rotating sphere with thermodynamic terms for various energy sources (radiation, latent heat).

Intertropical Convergence Zone: The boundary zone separating the northeast trade winds of the Northern Hemisphere from the southeast trade winds of the Southern Hemisphere.

Isothermal: Of equal or constant temperatures.

Leaf Area Index: The ratio of total upper leaf surface of vegetation divided by the surface area of the land on which the vegetation grows.

Multi-Layer Perceptron: A network composed of more than one layer of neurons, with some or all of the outputs of

each layer connected to one or more of the inputs of another layer.

Partial pressure: Pressure of a gas if it alone would occupy the volume of a gas mixture.

Radiosonde : A balloon-borne instrument that measures and transmits pressure, temperature, and humidity to a ground-based receiving station.

Regional Climate Models: See Global Climate Model, but applied on a regional-scale.

Relative humidity: The ratio of the amount of water vapour in the air compared to the amount required for saturation (at a particular temperature and pressure). The ratio of the air's actual vapour pressure to its saturation vapour pressure.

Savannization: A general increase in the savannah conditions of a region.

South American Low Level Jet Stream: A narrow stream that channels the near-surface flow between the Tropics and mid-latitudes east of the mountain range. It is related to the transport of moisture from the Amazon region into the fertile lands of southern Brazil and northern Argentina.

Time-Delayed Neural Networks: These networks are similar to feed-forward networks, except that the input weight has a tap delay line associated with it. This allows the network to have a finite dynamic response to time series input data.

Transfer function: The relationship between the input and the output of a system, subsystem, or equipment in terms of the transfer characteristics.

Troposphere: The layer of the atmosphere extending from the earth's surface up to about 10 km above the ground.

Zonal wind: A wind that has a predominant east-to-west component.

Appendix I

Calculation of relative humidity

By definition, relative humidity RH (%) is the amount of water inside a mixture of air relative to the maximum amount that mixture can hold. It is defined as the partial pressure of water vapour p_v (mbar) relative to its saturation pressure p_{sat} (mbar; Tsonis, 2007):

$$RH = 100\frac{p_v}{p_{sat}} \quad [1]$$

, and

$$\begin{aligned}\log p_{sat} = {} & -7.90298\left(\frac{T_{stm}}{T}-1\right)+5.02802\log\left(\frac{T_{stm}}{T}\right) \\ & -1.3816\times10^{-7}\left(10^{11.344(1-T/T_{stm})}-1\right) \\ & +8.1328\times10^{-3}\left(10^{-3.49149(T_{stm}/T-1)}-1\right)+\log p_{stm}\end{aligned} \quad [2]$$

, where p_{stm} (mbar) is the steam-point pressure (1013.25 mbar), T (K) is air temperature and T_{stm} (K) is steam-point temperature (373.16 K).

Reference: Tsonis, A. A. (2007) Moist air. *An Introduction to Atmospheric Thermodynamics*. Second ed. New York, USA, Cambridge University Press.

Appendix II

Sounding the Amazon basin

(To be submitted for publication in the journal of the American Meteorological Society)

Authors:

Thomas Lafon[1,2]*, Jennifer Fowler[3], John Fredy Jiménez[4], Gabriel Jaime Tamayo Cordoba[4]

Affiliations:

[1]Fundación Entropika, Apartado Aéreo N° 20, Leticia, Amazonas, Colombia.

[2]Oxford Brookes University, Headington Campus, Gipsy Lane, Oxford OX3 0BP, UK.

[3]University of Montana, 32 Campus Drive, Missoula, MT 59812, USA.

[4]Instituto de Hidrología, Meteorología y Estudios Ambientales, Avenida Vásquez Cobo, Leticia, Amazonas, Colombia.

*Correspondence to: tlafon@entropika.org

Abstract:

Radiosonde-collected data over the Amazon rainforest are of vital importance to a wide variety of studies that aim at understanding the interaction of the forest's vegetation on local and global climatology. However, atmospheric measurements in the Amazon basin are sparse due to the lack of funding allocated towards collecting such data and meeting the standards set by the World Meteorological Organization. We here review current radiosonde technologies and an alternative that aims at lowering sounding costs by recovering the sondes: the glidersonde. Two major issues currently hamper future developments and commercialisation of this technology: 1) how to have reusable radiosondes while keeping the market viable for the sonde manufacturers and 2) the need for governmental aviation authorizations for flying glidersondes that are difficult to obtain. We conclude this review with an alternative consideration as an incentive for cooperation in the development and implementation of cost-effective sounding equipment.

One Sentence Summary:

Possible viable alternative for conducting cost-effective daily atmospheric soundings.

Main Text:

The importance of atmospheric soundings in the Amazon

Although the importance of the Amazon rainforest has been emphasized in previous research (e.g. 1, 2-4), the interaction of its vegetation on precipitation and global water

vapour circulation is still poorly understood (e.g. 5, 6-8). Marengo (9) attributes this large amount of remaining uncertainties to the “lack of basic hydrometeorological observational data systematically collected over time and space”. In order to study surface-atmospheric exchanges in the Amazon basin with a certain degree of confidence, it is imperative to have both ground data and the more costly upper-air observations, such as that of radiosondes and satellites.

While accuracy and coverage of satellite imagery have improved markedly in recent years, radiosonde-collected data still provides us with the most detailed measurements of the troposphere due to their fine vertical resolution (10, 11). However, biases in data (temperature, relative humidity, pressure, wind direction and wind speed) are compounded in the tropics by relying on such sparse, historic and continued, sampling of important variables (12, 13). Fig. 1 shows the locations of operational sounding stations that form the upper-air network (GUAN) of the Global Climate Observing System (GCOS) as recorded by the World Meteorological Organization (WMO) in February 2011. The figure also illustrates the scarcity of sounding stations in the Amazon basin compared to other parts of the globe.

Fig. 1. *Global positions of operational sounding stations (▲) as recorded by the WMO in February 2011. The Amazon basin is represented by the shaded area.*

As of 2011, only 50% of the South American GUAN stations transmitted regular reports (14). This lack of data is further highlighted by the GCOS goal for establishing a reference network for upper-air climate observations (GRUAN) as an extension to GUAN. GRUAN standards are more rigorous than that of GUAN and aim at addressing historic biases in the data. Of the 15 initial GRUAN sites, none were in South America. There is a current plan for expansion, as the importance of monitoring the Amazon is recognized, however the budgets to support this research have not yet materialized.

Why so few measurements?

The primary factor in the lack of sounding data availability is, quite simply, cost. Here, we take as example the Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM)

which, being the governmental institute for meteorological and hydrological services of Colombia, is responsible for recording the climate at a national level (15). In 2011, the meteorological station of the IDEAM in Leticia, based in the southern tip of the Colombian Amazon, received funding for a total of 119 flights, while the recommended amount is about 730 flights a year (i.e. two flights a day; 16).

It is important to recognize that this case study is representative of issues that are recurrent in developing countries of Africa and South America, where budgets allocated to soundings is often limited or a minute part of a bigger budget (e.g. 17, 18). This engenders the inability of a vast majority of developing countries to comply with the standards set by the WMO in their 2007 revision of the WMO Convention originally published on the 11th October 1947.

Current radiosonde technologies

Sensor accuracy has made improvement over the years, however the standards for long-term climate studies require even higher accuracy and more rigorous documentation of instrument changes in order to reduce temporal and spatial inhomogeneity (19). Working towards correcting these issues is WMO's GRUAN programme, formed to provide long-term, high quality and accurate climate data. The 2010 WMO evaluation of radiosonde performance quantitatively ranked radiosonde technologies against GRUAN requirements (20). The top three companies closest to meeting these standards are Vaisala (Finland), GRAW (Germany), and Lockheed Martin Sippican (USA).

All three manufacturers have pricing levels that depend on the specifications of the radiosonde sensors. Assuming that a) the sondes are being acquired by a large agency such as the WMO or a national weather institute and b) a quantity corresponding to two flights a day per year is being ordered, the price is on average 155 USD per sonde. Here, we do not consider the cost of shipping, which may or may not be included in the sonde cost. For new stations, or stations that are changing equipment, additional charges will result from buying the corresponding receiving device (i.e. used in the collection of the data radio-transmitted by the sonde), which ranges from 12.500 USD to 80.000 USD.

Considering only well-established sounding stations, and therefore discounting the costs of the receiving devices, lifting gas and infrastructure, each station will spend approximately 113.000 USD per year on sondes alone; this does not include expenses for balloons, lines and parachutes (see Fig. 2). Costs of sondes increase by as much as 35% when quantities of supplies decrease.

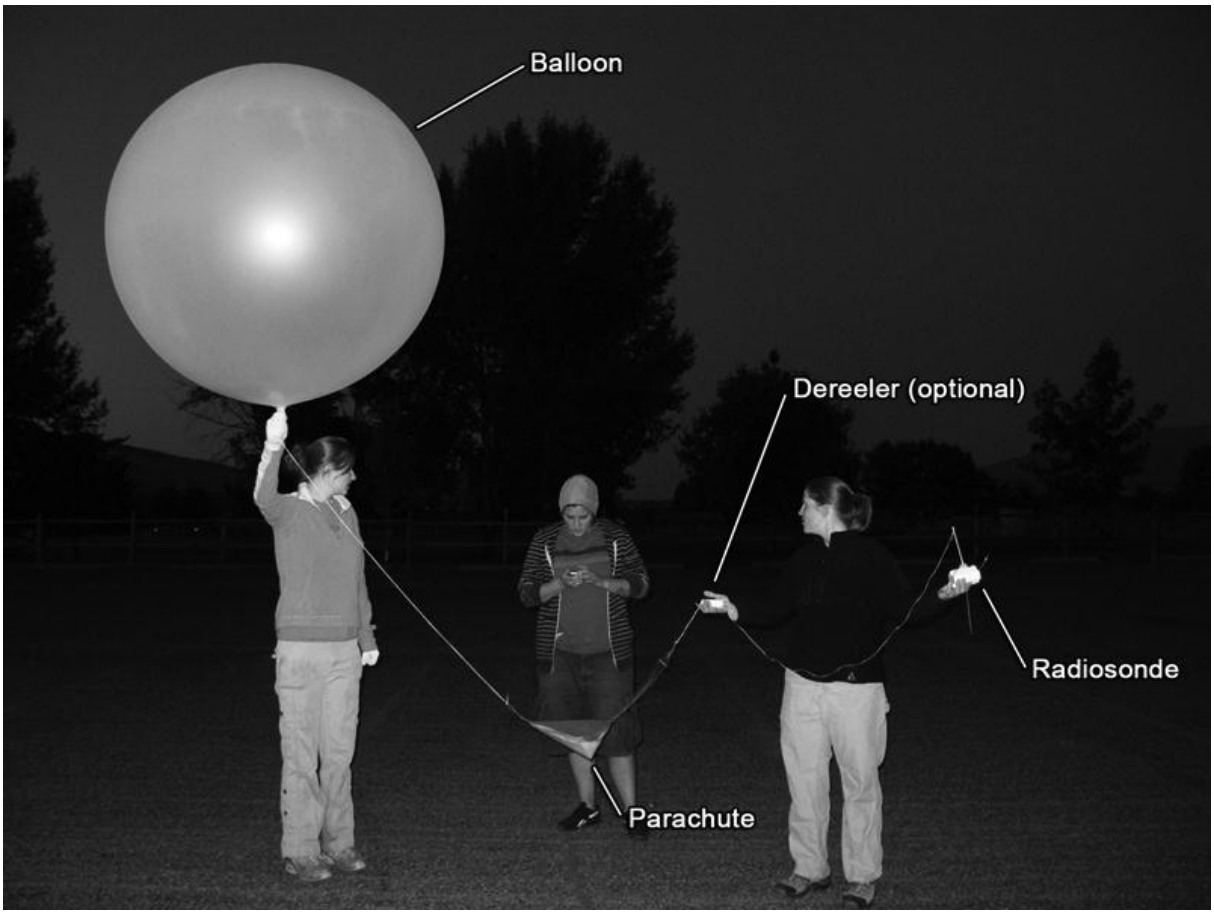


Fig. 2. *Set up of the GRAW sounding equipment for a flight.*

Possible alternative: glidersondes

An alternative to lower the costs of conducting daily soundings would be to have the radiosonde come back and be reused. Indeed, once the balloon has been launched it is often impossible to get the equipment back; the sonde can drift tens of kilometres away (21). The issue of recovering the radiosonde becomes even more prominent in parts of the globe where topography and/or land-cover render the landing site difficult to access (e.g. the dense vegetation of the Amazon rainforest).

With this objective in mind, the development of a new technology was initiated in 1997 by the National Severe Storm Laboratory (USA): the radiosonde would glide back to a designated landing point. This recoverable system sees the lifting of a miniature plane (the "glidersonde") using the same balloon and lifting gas method as that of a regular sounding flight. Once the balloon reaches its upmost altitude and pops, the plane free-falls until stabilizing and then glides back to a designated point using an onboard GPS and flight navigation computer.

Fig. 3. *The "DataBird", glidersonde model developed by GPSBoomerang.*

A recent model developed by GPSBoomerang (New Zealand), the “DataBird” (see Fig. 3), has shown promising results with the possibility of ascending to 35km and coming back within a 100m radius of the predetermined landing site (22). Unfortunately, glidersonde devices are still prototypes and therefore not readily available on the commercial market. Other than the lack of funding, a major issue currently impeding their development are strict air space restrictions exemplified in the US with Federal Aviation Administration (FAA) regulations: a Certificate Of Authorization (COA) is imperative but difficult to obtain as a glidersondes are considered to be Unmanned Aircraft Systems (UAS) and hence fall in the same category as military drones. In general, International Civil Aviation Organization rules should be considered; their document on UAS use similar requirements as the US FAA guidelines to acquire certificates (23).

Future developments

To conclude this review we here underline two major issues that currently hamper the development and commercialisation of the glidersonde technology: 1) how to have reusable radiosondes while keeping the market viable for sonde manufacturers and 2) the need for permits that are difficult to obtain for flying glidersondes.

We propose an alternative consideration to address the first issue, namely that of providing users with reusable equipment while maintaining manufacturer’s market share. One way to achieve this is to develop glidersonde systems that fit current radiosonde models and have the radiosondes be reusable for the number of flights needed to break even on the cost of

the glider, plus make an additional cost savings over the disposable radiosondes. For example, a package of six DataBird modules, capable of going into the stratosphere, costs approximately 3.000 USD. This would require each module to complete four flights with one reusable radiosonde to break even with current radiosonde costs and realize an additional 23% cost savings. In this way, the total budget spent on radiosondes could remain consistent while the availability of data increases. Applying this to our case study, the IDEAM of Leticia in the year 2011 would have purchased the same amount of radiosondes, been able to use them for four times as many flights (with the proposed glider system), and hence collect quadruple the amount of atmospheric data while having their sonde provider maintaining their sales level. In developed countries, this increase in flights could lead to a greater temporal resolution of the data.

It could also be hypothesized that such properties of reusability will attract more occasional users. Additionally, this addresses the issue mentioned by Douglas (18), namely the need for the sensors on the glider to be periodically reconditioned and replaced due to possible damages caused by landing and reiterative use; these will be replaced every four flights, assuming the glider is capable of consistent performance. Furthermore, recovering the sonde reduces the environmental impact caused by the non-biodegradable equipment of the radiosonde.

In order to overcome the second issue, that of overly strict aviation regulations, it is imperative to broaden interagency and private industry cooperation. Although in the US the FAA streamlined the permit process for its government partners in March 2012 (24), private companies cannot obtain a COA; issuing permits such as that of the COA to meteorological services has been made straightforward but the current process is such as to deter private

companies from developing the necessary glidersonde technology. It should be clearly understood that the technology and acquired data sets are equally beneficial for all involved parties.

We urge cooperation in the development of an operational glidersonde system and in the collection of more extensive and continuous data over areas such as the Amazon basin that are of vital importance to global climate. Climate models are key and powerful tools with which most future predictions are achieved today and we strongly believe their simulation and prediction performances will benefit from refined atmospheric data sets. However, as long as our observed data remains uncertain, so will our model predictions.

References:

1. Soares-Filho, B.S., D.C. Nepstad, L.M. Curran, G.C. Cerqueira, R.A. Garcia, C.A. Ramos, E. Voll, A. McDonald, P. Lefebvre, P. Schlesinger, Modelling conservation in the Amazon basin. Nature, 440, 520-523 (2006).
2. Werth, D., R. Avissar, The local and global effects of Amazon deforestation. Journal of Geophysical Research, 107, 8087-8094 (2002).
3. Gedney, N., P.J. Valdes, The effect of Amazonian deforestation on the northern hemisphere circulation and climate. Geophysical Research Letters, 27, 3053-3056 (2000).
4. Poveda, G., D.M. Álvarez, Ó.A. Rueda, Hydro-climatic variability over the Andes of Colombia associated with ENSO: A review of the most critical hotspot for biodiversity on Earth. Climate Dynamics, 36, 2233-2249 (2010).

5. Avissar, R., D. Werth, Global hydroclimatological teleconnections resulting from tropical deforestation. Journal of Hydrometeorology, 6, 134-145 (2005).
6. Makarieva, A.M., V.G. Gorshkov, B. Li, Precipitation on land versus distance from the ocean: Evidence for a forest pump of atmospheric moisture. Ecological Complexity, 6, 302-307 (2009).
7. Meesters, A.G.C.A., A.J. Dolman, L.A. Bruijnzeel, Comment on "Biotic pump of atmospheric moisture as driver of the hydrological cycle on land" by A. M. Makarieva and V. G. Gorshkov, Hydrol. Earth Syst. Sci., 11, 1013-1033, 2007. Hydrology and Earth System Sciences, 13, 1299-1305 (2009).
8. Myneni, R.B., W. Yang, R.R. Nemani, A.R. Huete, R.E. Dickinson, Y. Knyazikhin, K. Didan, R. Fu, R.I. Negrón Juárez, S.S. Saatchi, H. Hashimoto, K. Ichii, N.V. Shabanov, B. Tan, P. Ratana, J.L. Privette, J.T. Morisette, E.F. Vermote, D.P. Roy, R.E. Wolfe, M.A. Friedl, S.W. Running, P. Votava, N. El-Saleous, S. Devadiga, Y. Su, V.V. Salomonson, Large seasonal swings in leaf area of Amazon rainforests. Proceedings of the National Academy of Science, 104, 4820-4823 (2007).
9. Marengo, J.A., On the hydrological cycle of the Amazon basin: A historical review and current state-of-the-art. Revista Brasileira de Meteorologia, 21, 1-19 (2006).
10. Kivi, R., J. Kujanpää, O. Aulamo, S. Hassinen, P. Heikkinen, X. Calbet, F. Montagner, H. Vömel, "Observations of water vapor profiles over northern Finland by satellite and balloon borne instruments" in EUMETSAT Meteorological Satellite Conference (Bath, UK, 2009).
11. Gettleman, A., H. Selkirk, M. Fujiwara, M. Shiotani, F. Hasebe, J. Alexander, H. Vömel, D. Hurst, T. Peter, G. Morris, "Balloonsonde Tropical Tropopause Experiment (BATTREX) Science Plan" (NASA, Washington DC, USA, 2011).

12. Sherwood, S.C., J.R. Lanzante, C.L. Meyer, Radiosonde daytime biases and late-20th century warming. Science, 309, 1556-1559 (2005).
13. Randel, W.J., F. Wu, Biases in stratospheric and tropospheric temperature trends derived from historical radiosonde data. Journal of Climate, 19, 2094-2104 (2006).
14. CIIFEN, "GCOS assessment of the status and needs for climate observations in South America 2003-2011" (Centro Internacional para la Investigación del Fenómeno de El Niño, Geneva, Switzerland, 2012).
15. WMO, "WMO Statement on the Role and Operation of National Meteorological and Hydrological Services for Directors" (Cg-XVI, World Meteorological Organization, Geneva, Switzerland, 2007).
16. GCOS, "Report of the Third GCOS Reference Upper Air Network Implementation and Coordination Meeting (GRUAN ICM-3)" (World Meteorological Organization, Geneva, Switzerland, 2011).
17. Cheruiyot, J., "Performance of two types of meteorological balloons in upper air soundings in a tropical region" in WMO Technical Conference on Meteorological and Environmental Instruments and Methods of Observations (Geneva, Switzerland, 2006).
18. Douglas, M.W., "Progress towards the development of the glidersonde: A recoverable radiosonde system" in WMO Technical Conference on Meteorological and Environmental Instruments and Methods of Observation (St. Petersbourg, Russia, 2008).
19. GCOS, "GCOS Reference Upper-Air Network (GRUAN): Justification, requirements, siting and instrumentation options" (GCOS-112, WMO/TD No.1379, World Meteorological Organization, Geneva, Switzerland, 2007).
20. Nash, J., T. Oakley, H. Vomel, L.I. Wei, "WMO intercomparison of high quality radiosonde systems" (107, World Meteorological Organization, Yangjiang, China, 2011).

21. Seidel, D.J., B. Sun, M. Pettey, A. Reale, Global radiosonde balloon drift statistics. Journal of Geophysical Research, 116 (2011).
22. GPSBoomerang, DataBird-300 Specifications (cited 21 May 2012, http://www.gpsboomerang.com/content/view/16/30/).
23. ICAO, "Unmanned Aircraft Systems (UAS)" (Cir 328 AN/190, International Civil Aviation Organization, Montreal, Canada, 2011).
24. Turner, A. FAA streamline unmanned permit process. Air Traffic Management (published 15 May 2012, http://www.airtrafficmanagement.net/2012/05/us-streamlines-unmanned-permit-process/).

Acknowledgments:

This work was partly funded by the Montana Space Grant Consortium. We would like to thank Antonio Garcia-Mendez from the European Centre for Medium-Range Weather Forecasts for the global list of operational sounding station, GPSBoomerang for the picture of the DataBird glidersonde and the personnel of the Instituto de Hidrologia, Meteorologia y Estudios Ambientales of Leticia for their participation in this review.

Appendix III

Other research conducted over the MPhil period

("Bias correction of daily precipitation simulated by a regional climate model: a comparison of methods"

Accepted on 28th April 2012 and published online in the International Journal of Climatology)