
PhD's thesis cycle XXXII

---

# Neural-Network modelling for meteorological and climatological applications

Stefano Amendola

PhD Program Coordinator
Prof. Giuseppe De Grassi

Supervisor
Prof. Elena Pettinelli
Prof. Antonello Pasini

Department of Mathematics and Physics
Roma Tre University
Rome, Italy 2020

## Acknowledgments

*I would like to thank my supervisors, Antonello Pasini and Elena Pettinelli, for all the help received during this three-year period. In particular, I would like to thank Antonello for his patience, motivation, enthusiasm, immense knowledge and mostly for his friendship.*

*A special thank is for my wife Laura and our daughter Caterina, without them all this would not have been possible.*

# Index of contents

# 1. Introduction

During the last decades the field of the physics of the atmosphere obtained a rising interest by research centres worldwide and by the general public. Meteorology and climate are of primary importance for our lives, just think about the influence on sectors as civil protection, agriculture, tourism, transport or industry. Furthermore, climate change is a matter of deep discussion for both scientists and policymakers [1].

The study of climate requires long data series, so this science rapidly grows hand by hand with the availability of meteorological data, supplied by weather stations, atmospheric soundings, ships, buoys, satellites and so on. Furthermore, thanks to the improvement of computer capabilities, in the second part of the last century meteorological-climatic models were born [2]. They can be considered as a virtual laboratory in which the (complex) climate system is reproduced – with all its subcomponents and related interactions – by the equations of fluid dynamics and thermodynamics. In the application of these models we must keep in mind the limitations typical of numerical simulation of chaotic systems [3]. In general, topics as weather, monthly and seasonal forecasts, climate projections and related impacts are approached by dynamical modelling. In some respects, for example for climate change attribution – i.e. the effort to scientifically establish causes for the recent warming of the Earth – this approach is a matter of discussion. In fact, the problem of attribution is mainly addressed by the so-called Global Climate Models (GCMs) [4] – a type of meteo-climatic model – and, although in last decades great improvements are achieved, these models still have some limitations.

In particular, the use of GCMs for attribution only partially allows to apply a robustness scheme. Robustness of the results relies on a condition of independence among the different models employed, i.e. reliability improves if the same results are obtained using different models, if independence between them holds. As in many cases GCMs have a common ancestor, the condition of independence is questioned [5]. Furthermore, the unavoidable abstraction and idealization included in the physics of dynamical models does not make them able to completely represent the climate system. Consequently, in order to obtain stronger results, data-driven models have been used in atmospheric sciences [6]. Among others, we underline the application of neural network (NN) analysis [7] and Granger causality [8], which have been shown to be particularly useful in many applications.

A multi-approach strategy permits to satisfy the just described condition of robustness: results obtained by different ways may complement each other and data-driven models can also improve the performance of dynamical models [6]. Based on these arguments, data-driven approaches have found large application for the study of different topics related to climate and meteorology.

An example is the application of Granger Causality (GC, [8]) to the atmospheric sciences. GC establishes a criterion to find causal relations between time series. It is based on vector autoregressive (VAR) models. In recent years different studies in atmospheric sciences based on GC have been developed ([9], [10], [11], [12]).

However, our interest is principally focused on another type of data-driven approach, i.e. neural network analysis, an artificial intelligence method. NNs can find nonlinear relationships between a set of predictors and a fixed target. The use of different set of predictors permits to investigate the causes that may have generated the behaviour of the target. Several works that make use of neural network analysis for climatic topics can be find in the scientific literature ([13], [14], [15], [16]). In many cases in atmospheric sciences we are faced with relatively short data sets. So, these problems are not deal with deep learning as for big data topics. In particular, a tool for small datasets has been recently developed [17].

Another field of application of statistical modelling is that of meteorological and/or seasonal forecasts. Usually dynamical models are used to obtain this kind of forecasts. Also, in this case some limitations exist, so one step further is necessary to improve forecasts, especially for local areas. Numerical models for meteorological/seasonal forecasts can suffer for their poor resolution and for the necessary oversimplifications that do not permit to well represent physical mechanisms at the basis of the observed variability. So *downscaling* techniques have been developed in order to solve these problems. Downscaling can be divided in dynamical and statistical ([18], [19]). The first one is based on the use of dynamical models with improved resolution and parameterizations, for example Regional Climate Models [20], to be nested into global models. The latter uses statistical techniques, in particular past data are analysed in order to find a relationship able to improve the performance of GCMs [19]. So, in this case we need long series of data. Often statistical downscaling makes use of neural network analysis, as for example in a model recently developed for seasonal forecast over the Italian peninsula [21], or for the downscaling of GCMs to a spatial scale comparable to that of RCMs [22]. Despite of its relative simplicity, statistical downscaling often overcome dynamical downscaling performances [19].

Finally, as last example about the use of neural network models in climate sciences, we mention some works that make use of NN models in order to obtain a "point forecast", i.e. forecast for a very restricted area (at limit for one point) for a particular meteorological event. Pasini et al. [23] developed a model for forecasting local visibility over Northern Italy. Marzban focuses on the application of neural network models for phenomena as tornadoes in a series of papers ([24], [25], [26], [27], [28]).

In the light of what has been said, the objective of this PhD thesis is to develop a neural network tool able to investigate some open topics about atmospheric sciences. In the context of the rapid changes to which we are witnessing, physical information of great usefulness can be achieved. We will see that this tool can be applied to a wide range of topics, covering a wide spatial scale from global to local. In fact, we will face with topic such as climate change attribution (Chapter 3), the analysis of a general circulation problem related to the so-called Atlantic Multidecadal Oscillation (Chapter 4) and, finally, the analysis of impacts related to climate/weather conditions (Chapters 5 and 6). Furthermore, the consideration of multi-linear regression analysis parallelly to the neural network tool will be useful to both underline the goodness of the choice of a non-linear method and to analyse the importance of linear or non-linear mechanisms.

So, this thesis is structured in the following way:

In **Chapter 2** we introduce the climate system. Here we furnish some generalities about climate, with considerations on its energy budget and the subcomponents in which it can be divided – obviously from a theoretical point of view. Then we introduce also the hot topic of climate change, related to the greenhouse effect. We conclude the first Chapter by introducing also the item of weather and climate modelling.

In **Chapter 3** we will explain the Neural Network tool. Due to the central importance of neural analysis for this thesis, we start with a historical explanation of the development of this kind of method, from the origins up to the typology related to this research. It is a technical Chapter, useful to fix the idea about neural network analysis. It is closed by a section that explains in detail the tool developed during this three-year activity, with its peculiarities and novel characteristics.

In **Chapter 4** we start to show the results obtained with our analysis. Here we will show the application to the debated theme of attribution of the mean global temperature behaviour on the last 150 years. We find a lot of interesting results with the confirmation of results obtained with a completely independent mean of investigation. Our tool will also be able to investigate about the

causes that have determined the behaviour of global temperature in sub-intervals into the considered period.

**Chapter 5** is dedicated to the analysis of the causes behind the behaviour of the Atlantic Multidecadal Oscillation (AMO). AMO is a mode of variability of the sea surface temperature anomalies over the North Atlantic Ocean. For many years it has appeared as a natural component of the climate system, but recently several works linked it to anthropogenic activity appears. We insert in this open debate with a completely different approach: in this chapter we will show interesting results.

In **Chapter 6** we will show the first application of our neural network tool to study impacts related to weather/climate conditions. Here we focus on the analysis of the causes that could influence forced migrations from the Sahelian countries to Italy. It is a hot topic in the political agenda of many European countries. We will see impressive results, that relate climate variables and harvest yields to the migration rate to Italy.

Finally, in **Chapter 7** we conclude this work with an application of our tool to studying the impact of meteorological variables on the dynamics of the observed quantity of a kind of sandfly responsible for the leishmanias spread. In fact, in recent years we observed an unusual expansion of such sandflies over areas previously considered immune. Interesting results are achieved, showing the power of our model.

# 2. The climate system

An introduction about the physics of the climate system is necessary to face with the themes discussed in this thesis. Climate is defined as the average weather conditions on a long period for a certain area. As prescribed by the World Meteorological Organization (WMO), at least 30 years of data are necessary to define a climatic mean. The availability of long-time data series for meteorological parameters as temperature, humidity, wind, cumulated rainfall and so on, provides information about the climate. In these terms climate might seem something changeless and stationary but, it is not so. In fact, climate has changed – due to different causes – in all the history of our planet and it will continue to change. The well-known recent *climate change* is an example of these changes.

Starting from a historical point of view, the study of the human environment (our Earth) and of the broader astronomic environment (the Sun, the Moon, the planets) has given the first strong impulse to the birth of the science of physics. Afterwards, for a long period the earth sciences have had less influence in comparison to the physics of the "infinitesimal" (quantum mechanics) and that of the "infinite" (astronomy). During last century, due to both theoretical and technological developments, earth sciences – and in particular atmospheric sciences – obtained new attention by the physical and the mathematical communities. The work by Lorenz of 1963 [3] on the chaotic behaviour shown by simple deterministic nonlinear flows may be considered as the Rosetta Stone for the development of the modern physics of climate. Also, the availability of a rising amount of measurements of the various meteorological parameters was a fundamental ingredient for this development. Furthermore, the birth of computing has started the field of numerical modelling for weather and climate forecasts. Thanks to all these achievements, during the last decades the physics of the atmosphere has obtained very important results and a continuously rising interest. In fact, weather and climate are fundamental actors in the everyday life of each of us, just think about on sector as agriculture, civil protection, industry and tourism.

So, in the following section an introductive description of the climate system is furnished while in section 2.2 we introduce the topic of weather and climate dynamical modelling.

## 2.1 Climate system description – physics, feedbacks, equilibrium states.

In order to understand the complexity of the climate system just look at Figure 1. The main driver in the climate system is the Sun, in fact, it is thanks to its different heating between the equatorial areas and the poles that the atmospheric dynamics may exist.

The climate system must be considered as a unique structure, but for scientific purposes it is convenient to idealize it as composed by sub-systems that continuously interact each other. These are the *atmosphere*, the *hydrosphere*, the *cryosphere*, the *lithosphere* and the *biosphere*. The *atmosphere* is the aeriform part, from the Earth surface up to the open space. The part strictly involved in the observed weather phenomena is the troposphere. Extended up to about 11 kilometres, it is the densest portion of the atmosphere and it contains most of the water vapour. The *hydrosphere* regards the liquid water on our planet. Oceans, lakes, rivers, subterranean water strongly influence the system on a wide spatial and temporal scales. Just think about at the heat storage role of the oceanic masses or at the local influence of a lake. The *cryosphere* is related to the water in its solid phase (ice). Mountain and land ices, marine ices, permafrost are all part of this component. The *lithosphere* consists in the rocky surface of the Earth and its manifestations (volcanoes in particular). Finally, the *biosphere* is all that concern the lively things that actively interact with the climate system.

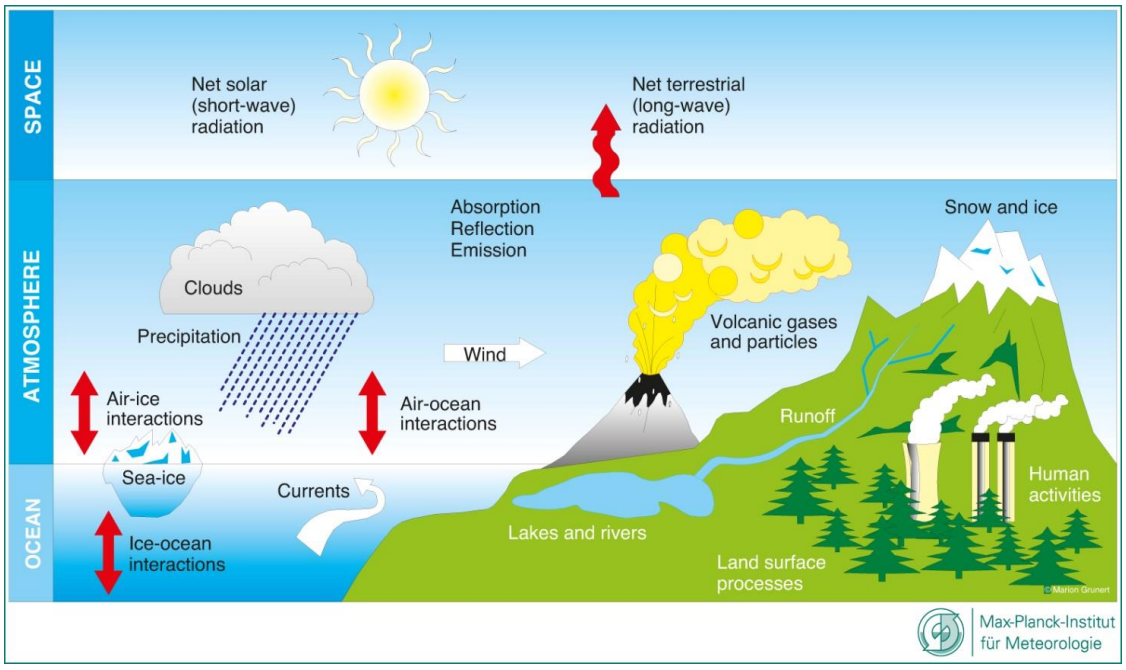


**Figure 1 Representation of the climate system components and their interactions, see details into the text. From Max Planck Institute for Meteorology.**

These five components – as shown in Figure 1 – are strictly related, with a continuous exchange of energy and mass due to heat fluxes, matter phase changes, radiative transfer, chemical reactions, anthropogenic emissions, and so on. In order to understand all these physical interactions, the wealth of knowledge must be large. This is understandable by the following list of some of the physical phenomena related to the climate system – starting from the sun, up to the microphysics of the clouds – and the knowledge necessary to understand them:

- the inner processes of the Sun, necessary to explain its electromagnetic spectrum;
- the interactions between the solar radiation – i.e. the photons – and the Earth atmosphere, including the radiative transfer and the effects on the atmospheric chemical composition (named photochemistry);
- the processes of radiative transfer: transmission, absorption, emission;
- the dynamical processes of the atmosphere and the oceans;
- the thermodynamic processes that happen in the atmosphere and oceans;
- the microphysics of the clouds;
- the part inherent to the measurements of the meteorological parameters – weather stations, buoys, probe balloons, remote sensing (radar, satellite measurements, ...) and so on;
- the part related to the numerical modelling, for the past climate reconstructions and forecasts.

In addition to these physical processes, the climate system owns a peculiarity related to the presence of many *feedback* mechanisms. In fact, the modification of a parameter can bring to a series of changes in the system up to a further alteration of the parameter itself. We now fix the idea with two simplified examples. Just think to a rising in the mean global temperature. This variation leads to an increased evaporation from the hydrosphere that in turn leads to a greater cloudiness in the atmosphere. The result is a decrease in the solar radiation at the surface due to this cloudiness with a consequent temperature reduction. This is a typical example of *negative feedback* – i.e. an attenuation of the initial variation. On the other hand, for example, suppose a decreasing of the polar ices due to melting. This leads to the decreasing of the surface albedo and so to a higher solar radiation absorption by the surfaces under the melted ice. The final consequence is a rising in the air temperature with an even larger ice melting. This is a *positive feedback* – an improvement of the initial variation. These two simple examples of feedbacks should be enough to understand the complexity and the variety of the processes involved in the "construction" of the climate system.

Furthermore, in the climate system multiple quasi-stable equilibrium states may exist. At first, assume that we can calculate an equilibrium state starting from certain fixed boundary conditions. The question is if that state is stable or not. Generally, a complex system with feedbacks shows a chaotic behaviour and climate lies in this category. We remember that a chaotic behaviour for a system means that we are not able to predict its state for a long time ahead. In fact, even if we know with arbitrary accuracy the initial state, after a certain time the system loses the knowledge of its initial conditions and goes toward unknown solutions. In the context of a chaotic behaviour, climate system can also have quasi-stable equilibrium states. We know that a stable state is one in which a small perturbation brings in any case to the restore of the initial conditions. On the other hand, in a quasi-stable state even a small perturbation can lead to an abrupt change of the state of the system. In summary, similar boundary conditions may mean very different states of equilibrium, it is a real trouble in the case a strong modification of one or more of the boundary conditions happens (as is the case for the anthropogenic emissions over last 150/200 years – see section 2.1.4).

In this context, appears clear that modelling and forecasting meteorological and climate dynamics is a formidable challenge. In the section 2.2 we deal with the issue of modelling; meanwhile other generalities about the climate system are supplied. Again, we underline that here only a preliminary introduction on the matter is furnished; for more quantitative details we refer to two texts – [29] and [30].

### 2.1.1 The Earth energy budget

The sun is the main source of energy in the climate system. The engine for the atmospheric motion is due to the differential heating of the Earth surface by the solar radiation. The solar spectrum presents its maximum at the visible wavelengths (VIS), with a wide percentage of radiation reaching the Earth also in the ultraviolet (UV) band. In this context the solar radiation is also called Short-Wave radiation (SW). At the mean Earth-Sun distance arrives an average amount of about 1370 $W/m^2$: it is the so called *solar-constant*. This quantity is not really constant, as the orbit of the Earth with its eccentricity leads to an annual variation of the incident solar radiation on the Earth surface – about 1435 $W/m^2$ on January and about 1345 $W/m^2$ on July. This difference is not relevant for everyday life, as the seasonal cycle is dominated by the tilt of the Earth rotation axis. Also, the eleven-years sunspot cycle leads to slight variations of the solar radiation. Starting from the instrumental era the observed variation of solar radiation is a very small fraction of the total (lower than 1%).

A fundamental variable related to the incoming solar radiation is the percentage that is absorbed by our planet. In average, the fraction that is reflected or back-scattered towards the space – i.e. not absorbed – is about 0.30 of the totals, this quantity is the so-called *albedo* (a perfect reflector has unitary albedo, while a perfect absorber has null albedo). The remaining fraction of the solar radiation is available as energy for the climate system. In particular, about 20% of the total is absorbed by the atmosphere (principally in the stratosphere by the Ozone layer) and the remaining 50% is absorbed at the surface. Obviously, these numbers are indicative as the real-time situation may be very different due mainly to variations in the cloud cover of the Earth. Since a real-time measure of the planetary albedo is a difficult issue, we do not have a long-term series of data and the actual trend is unknown.

At equilibrium – considering long time period – the absorbed energy by the climate system must be equal to the energy released towards the space. The latter is due to the radiation re-emitted by the Earth surface and by the atmosphere: it has its maximum in the infrared part of the electromagnetic spectrum – it is generally called Long-Wave (LW) radiation. Furthermore, as we have more solar radiation at low latitudes respect to the polar areas while the terrestrial re-emission happens not in the same way (Figure 2), some mechanism of transport of energy must exists. The transport of this equatorial surplus of energy is due in equal percentage to both atmospheric and oceanic circulation.

Detailed knowledge of the energy budget and of the related temperature vertical profile requires a deep knowledge of the radiative transfer processes. These are based on Planck's radiation law. Application of radiative transfer theory permits to: better understand sources and sinks of energy in the atmosphere; to simplify concepts as the atmosphere in radiative equilibrium; and to understand the energy budget at top of the atmosphere (TOA). About the latter point, the knowledge of the wavelength spectrum of the LW radiation leaving the Earth is fundamental for the energy balance. In fact, each gaseous component of the atmosphere interacts with LW radiation depending on the wavelength of the radiation itself.

Some gaseous species are responsible for the well-known *greenhouse effect*, the process to which is due the habitability of the Earth. In fact, these gases "trap" part of the LW radiation, with the result to increase the global mean temperature (see section 2.1.4 for more details).

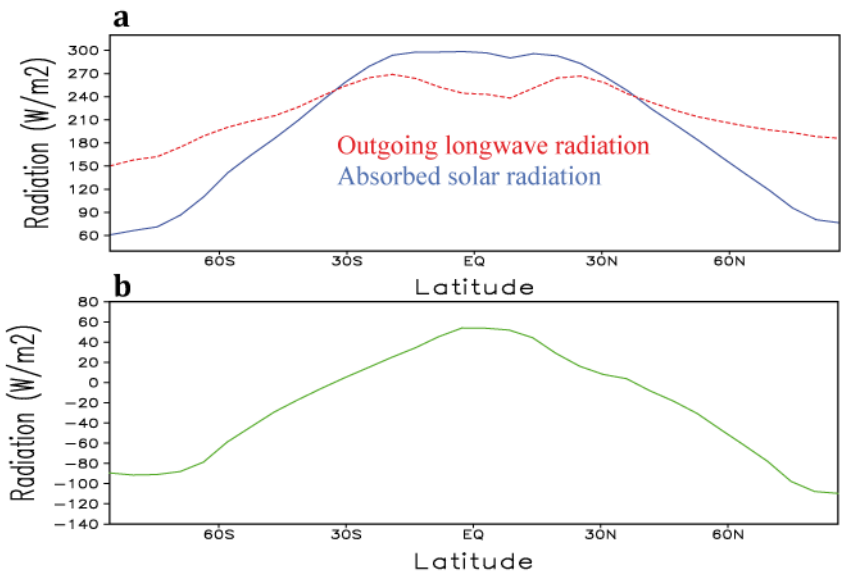


**Figure 2 (a) Zonal mean of the absorbed solar radiation (blue) and the outgoing longwave radiation (dashed red) at the top of the atmosphere in annual mean (in $Wm^{-2}$). (b) Zonal mean of the difference between the absorbed solar radiation and the outgoing longwave radiation at the top of the atmosphere in annual mean (in $Wm^{-2}$). From [31].**

### 2.1.2 Atmosphere

The atmosphere is the central component of the climate system. It is a relatively thin layer of a gaseous mixture, distributed over the Earth surface. A large percentage of the mass of the atmosphere is found in the first kilometres of heights. More precisely about 99% of the mass is "contained" below 30 km above the Earth surface. Considering the horizontal dimensions of our planet – about 20000 km from pole to pole – the atmosphere is a narrow layer. Despite to this a great amount of detail and a lot of interesting and often complex physical phenomena are present into this thin layer.

The classical division of the atmosphere in vertical layers (Figure 3) is possible thanks to the different properties about the chemical composition, the temperature gradient, the stability of the layer and energy considerations. Starting from the surface we find the troposphere, the stratosphere, the mesosphere and the thermosphere. They are conceptually separated by the so-called pauses. Below the mesopause (about 80 km) we find constant concentrations for the inert gases as nitrogen and oxygen (that together constitute the 99% of the total), also carbon dioxide is well mixed below this height. Variable components are present at different heights. The water vapour is present in most part in the troposphere (especially in the lower part) and the ozone in the middle of the stratosphere. The situation is complicated by the presence of rapidly changing substances as water in liquid and solid phases, dust particles, sulphate aerosols, volcanic ashes and so on. With a highly

variable concentration in time and space, aerosols play an important – and not well understood – role in the climate system. Due to the gravity force, the atmosphere is vertically stratified (denser layers are present near the surface) – it is the expression of the well-known hydrostatic equilibrium. Despite to their low percentage with respect to the main components, the minor constituents may play a very important role. Water vapour, carbon dioxide, and methane are strong absorbers of the LW radiation – i.e. the well-known greenhouse gases cited in the previous subsection. So, the atmosphere is important not only for the dynamic part but also for the radiative transfer of the radiation and then for the Earth energy budget.

The troposphere, in which the temperature drops off with the height up to about 10 km, is the place of the main meteorological phenomena. It shows a wide variety of phenomena at very different spatial and temporal scales. Starting from the largest Rossby waves (about 5-6000 km) up to local turbulent phenomena in the planetary boundary layers (on the scale of centimetres or less), the troposphere shows complex and often unstable dynamics. Convection is a phenomenon of fundamental importance in the troposphere, such as the vertical temperature profile at equilibrium is driven by both radiative transfer and convection.

At higher heights, the stratosphere shows a rising temperature gradient in the vertical direction. This is due to the absorption by the UV component of the solar radiation by molecular oxygen that brings to the formation of the well-known Ozone layer – triatomic oxygen – a minor but fundamental constituent of the planetary atmosphere. Ozone is a photochemically produced specie, it is present in all the atmospheric column, but it has a maximum around 25 km of altitude as said, in the stratosphere. The Ozone is the responsible for the already mentioned inverse temperature gradient of the stratosphere and has a role of fundamental importance as shield for the life-dangerous ultraviolet radiation. Also, the gases responsible for the Ozone layer disruption (the so-called Chlorofluorocarbons, CFCs) are minor components – of the order of part for billion or less. Despite to such very low concentration, CFCs are responsible for the well-known Ozone layer depletion. Even a minor constituent of the atmosphere may have a fundamental role for climate dynamics, it is the same for the above-mentioned greenhouse gases (in section 2.1.4 we will see deeper details). Furthermore, in the stratosphere also very interesting phenomena related to the observed circulation – i.e. the winter polar stratospheric vortex – may have an influence on tropospheric dynamics.

Even higher, due to very low air density, we have no more absorption of the solar radiation, so in the mesosphere temperature drops again with the height. Finally, in the upper atmosphere temperature rises again due to photo-dissociation of the air molecules.

As said, the engine of the atmospheric motions is the differential heating by the solar radiation on the Earth surface. Due to this, generally we have air convection at low latitudes and descending motion in the polar regions. In terms of fluid dynamics, to solve the "problem" of the Earth atmosphere, we are faced with a stratified fluid subject to convection and in rotation with the planet. Things are made more complex due to irregular boundary conditions – alternation of different surfaces as land and ocean, presence of orographic obstacles – that bring to different thermodynamic and mechanic properties. Despite of all this complexity – that sees its expression in the weather features – when we consider long time average, atmospheric dynamics has a globally coherent structure – Figure 4. Due to the combined effect of convection and Coriolis force, we have three big circulation cells in the atmosphere. To fix the ideas think about the northern hemisphere. Starting from the equator, we have convection of air masses up to the tropopause, here the motion is forced to proceeds toward the North Pole but – due to the Coriolis force – the route is broken with descending motion at tropical latitudes (around 30N) – this is the so called *Hadley cell*. It brings to the well-known sub-tropical high-pressure systems. Then we have the *Ferrel cell* up to middle latitudes, where we have again convection in correspondence with the so-called *polar front*. Finally, with the polar cell the path is completed, with descending motion at the North Pole. In proximity of the tropopause, at the latitude correspondent at the closure of the Hadley and Ferrel cells, we find a peculiar – relatively narrow – belt of strong winds, the subtropical and the polar *jet streams*. The jet streams are around 11 kilometers and can reach speeds of more than 300 km/h.

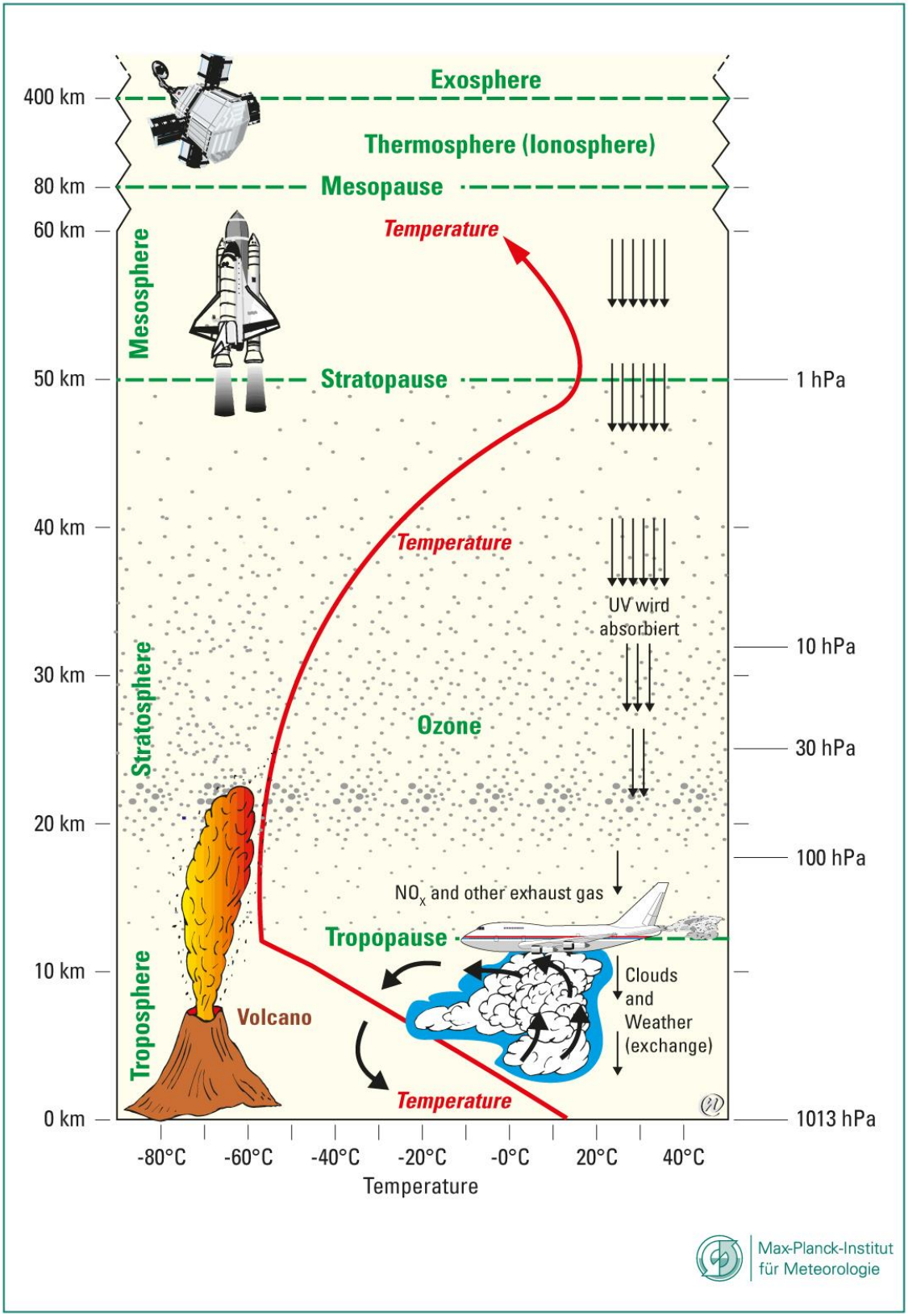


**Figure 3 A schematic representation of the vertical structure of the atmosphere. From Max Planck Institute for Meteorology website.**

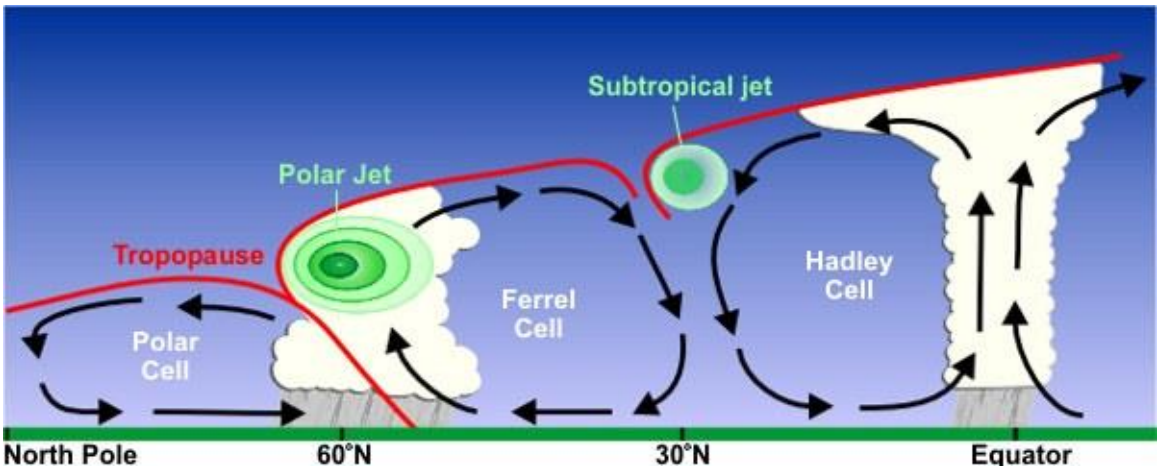


**Figure 4 A representation of the mean global atmospheric circulation, from the equator to the north pole (the same is for southern Hemisphere). Here we can find the Hadley, Ferrel and Polar cells, a highlight of the jet stream is also represented. More details in the text. From United States National Weather Service: https://www.weather.gov/jetstream/jet**

### 2.1.3 Hydrosphere

The hydrosphere consists in all the water in the liquid state on the Earth surface. Oceans, interior seas, lakes, rivers and subterranean waters are the "components" of the hydrosphere. Obviously, the oceans play the most important role. They cover almost two thirds of the Earth surface, so the solar radiation arrives mostly on them. Due to their huge mass and to their high heat capacity they must be considered as the climate system energy reservoir. Oceans are considered as thermal regulators for the global temperature. In terms of dynamics, as water is denser than the atmosphere, the most active part of the oceans is the surface mixed layer (the first 100 meters). Below this one, another layer with peculiar characteristics – i.e. the thermocline – is present. Further down we find the deep ocean: it has more constant properties respect to the layers above, in fact it reacts to the external perturbations in a very long time (on the order of decades up to centuries).

Oceans show a slower circulation than that of the atmosphere, a global circulation gyre is present, and a complete cycle is accomplished in about one century. The ocean circulation in average conveys heat from tropical areas to the poles due to the different heating by the solar radiation – see again Figure 2. We have not only horizontal circulation. Overturning of big ocean masses is present at several locations in the ocean and in the interior seas. These motions, with descending and rising currents are the engines for the ocean circulation itself and they are due to mechanisms related to temperature and salinity features of the oceanic masses. So, a strong influence in these phenomena is due to events as ice melting and rain precipitation, in fact, both

modify the surface salinity conditions. Circulation at smaller scale is present and it influences climate at a regional level. The response time to an external perturbation varies from weeks in the mixed layer up to centuries for the deep ocean.

Ocean and atmosphere are strongly coupled. Air-sea interactions occur at several temporal and spatial scales through exchange of matter, energy and momentum. Evaporation of water vapour is a relevant component of the hydrological cycle, together with cloud formation, precipitation and runoff towards the oceans. The global amount of water vapour depends strictly on the balance between evaporation and precipitation over the oceans.

Inner waters, as lakes, rivers and subterranean waters are also a relevant part of the hydrological cycle and furthermore can have an impact on climate at the local scale, e.g. rivers are important for the salinity of water near the coasts, or a lake can influence long term precipitation and temperature trend in the neighbouring areas.

We conclude this subsection by talking about another important role for the oceans. They may be source, sink or reservoir for other important atmospheric components as, for example, carbon dioxide. Since $CO_2$ is slightly soluble in water, about one-third of the total global emissions from fossil fuels burning is washed out of the atmosphere, with a consequent reduction of the greenhouse warming due to this gas. However, we do not know how long the oceans will continue to provide this service. We know that the solubility of $CO_2$ in water decreases with increasing temperatures. So, oceans could become saturated with a consequent slowdown or even an inversion of this uptake. If at last it will become a source of carbon dioxide, a strong positive feedback for global warming will occur.

### 2.1.4 Greenhouse effect

The above-mentioned greenhouse effect is a popular term given to the role of the atmosphere in the energy budget of the Earth, that brings to a rising of the mean global surface temperature. This name derives from the analogy with the garden greenhouse. It permits to the solar radiation to enter and warm the soil, but the LW radiation emitted by this latter is entrapped because the greenhouse is opaque to the infrared wavelengths. Obviously, the parallelism is a bit forced but this term is widely used in both the scientific and the popular lexicon.

The role of the panels of glass of the garden greenhouse is replaced by some gases in the Earth atmosphere. Almost all the atmospheric gases are transparent to the SW radiation. The main constituents (molecular Oxygen and Nitrogen) are fairly transparent also to the terrestrial radiation,

but some – minor – components are not. This opacity is due mainly to the absorption bands of water vapour, carbon dioxide, methane, ozone and other minor components that are really present in the IR frequencies. Although they represent a low percentage of the total of the atmosphere, these gases can be easily measured – except for the water vapour due to its very low lifetime – and all of them show an unequivocal increase starting from the industrial era.

Reconstructions from ice core analysis show values for carbon dioxide comprises between 180 and 280 parts per million in volume (ppmv) during at least the last 400 thousands of years [32], while for methane on the same period we are between 400 and 700 parts per billion in volume (ppbv) [32]. In the last 150 years we passed by 280 ppmv up to 410 ppmv for $CO_2$ and by 700 ppbv up to 1850 ppbv for methane. Also, for other greenhouse gases (GHGs) we see similar increases. Furthermore, this rising is unequivocally due to human activities, since the origin of the atmospheric carbon can be discerned by isotopes measurements [33]. These two gases are two of the more powerful GHGs, as they have strong infrared absorption bands located in correspondence to wavelengths typical of the terrestrial and low atmosphere emission spectra, so they block part of the radiation emitted towards the space.

Other pollutants, in particular sulphur-containing gases like $SO_2$ and $H_2S$, are responsible for aerosol production that is related to global cooling production and in part have limited the warming during last decades. So, the situation is complex, we have various conflicting contributes to the global mean temperature behaviour.

In term of physics, simple energy budget calculations permit to estimate that the “natural” contributions of greenhouse gases at the levels of pre-industrial era, is responsible for around 30-35K of increase with respect to an Earth without atmosphere. The greenhouse effect is a fundamental – and positive – ingredient for the birth of life on our planet as we know it. In fact, without atmosphere the Earth would be completely frozen. If a minority percentage of gases is responsible for an effect of around 30-35K, it should be evident that a strong modification of their atmospheric concentrations it is enough to have serious problems. Furthermore, the climate system does not behave in a linear way; on the contrary it is a complex system with non-linear dynamics and many feedbacks, so even the rising of concentration of the GHGs does not involve a proportional temperature increase. A strong modification of the boundary conditions could bring to unexpected new equilibrium states as mentioned in the introduction of section 2.1.

The theory related to the increase of the GHGs finds confirmation in the observed temperature, that shows a warming trend starting from the pre-industrial era, although with some exception (Figure 5) [34], [35], [36].

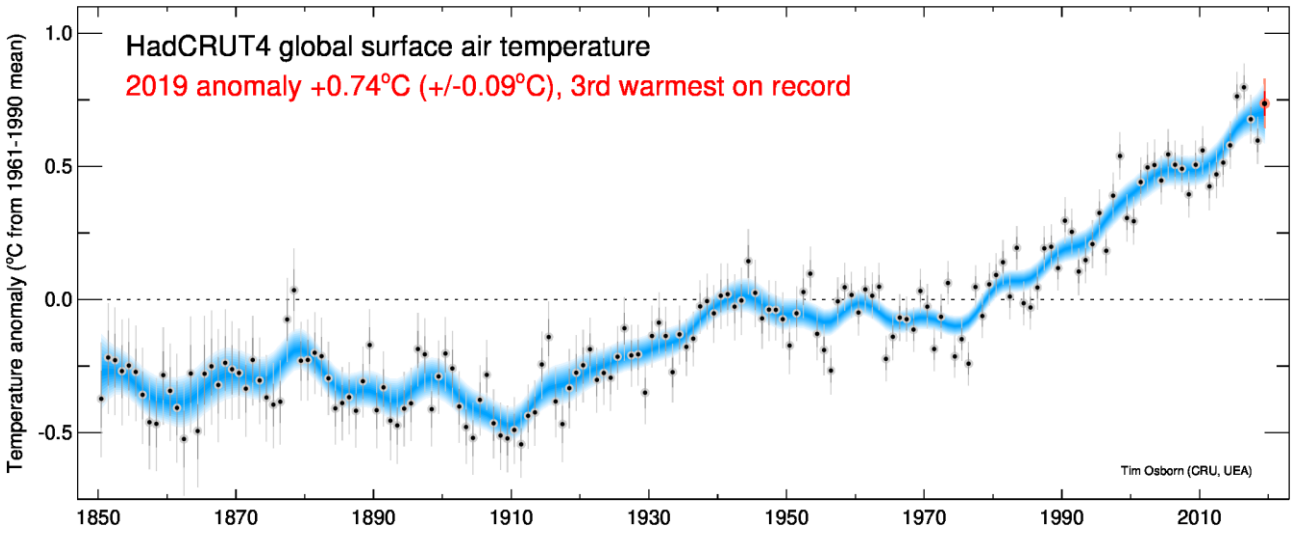


**Figure 5 Global mean surface temperatures from 1850 up to 2019. Yearly (black dots with uncertainty bars) and decadal smoothed (blue curve) temperature series are shown. Data from HadCRUT4 dataset. Image by Tim Osborn (CRU, UEA).**

If global warming it is due to the rising in GHGs concentration is a question traditionally faced with the global climate models (see section 2.2 for a description about their functioning). The scientific community agrees to attribute the recent warming to GHGs emissions by anthropogenic emissions. As said in the introduction, a discussion on the theme of the attribution is faced with an original method in chapter 4.

### 2.1.5 Climate Observations

As is typical for many sectors of physics, measurements are of fundamental importance for the development of meteorological and climate research. For both past climate analysis and for future climate forecasts we need the data. Historically the first quantitative measures related to atmospheric sciences have been that of temperatures. Up to the middle of the 20$^{th}$ century, data were furnished by in situ measurements or, at most, by atmospheric soundings by weather balloons.

In order to determine past climate trends – in particular temperature variations – a lot of indirect proxies exist as ring trees analysis, geological analysis and – probably the most important – ice core measurements. Especially the last, as said in the previous subsection 2.1.4 [32], are a fundamental instrument for the temperature determination over the last hundreds of thousands of years.

The measure of actual climate parameters is important also for forecasting. Since the system under observation is huge and very complex, the measures should span the globe, as in the vertical that in the horizontal dimensions, in the deep of the oceans and should also have a high temporal resolution. Of fundamental importance for these purposes is the birth of the satellite era in the 1960s: nowadays most of the climate research is based on satellite measurements. The techniques in use involve the field of *remote sensing*. The measure of the parameter of interest is performed remotely – in an indirect way – exploiting the properties of the electromagnetic radiation. This last can be reflected or scattered sunlight, thermal emission by the earth surface or by the atmosphere molecules or even backscattered by a laser source on the satellite itself. Measures of vertical profiles of temperature, humidity, chemical composition and so on are possible.

Once obtained the radiometric measure another complex field joins the game, that of the *retrieval theory*. In order to obtain the measure for the parameter of interest, we must apply some complex inversion techniques to the mathematical formulae that relate the radiation to the parameter itself.

In this chapter a basic description of the climate system was supplied. Obviously here we cannot be exhaustive on this topic, so we again remand to the bibliography for more details [29], [30]. In the next section we face with the topic of weather and climate modelling. This will give us a clearer view on the differences between meteorology and climate.

## 2.2 Weather and climate modelling

Despite to their strong relationships, climate and weather present also strong differences. The objective of the scientists is to understand the physics of the two systems in order to make models able to do weather/climate forecasts and to reconstruct the past. Although they are strictly related, some differences exist, so also the approaches for modelling are different. Here we want to give some qualitative information on how it is possible to perform such kind of forecasts.

Weather and climate models try to reproduce in a “virtual laboratory” the behaviour of the two systems by solving the equations that represent the physical mechanisms involved [37]. It is a real demanding computational problem, in fact, it requires the resolution – often approximated – of complex, non-linear equations for each point of a global latitude-longitude-vertical grid and for many time steps. Both weather and climate modelling are related to data. Data play a double role; they are used as initial condition in the starting phase of the models and they are used also to validate the output of the models themselves.

Weather forecasts try to simulate the future weather conditions up to a maximum of 10-15 days. In fact, as written at the beginning of this chapter [3], the system is chaotic with a strong dependence on the initial conditions. Furthermore, some of the equations are solvable only in an approximate way, so we have the so-called predictability barrier. Even if we could know with absolute precision the initial conditions (i.e. the most important meteorological parameters) of our system – and it is not our case –we cannot perform reliable weather forecasts up to 10-15 days. The equation of motion regards temperature, pressure, wind, air humidity and they are the so-called *primitive* equations:

$$\frac{\partial V}{\partial t} + V \cdot \nabla V + W \frac{\partial V}{\partial z} = f(y) V \times k - \frac{1}{\rho_r} \nabla \mathrm{p} + \mathrm{F}$$

$$\frac{\partial p}{\partial z} = -g\rho$$

$$div(V) + \frac{1}{\rho_r} \frac{\partial W \rho_r}{\partial z} = 0$$

$$p = \rho R T$$

$$\frac{\partial \theta}{\partial t} + V \cdot \nabla \theta + W \frac{\partial \theta}{\partial z} = 0$$

In these equations $V = (u, v)$ and $W$ are, respectively, the horizontal and the vertical wind, $f = 2\Omega \sin \phi$ is the Coriolis term, $\rho$ is the air density, $p$ is the atmospheric pressure, $F$ encloses the forcing terms and finally $\theta = T(p_0/p)^{0.28}$ is the potential temperature. It is a deterministic equations system, if we know its state at a fixed time, we can calculate the state for a future time. We note that for the weather forecasts the conditions of the oceans are fixed at the starting time, in fact the changes in such a short range (5-7 days) do not modify in a significant way the atmospheric circulation.

Climate modelling must consider more physical processes than weather forecasts. In fact, now the evolution of the ocean is important as the interactions with the biosphere, the variation in the chemical composition of the atmosphere matters, we have also to consider the radiative transfer processes and so on. So, for this topic we have a different model for each component of the climate system with a "coupler" that can coordinate all the components in order to represent properly all the

physical interactions between them. These climate models are also called Global Climate Models (GCMs).

A question is immediately raised by the comparison with the weather forecasts. If the latter are possible only up to 10-15 days, how we can perform forecasts up to decades? First of all, the request is different, while weather forecasts concern the knowledge of the detailed weather conditions for a very small spatial scale, with a climate forecast we are interested in a mean state of a variable as, for example, the global mean temperature. Furthermore, if weather forecasts are a initial conditions problem, climate forecasts are a boundary conditions one. For climate forecasts the initial conditions are not important as the system – as said – loses quickly memory of the initial state. Instead, a fundamental role is played by the boundary – or external – conditions as the solar radiative forcing, the greenhouse gases concentrations, the features of the Earth orbit and so on. In this terms, not only climate forecasts make sense, but they are also possible. The most important climate forecasts are that related to the global mean temperature. Several climate models show a good agreement with the past temperatures [1]. Furthermore, also some tests have been performed for future projections. In particular, the idea is to change the value of some of the most influential variables as the greenhouse gases concentration – e.g. keeping fixed the concentration of GHGs at the pre-industrial level – and perform the forecast [1]. The results confirm that the recent warming of our planet seems due to the increasing of the GHGs concentrations.

As we said in the introduction, and how we will see in Chapter 3, a note moved against the GCMs is that they have a common ancestor and so must be considered too similar to consider their results sufficiently reliable. In Chapter 3 we focus on this point and show that with a completely different and independent approach we are able to obtain the same results and also deeper details in the same and also in other open topics of climate sciences (as we will see in the following of this thesis).

To conclude this introductive chapter about the climate system, we want to underline the complexity of the problem that is generally faced with the dynamical models. Also related problems are generally faced with this type of model, although the goodness of this approach some questions are raised, so the idea at the basis of this thesis is to follow a completely independent way – i.e. with data-driven neural network models – to gain knowledge in this field of research.

# 3. Methods – Neural Network analysis

The objective of this thesis is to study open problems of the atmospheric sciences using artificial intelligence methods. The idea is to make use of Neural Network (NN) analysis. So, in this chapter NN modelling is explained, starting from a historical point of view up to the detailed description of the NN model used for this work.

Neural Network modelling is inspired by neurosciences. In fact, the human brain with its $10^{11}$ interconnecting neural cells (or neurons), represents the most efficient parallel calculator, even if a single neuron can be compared to a processor with a modest calculation speed. Despite to its relatively slow 'clock speed', yet human brain beats computers on many tasks. So, its extraordinary capability consists on the presence of a massive network connection between the neurons themselves. So, what computational capability could be reached with a massive network of connected neurons? This question has inspired a lot of scientists and it is at the basis of the lively field of neural network modelling. From speech and image recognition up to geophysical research applications, NN modelling has been a hot topic during last decades. In particular, physicists look at NNs for data analysis, modelling and prediction.

NNs can be separated into two large categories based on the learning typology. We talk about *supervised learning* when we try to reproduce a series of data $\boldsymbol{y}$ $(y_1, y_2, \dots, y_n)$ starting from a set of $j$ predictors $\boldsymbol{x}^j$ $(x_1^j, x_2^j, \dots, x_n^j)$ - $n$ is the number of observations. The output of the NN $\boldsymbol{y}'$ $(y_1'$, $y_2', \dots, y_n')$ is *supervised* in order to reproduce the response (or target) data $\boldsymbol{y}$, minimizing an *error function* (or *objective* or *cost function*). Usually, regression problems are approached with supervised learning. At the opposite, in *unsupervised learning*, the model is supplied only with the input data and then tries to investigate about the nature of these data. Cluster analysis or classification problems are typical unsupervised learning applications.

There are many types of NNs. The most common type is the *feed-forward* one, in which the signal proceeds forward from the input to the output (passing through intermediate layers) without any feedback. This typology is the most suitable for the purposes of this work, in particular the *multilayer perceptron* – consisting in one or more hidden layer(s) inserted between the input and the output (see next section) – is the typical approach for regression problems, see [7] for a review of

similar geophysical applications with artificial intelligence methods. Other types of NNs exist. Among others, we cite *recurrent neural* networks (RNN), in which internal loops are present in order to keep memory of previous computations. In this way previous information can persist and drive training of NN towards better results, so RNN may be a good instrument especially for forecasting.

This chapter is organized as follow: in section 3.1 we introduce the argument of NN from a historical point of view, this is necessary in order to introduce the argument with a criterion of increasing difficulty order; in section 3.2 some generalities about the central problem of the nonlinear optimization in NN analysis are given and we also talk about the problem of overfitting and on the methods for prevent it; then, in section 3.3 several optimization methods will be described; in section 3.4, we will talk on how we may organize the available data for obtaining better NN training; finally in section 3.5 a deep discussion on the general structure of the NN model in use for this thesis is supplied.

## 3.1 From McCulloch and Pitts to multi-layer perceptron: a historical summary

The model of McCulloch and Pitts in 1943 [38] can be considered the first NN model. Following the rationale of neurosciences, we know that a neuron is physically connected to its neighbours, each of which sends an electric impulse to it. If the total stimulus (given by the sum of the single impulses) overcome a certain threshold, the neuron becomes active and in turn sends a signal in output. In a similar way a neuron in the McCulloch and Pitts model consists of a binary threshold, i.e. it receives a weighted sum of inputs from its neighbours and if this sum is higher than a fixed threshold the neuron gives 1 as output, otherwise 0. In formulae, if $x_k$ is the signal from the $k-th$ neighbour, the neuron will give as output:

$$y = H\left(\sum_{k=1}^{N}(w_k x_k + b)\right), \qquad (1)$$

where $w_k$ are the weights and $b$ is the *bias* or *offset parameter*. $H$ is the Heaviside step function: $H(z) = 1$ if $z \geq 0$ and $H(z) = 0$ if $z < 0$. With a tuning on the bias $b$ we can change the level of the threshold. An algorithm for the calculation of the best weights and bias still did not exist so, from a practical point of view, the model was not very useful.

The next step of NN modelling was achieved thanks to the *perceptron* model of Rosenblatt [39], [40] and by a similar work by Widrow and Hoff [41]. This perceptron model consists in a

layer of input neurons connected to an output layer (Figure 6). The improvement respect to the McCulloch and Pitts model consists in the introduction of a *learning algorithm*, thanks to which the weights and the bias parameters can now be calculated.

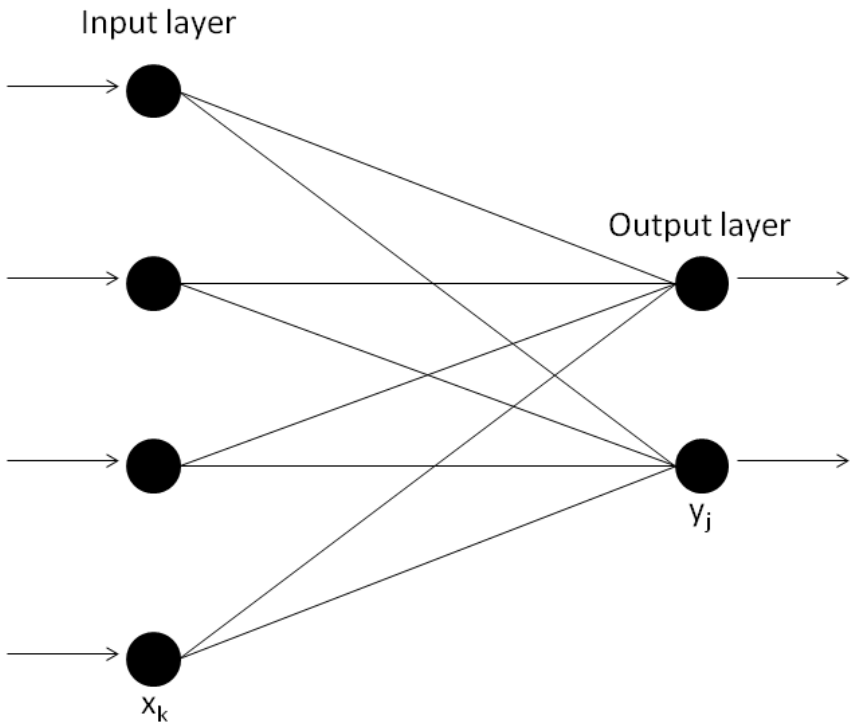


**Figure 6 An example of perceptron model with four inputs and two outputs.**

An output neuron is given by:

$$y_j = f\left(\sum_{k=1}^{N}(w_{jk}x_k + b_j)\right), \qquad (2)$$

where $x_k$ is an input, $f$ is the *transfer* (or *activation*) function to be specified, $w_{jk}$ are the weights connecting the input $k$ to the output $j$ and $b_j$ is the offset related to the output $j$. Once the inputs and outputs are fixed, the NN is trained to give the output $y_j$ as close as possible to the target $y_{tj}$. This is done by adjusting the weights in equation (2), so that the problem consists in the minimization of the error function with respect to the weights and bias themselves. More details about the minimization (also called *optimization*) will follow in the next.

It is important to highlight the possibility to use a different transfer function than the Heaviside step one. The latter is useful when the output is a binary variable, for example categorical (rain/no rain). So, in general different transfer functions can be used, as the *logistic function*, that is a so-called *S-shaped* function:

$$f(x) = \frac{1}{1 + e^{-wx}}. \qquad (3)$$

For $x \to -\infty$ this function has an asymptotic value of 0, while as $x \to +\infty$ it approaches the value of 1. The transition from 0 to 1 is gradual, as the logistic function is also nonlinear and differentiable it is more useful than the Heaviside step. Furthermore, now we can analyse the role of the weights and bias. Considering the univariate form $f(wx + b)$, we see that large values of $w$ give a steeper transition from 0 to 1, and as $w \to +\infty$ the activation function approaches the Heaviside step. Furthermore, small values of $w$ gives a quasi-linear transition. While increasing $b$ slides our function along negative values of the x-axis and vice versa. Analogous considerations can be done with different transfer functions. Another commonly used sigmoidal shaped function is the *hyperbolic tangent*:

$$f(x) = tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}. \qquad (4)$$

For $x \to +\infty$ this function has an asymptotic value of +1, while for $x \to -\infty$ it tends towards -1. So, the range of the hyperbolic tangent is between $\pm 1$. It can be demonstrated (LeCun et al., [42]) that the use of the logistic function brings to a slower convergence in comparison to the use of the hyperbolic tangent, so the latter is usually preferred.

After an initial excitement, some limitations of the perceptron model appeared. In particular, this approach was useful for linearly separable problems only. For instance, the perceptron model fails if it is applied to the Boolean logical operator XOR (the exclusive OR), a detailed explanation can be found in [43]. So, scientists realized that a further step was necessary. In particular the placement of additional layer(s) of neurons between the input and output layer permits to give more power to the neural network, in particular to fit any nonlinear function, the *multi-layer perceptron* (MLP) was born.

At this point of the history the problem was again the lack of an algorithm for the minimization of the error function for this new architecture. So, the interest about NN vanish, until 1986, when Rumelhart [44] rediscovered an old algorithm by Werbos [45] and apply it to the multi-

layer perceptron problem. In reference to Figure 7, in a MLP with one hidden layer the input signal $x_k$ is mapped into the hidden layer of neurons $h_j$ by:

$$h_j = f\left(\sum_k w_{jk} x_k + b_j\right), \qquad (5)$$

And then it is mapped on the output $y_m$:

$$y_m = g\left(\sum_j \tilde{w}_{mj} h_j + \tilde{b}_m\right), \qquad (6)$$

where $f$ and $g$ are the transfer functions for the hidden and output layer respectively; $w$ and $\tilde{w}$ are the respective weights and $b$ and $\tilde{b}$ the bias parameters.

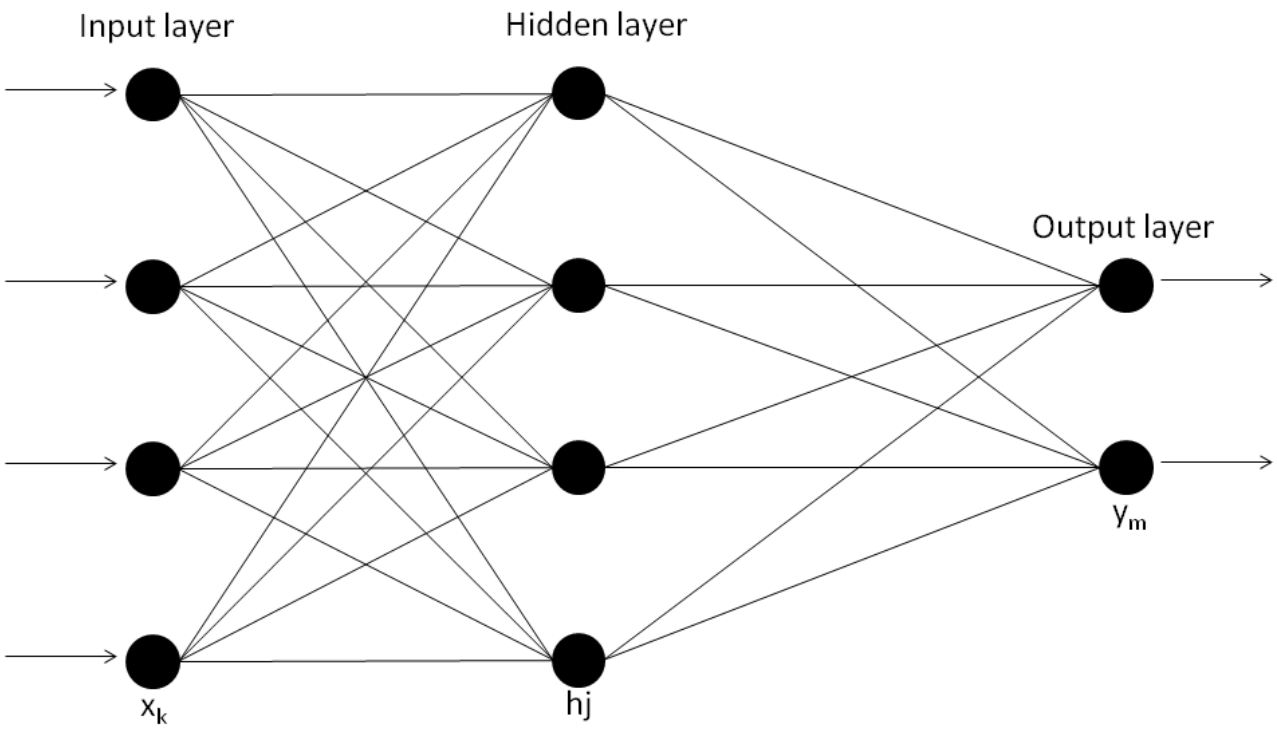


**Figure 7 A Multi-Layer Perceptron NN with one hidden layer**

As said, a supervised NN is trained in order to reproduce a fixed target. This is done by minimizing an error function, that can be defined in different ways. For example, a good choice is the *Mean Squared Error* (MSE) calculated on the output of the NN model and the target:

$$C = \frac{1}{N}\sum_{k=1}^{N}\left\{\frac{1}{2}\sum_{m}\left(y_m^k - y_{tm}^k\right)^2\right\}, \qquad (7)$$

here, $y$ is the output of the NN model while $y_t$ is the target data with $n = 1, 2, \dots, N$ different observations, $m$ is the summation on the number of outputs. The constant $1/2$ is traditionally adopted, it is helpful in the calculation of the derivative of $C$. In fact, the problem is to have an optimization algorithm in order to minimize $C$ respect to the weights and the bias, i.e. to find the best value for these parameters and so the best regression law. Afterwards we will talk about the details for this optimization procedure, now we can say that it is a very difficult problem. In fact, it consists in the research of the minimum for a multivariable function and such a research can be performed analytically only in a low number of cases. Furthermore, depending on the input variables scaling, the convergence appears to be inaccurate and/or too slow for operational purposes. So, a common procedure is to standardize data before to apply them to NN analysis. More scaling typologies exist, for example the "classic" normalization of data (in order to obtain data with zero mean and unitary variance), but in several cases also a rescaling of data is used (in order to obtain a series between $\pm 1$ or between 0 and 1). In the following we will show also another good reason for normalizing data.

The scaling of the output is also important, in particular the best choice is related to the problem into exam, i.e. if it is a regression or a classification. This is related to the choice of the transfer function for the output layer (the $g$ in formulae 6). For example, if we try to classify the output into two classes, it would be useful place $g(x) = \tanh(x)$. On the other side, if the output date is unbounded, it could be a problem limit data between $\pm 1$ because these limit values are only asymptotic for *tanh*. In this case is good practice to normalize data between $[-0.9 + 0.9]$ in order to improve the convergence velocity. Another possibility is to make use of a linear output transfer function, for example for the MLP with one hidden layer it means that the output will be a linear combination of sigmoidal shaped functions.

In literature, a univocal definition for counting the number of hidden layers does not exist. This number is considered sometimes as the effective number of hidden layers, while in other cases it is obtained as the sum of the effective number and the output layer. In this thesis the first

convention is adopted. An immediate representation of the structure of a NN can be given by sequentially listing the number of neurons of each layer, for example a 4-4-1 NN is a neural network with 4 input, 4 neurons in the hidden layer and 1 output.

The total number of weights and bias coefficients in a NN with structure $n_1 - n_2 - n_3$ can be calculated as:

$$N = (n_1 + 1)n_2 + (n_2 + 1)n_3. \qquad (8)$$

Where the number of weights is $n_w = n_1 n_2 + n_2 n_3$, and the number of bias is $n_b = n_2 + n_3$. In a multilinear regression with *n* predictors, the corresponding number of coefficients is $n + 1$. As we can see from the above formulae, the number of parameters in a MLP NN $n_1 - n_2 - 1$ is widely greater than the "correspondent" multilinear regression. Not exist a law that fixed the optimal number of parameters *N* to be used in relationship to the number of the observation in a dataset $N_D$. Generally, it is better to fix a NN structure with $N \ll N_D$, but in many cases (especially in geophysical applications) this in not possible, as we are faced with limited datasets. Furthermore, we can have strong correlations between predictors, so a Principal Component Analysis [6] is sometimes applied before NN analysis in order to work with a smaller (the leading few Principal Components) number of inputs. In any case, we underline that generally it is a good practice to fix $N \ll N_D$.

Regarding the case with more than one output, two approaches are possible. A NN with multiple output or more NN with one single output. The choice depends on the correlation between the outputs themselves, for example if they have zero correlation then the best choice is to work with more NN with a single output. Obviously, the multiple output architecture is the best in case of correlated outputs, in this case the NN can be able to consider such relationship.

Finally, we note an interesting fact. If in a NN with one hidden layer we place two linear transfer function, then the output will be simply a linear combination of the inputs. So, we see that the presence of a nonlinear transfer function, at least for the hidden layer, is fundamental to obtain a regression law with nonlinear characteristics.

## 3.2 Optimization – generalities

As said, the central problem in NN analysis consists in the research of the optimal value for the weights and the offset parameters. This is done by minimizing the error function *C*, so it is a mathematical analysis problem, i.e. the research of the minimum for a multivariable function (in our

case these variables are the weights and bias) – it is a *nonlinear optimization* problem. Generally, it is addressed with the well-known *back-propagation* algorithm. It consists of two different steps: the first one permits the estimation of the gradient of *C*, just thanks to the back-propagation of the NN model error; while the second step is the descent towards the minimum of *C*. For a multivariable function it is a complex problem. While the first part of the algorithm is a standard one, the second can be performed with several methods, in this section we describe the most important one, from both a historical and a practical point of view. Usually the term back propagation is referred to the entire algorithm (independently from the method used for the descent) or only for the first part of it (the computation of the gradient), so it could bring to some ambiguities. Here, with back propagation we refer to the first part of the algorithm.

First, to show what is the meaning of nonlinear optimization, we now describe both the linear (e.g. the research for the coefficients of a polynomial fit) and nonlinear optimization problems. Consider the relation:

$$y = w_0 + \sum_k w_k \, g_k, \qquad (9)$$

$g_k = g_k(x_1, \dots x_n)$ is a function (also nonlinear) of the predictor variables $x_1, \dots x_n$ and $w_1, \dots w_n$ are the parameters (the analogous of weights for NN) to be determined. As we can see, $y$ is a linear function of the parameters $w_n$. So, in this case the error function to be minimized is a quadratic function of the parameters:

$$C = \sum (y - y_t)^2, \qquad (10)$$

so, it has only one global minimum. For NNs we have seen that the weights $w_n$ are the variables of a more complex function as the hyperbolic tangent (in any case a nonlinear function) so, the error function is related in a nonlinear way to the weights themselves. In these terms, appears clear that $C$ will present a more complex shape, with a lot of secondary minimum, so that we have no guarantee that the optimization algorithm will be able to find the right minimum – i.e. the correct value for the weights. For NN analysis the research for the value of the "regression coefficients" is not a trivial issue.

Historically, for the descent the well-known *gradient or steepest descent* method is used. It is a slow method not more used for computational purposes, but for its historical importance we now describe it. In the next section we will describe other interesting (and more used) optimization

algorithms. For the sake of simplicity, in the following derivation it is convenient to define the error function $C^n$ in reference to a single observation of the training dataset:

$$C^n = \frac{1}{2}\sum_m (y_m^n - y_{tm}^n)^2, \qquad (11)$$

the "original" *C* (equation (7)) is the mean of $C^n$ over the *N* observations, $m$ is the summation on the number of outputs. Considering now the MLP with one hidden layer defined above by equations (5) and (6). If we place $w_{j0} = b_j$, $\widetilde{w}_{m0} = \tilde{b}_m$ and $x_0 = h_0 = 1$, we can redefine these equations in a more useful format:

$$h_j = f\left(\sum_k w_{jk} x_k\right), \qquad (12)$$

$$y_m = g\left(\sum_j \widetilde{w}_{mj} h_j\right), \qquad (13)$$

now the summation over *k* and *j* starts from zero (the bias term). Furthermore, we can represent all the weights and bias of the two layers of our NN with a vector $\boldsymbol{w}$. So, in back propagation first of all the weights and bias are randomly fixed (generally following a certain algorithm, as we will specify successively), and then are updated of the quantity $\Delta\boldsymbol{w}$ related to the gradient of the error function $C^n$ by:

$$\Delta\boldsymbol{w} = -\eta \frac{\partial C^n}{\partial \boldsymbol{w}}, \qquad (14)$$

$\eta$ is a scale factor called *learning rate*, it is the step, fixed by the algorithm, along the direction of the gradient of $C^n$. Explicitly for the hidden and the output layers, (14) can be rewrite as:

$$\Delta w_{jk} = -\eta \frac{\partial C^n}{\partial w_{jk}}, \qquad (15)$$

$$\Delta \widetilde{w}_{mj} = -\eta \frac{\partial C^n}{\partial \widetilde{w}_{mj}}. \qquad (16)$$

So, at the *n+1* step the weights and bias are updated as:

$$\boldsymbol{w}(n+1) = \boldsymbol{w}(n) + \Delta\boldsymbol{w}(n). \qquad (17)$$

Now we see how to explicitly calculate $\Delta \boldsymbol{w}$, for both the hidden and the output layers. First of all, we define:

$$u_j \equiv \sum_k w_{jk} x_k, \qquad \tilde{u}_m \equiv \sum_j \tilde{w}_{mj} h_j. \qquad (18)$$

In this way equations (12) and (13) become:

$$h_j = f(u_j), \qquad y_m = g(\tilde{u}_m). \qquad (19)$$

Now, in order to calculate $\Delta \boldsymbol{w}$, we can start with the explicit calculation of the derivatives of $C$:

$$\frac{\partial C^n}{\partial \tilde{w}_{mj}} = \frac{\partial C^n}{\partial \tilde{u}_m} \frac{\partial \tilde{u}_m}{\partial \tilde{w}_{mj}} = -\tilde{\delta}_m \frac{\partial \tilde{u}_m}{\partial \tilde{w}_{mj}}, \qquad (20)$$

here we have used the well-known chain rule for the derivative calculation. Furthermore, we have defined $\tilde{\delta}_m$, that is called *sensitivity* of the $m-th$ output neuron. Again, applying the chain rule we can calculate the sensitivity:

$$\tilde{\delta}_m = -\frac{\partial C^n}{\partial \tilde{u}_m} = -\frac{\partial C^n}{\partial y_m} \frac{\partial y_m}{\partial \tilde{u}_m} = (y_{tm} - y_m)\, g'(\tilde{u}_m), \qquad (21)$$

where we explicitly differentiate the error function $C^n$ (equation (11)). Considering the second of the equations (18):

$$\frac{\partial \tilde{u}_m}{\partial \tilde{w}_{mj}} = h_j. \qquad (22)$$

At this point we have all the ingredients to complete our recipe, using equations (20), (21), and (22) into (16) we obtain the expression for the weights and bias update relative to the output layer for our MLP network:

$$\Delta \tilde{w}_{mj} = \eta \tilde{\delta}_m h_j = \eta (y_{tm} - y_m) g'(\tilde{u}_m) h_j. \qquad (23)$$

If we use as transfer function for the output layer a linear one, we have $g' = 1$. Following the same rationale - i.e. applying the chain rule - we can calculate also the expression for the update of the weights connecting the input to the hidden layer of the MLP:

$$\Delta w_{jk} = \eta \delta_j x_k = \eta \left( \sum_m \tilde{\delta}_m \, \tilde{w}_{mj} \right) f'(s_j). \qquad (24)$$

The equations (23) and (24) permit to complete our algorithm. At the first iteration, the MLP model starts with random values for the weights, it gives as output $y_{tm}$, from which the error is calculated by $C^n$. At this point the error is back propagated in order to update the weights connecting the hidden to the output layer and then again back propagated to update the weights connecting the input to the hidden layer. The procedure is iterated until a certain convergence criterion is reached; the single iteration is called *epoch*. So, the process consists in mapping the inputs *forward* and then *back propagating* the error to update the weights and this is repeated until $C^n$ satisfies a fixed convergence criterion. As said, we have applied this procedure to a single pattern of data, i.e. data relative to a single observation. We can also consider more observations in order to better exploit all the information contained in the dataset (we see deeper this aspect in the following). The procedure remains the same, with the global error function $C$ calculated by averaging on all the single error functions $C^n$ (each calculate separately for every $n$). What was said above holds also for a more complex NN with more than one hidden layer.

The convergence criterion is related to the value of the error function $C$: it is a very subtle matter. A MLP with a single hidden layer and a sufficient number of hidden neurons is able to fit a training dataset with arbitrary accuracy – i.e. until reaching a $C = 0$ value. But we have to keep in mind that data contain both signal and noise, so we do not need for a NN able to reproduce both (a condition known as *overfitting*) but we search for a model able to reproduce quite well the signal and mostly to generalize to data not used in the training procedure. To prevent overfitting, the classical procedure consists in dividing the data into two parts, one for training and the other for *validation*. The error function for the weights calculation is estimated on the training set, so at each epoch its value will be decreasing. Correspondingly $C$ is calculated also on the validation set (not used for training), on which it will first have a decrease and starting from a certain epoch an increase (Figure 8). This indicate that, outdated the epoch corresponding to the minimum, we are in overfitting. So, that epoch is considered a good choice for stopping the training, this is the *early stopping* approach. A fixed law for the fraction of data to be dedicated for validation does not exist, empirically about 10% of data is used as validation set (more details about the question of the dimension of validation set will be given is section 3.4).

A further training related problem is due to the presence of several secondary minimum in the error function. The risk is to income in one of this relative minimum with a consequent poor

fitting. On the other hand, reaching the absolute minimum can give overfitting. One way to approach this problem is using an ensemble strategy. The training-validation procedure is repeated several times, each with different initial values for the weights of the NN, in order to better explore the multidimensional space of the error function. The choice of the initial weights can be done randomly or following an algorithm as that of Nguyen and Widrow [46]. In this way there is a high probability to obtain a good regression law with no noise fitting.

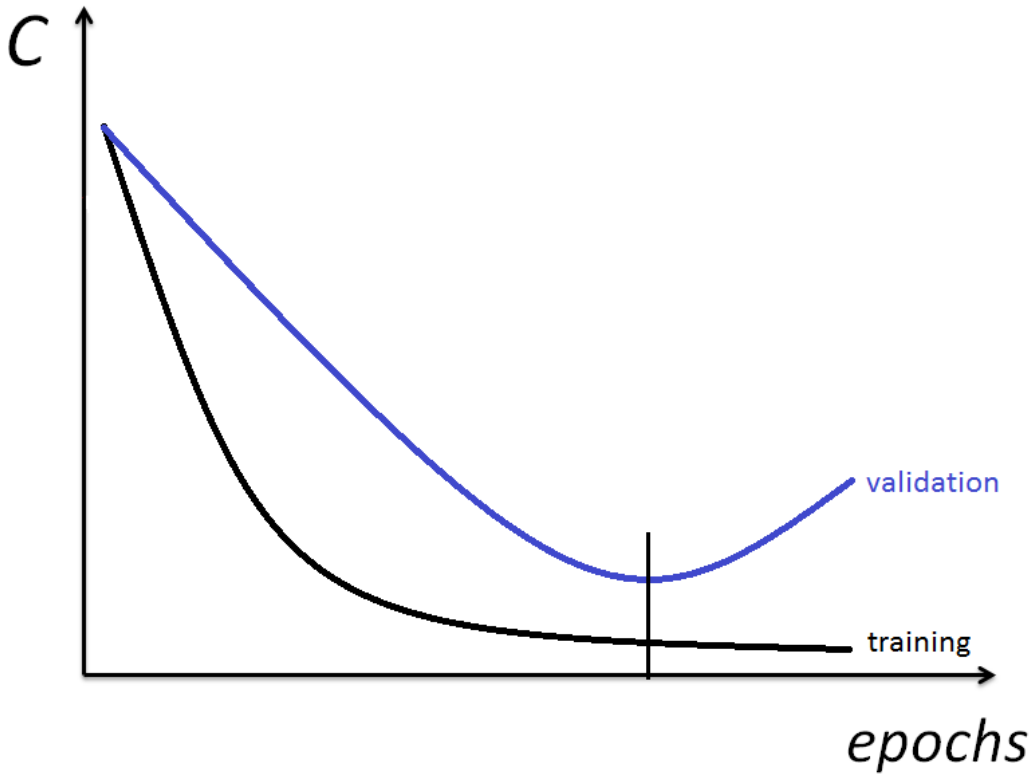


**Figure 8 Cost function calculated respect to the number of epochs for training (black curve) and validation (blue curve) sets. On the training set *C* is a monotonically descendent function, while on the validation set it has a minimum. In order to prevent overfitting, it is a good practice to stop the training at the epoch correspondent to this minimum (early stopping procedure).**

The number of hidden layers and neurons has also an influence on overfitting. In this chapter we have described a MLP with one hidden layer, but the same reasoning holds for more complex structures. So, the questions are: how many hidden layers are necessary? How many hidden neurons for each layer? Several studies [52], [53] and [54] have shown that a MLP with one single hidden layer is able to reproduce arbitrarily well any continuous function, provided that a sufficient number of hidden neurons is used. A simple single hidden layer MLP can be a universal approximator. However, there is no general theoretical indication on the exact number of neurons, only with empirical tests is possible to fix a congruous number. Intuitively a high number of

neurons will bring to better interpolation capabilities but also to a higher risk to fall into overfitting. So, the practical indication is to perform several tests with the same training-validation procedure earlier described, making use of a different number of neurons in order to fix the ideal number. In the next sub-section, we introduce also the most common approach used for overfitting prevention, the so-called regularization.

Overfitting is the most delicate issue of NN analysis. As explained, several cautionary methods exist; now we introduce the most common approach, named regularization of the error function. This method acts in a direct way on the error function previously defined (7). The idea is to add penalty parameters to the error function. In terms of formulae, equation (7) is modified as:

$$C = \frac{1}{N}\sum_{k=1}^{N}\left\{\frac{1}{2}\sum_{m}\left(y_m^k - y_{tm}^k\right)^2\right\} + P\frac{1}{2}\sum_{j} w_j^2, \qquad (7b)$$

here $P$ is a fixed positive constant called the penalty parameter (or regularization parameter or hyperparameter), $w_j$ are all the weights/bias of NN. We see that with a positive and constant $P$ the selection of larger $|w_j|$ is not encouraged because this would increase the value of $C$. In other words, if we fix a large $P$ we penalize larger weights. The effect of the penalty can be better understood in terms of the transfer function. For example, considering the hyperbolic tangent, if $|wx| \ll 1$ we can approximate the function with its first term of Taylor expansion:

$$y = \tanh(wx) \approx wx$$

So, if we adopt small weights, the nonlinear activation function become a linear one. Using large $P$ for penalize $w$ brings to decrease the nonlinear capabilities of the NN model and so to prevent overfitting.

The scaling of the predictor variables also has a strong importance. In fact, if we have two predictors, the first of which is much larger in magnitude that the second one, this latter must have higher weights than the first in order to have a comparable influence in the NN model. But the penalty parameter $P$ is the same for all the weights (7b) and then it will act more strongly on the second input, not allowing it to take large weights. A similar analysis holds for multiple target cases. So, the general procedure is to standardize data before to use them for NN training. Classical standardization (zero mean and unitary variance) may be used, but also other methods exist – e.g. the normalization of data between $\pm 1$ or between $+1/0$. Using this shrewdness, we may be sure that the regularization will works properly.

A way to fix the value of the penalty parameter is by an approach like the early stopping previously described. The dataset is divided into training and validation, several trials are performed with different values of $P$. Again, the value of $P$ correspondent to the minimum on the validation set will be the optimal one.

In this section we have presented a quite exhaustive introduction about NN analysis, starting from the structure of the network up to the methods for prevent overfitting. Now, in the next one, we will go deeper into the details of the methods for the optimization of the error function, the core of NN analysis.

## 3.3 Optimization methods

As seen in the last section, the minimization of the error function $C$ with respect to the weights/bias $\boldsymbol{w}$ is a problem of nonlinear optimization. Usually it is common to solve this problem with an iterative approach (useful for computational purposes). Now, a general description of this approach is shown, then, in the next subsections, a detailed description of the most used methods is supplied. Here, we show optimization methods called *deterministic*, in that each step of the iterative procedure of optimization is well fixed by deterministic formulas. Also, stochastic optimization methods exist (see [43] and references therein for more details).

Considering an iterative approach, we name $\boldsymbol{w}_0$ the actual value of our weights/bias. If we expand $C$ as a Taylor series around $\boldsymbol{w}_0$ we have:

$$C(\boldsymbol{w}) = C(\boldsymbol{w}_0) + (\boldsymbol{w} - \boldsymbol{w}_0)^T \nabla C(\boldsymbol{w}_0) + \frac{1}{2}(\boldsymbol{w} - \boldsymbol{w}_0)^T \boldsymbol{H}(\boldsymbol{w} - \boldsymbol{w}_0) + \ \ldots, \qquad (25)$$

here $\nabla C$ is the gradient of the function $C$, $\boldsymbol{H}$ is the Hessian matrix composed by all the second derivatives of $C$. If we apply the gradient operator to (25):

$$\nabla C(\boldsymbol{w}) = \nabla C(\boldsymbol{w}_0) + \boldsymbol{H}(\boldsymbol{w} - \boldsymbol{w}_0) + \ \ldots \qquad (26)$$

For the optimal values of $\boldsymbol{w}$, $\nabla C(\boldsymbol{w}) = 0$ and ignoring the higher order terms we have:

$$\boldsymbol{w} = \boldsymbol{w}_0 - \boldsymbol{H}^{-1} \nabla C(\boldsymbol{w}_0). \qquad (27)$$

In these terms, we obtain the formulae for the iterative computation of $\boldsymbol{w}$:

$$\boldsymbol{w}_{n+1} = \boldsymbol{w}_n - \boldsymbol{H}_n^{-1} \nabla C(\boldsymbol{w}_n). \qquad (28)$$

This is the well-known Newton method (we will go in deeper details in section 3.3.3). When we are faced with a multidimensional problem – i.e. when we have many weights and bias parameters – the explicit calculation of the inverse of the Hessian matrix is a complicated issue. So, in order to limit the computational effort, several approximation methods have been developed. Now, once fixed the general problem, we will describe several optimization methods adopted in NN analysis.

### 3.3.1 Gradient descent method

This method has been already described into section 3.2. Also named steepest descent, the parameters of the NN are updated by the relation:

$$\boldsymbol{w}_{n+1} = \boldsymbol{w}_n - \eta \nabla C(\boldsymbol{w}_n), \qquad (29)$$

where $\eta$ is the learning rate. By comparison with (28) we see that $\eta$ is replacing the inverse of the Hessian matrix, so the steepest descent represents a simplification of the more general Newton method. In practice, one tries to find the minimum – i.e. the optimal value for the weights and bias $\boldsymbol{w}$ - by descending along the negative gradient of $C$. This is the direction of the steepest descent (so, this is the origin of the name of the method). At a first glance this could appear as a very efficient way for searching the minimum but it is not so.

The value of $\eta$ can be a fixed constant or calculated by a *line minimization algorithm*. The first approach simply consists in proceeding along the negative gradient direction with a fixed step at each iteration. In the second one, we proceed along the direction of the negative gradient and stop when the minimum along that direction is achieved. In detail, at step $n$ we have the estimation $\boldsymbol{w}_n$ for the parameters. From this point we move along the negative gradient direction $\boldsymbol{t}_n$ identified by:

$$\boldsymbol{t}_n = -\nabla C(\boldsymbol{w}_n). \qquad (30)$$

So, we move along the direction $\boldsymbol{t}_n$:

$$\boldsymbol{w}_n + \eta \boldsymbol{t}_n. \qquad (30b)$$

We move along the direction $\boldsymbol{t}_n$ until the minimum - along this direction - is reached, this is individuated by the condition:

$$\frac{\partial}{\partial \eta} C(\boldsymbol{w}_n + \eta \boldsymbol{t}_n) = 0, \qquad (31)$$

that permits to fix the value of $\eta$, explicitely:

$$\boldsymbol{t}_n^T \nabla C(\boldsymbol{w}_n + \eta \boldsymbol{t}_n) = \boldsymbol{t}_n^T \nabla C(\boldsymbol{w}_{n+1}) = 0, \qquad (32)$$

from which it is clear that $\boldsymbol{t}_n \perp \nabla C(\boldsymbol{w}_{n+1})$. Since that $\boldsymbol{t}_{n+1} = -\nabla C(\boldsymbol{w}_{n+1})$, we finally have:

$$\boldsymbol{t}_n^T \boldsymbol{t}_{n+1} = 0. \qquad (33)$$

So, the directions individuated at the steps *n* and *n+1* are orthogonal, this results in an inefficient zig-zag path for this approach (Figure 9). Using a fixed step $\eta$ bring to a similar bad situation, too small size for $\eta$ could result in too many steps (Figure 10a), while too large size brings to an even more strong zig-zag pattern (Figure 10b)

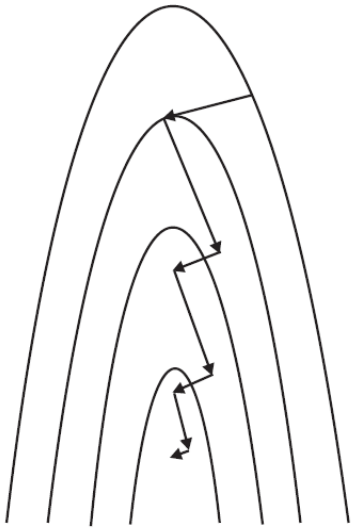

**Figure 9 The gradient descent method with line minimization. The research of the minimum is in this case very slow due to the zig-zag path. Adapted from Hsieh [43]**

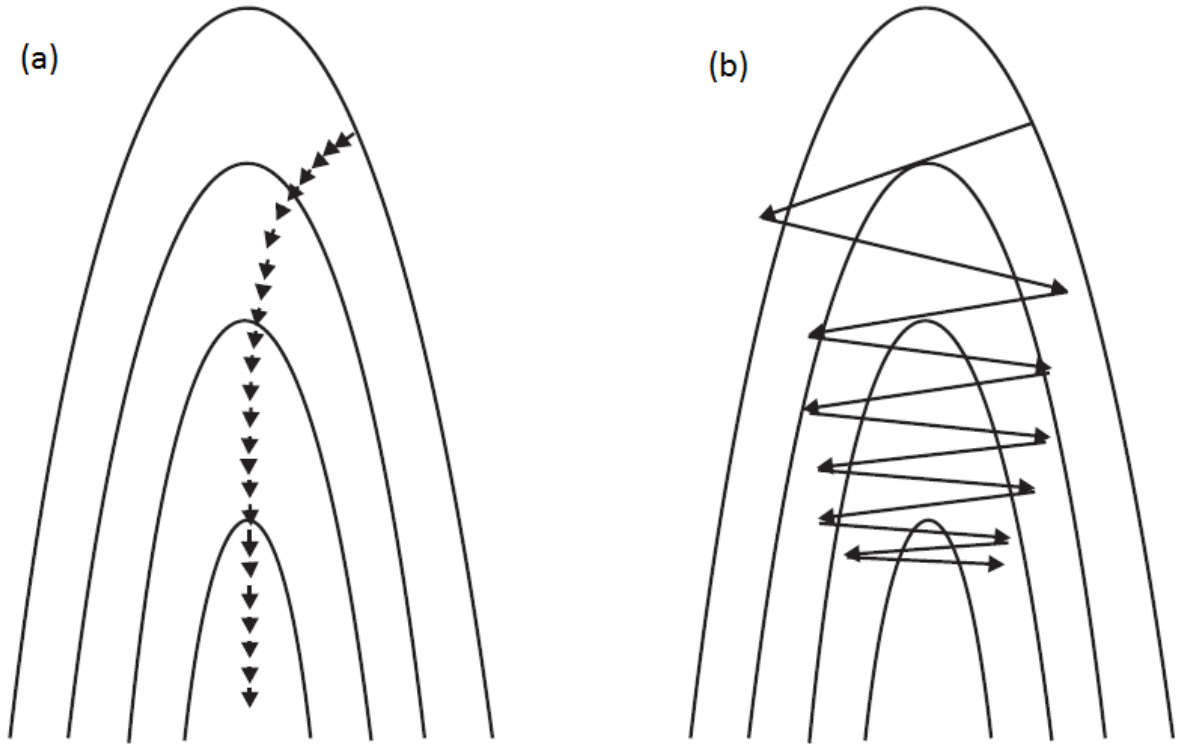


**Figure 10 Example of the gradient descent with a too small fixed step size (a), and with a too big fixed step size (b). Adapted from Hsieh [43].**

Ultimately, the way for optimize the method in order to make the steepest descent an efficient algorithm is to add the so-called *momentum* $\mu$ to the direction of descent:

$$\boldsymbol{t}_n = -\nabla C(\boldsymbol{w}_n) + \mu \boldsymbol{t}_{n-1}. \qquad (34)$$

The rationale is that $\mu$ prevents the new direction to be orthogonal to the previous one, decreasing the total length of the path. So, the steepest descent method is generally used with the momentum coefficient (Figure 11). Using the equation (34) into (29) gives the steepest descent method with momentum.

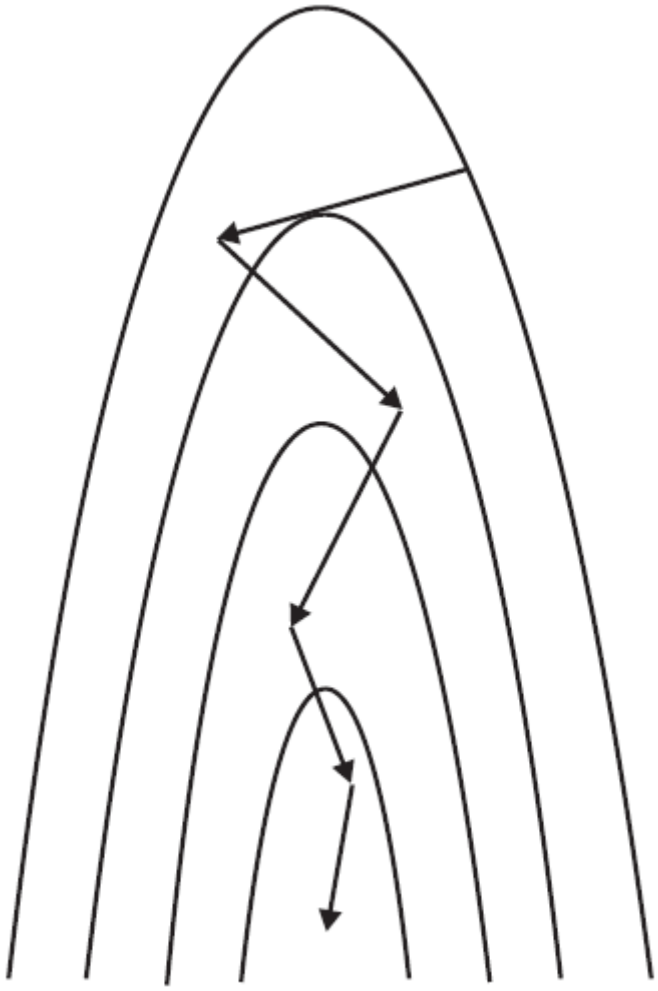

**Figure 11 Example of the gradient descent method with momentum. Adapted from Hsieh [43].**

### 3.3.2 Conjugate gradient method

This method has been developed thanks to the works of Hestenes and Stiefel in 1952 [47] and by Fletcher and Reeves in 1964 [48]. It consists in a gradient descent method with momentum (equation (34)), where this last is automatically fixed as we can see in the following. Starting from the weights determination $\boldsymbol{w}_{n+1}$ we want to fix the next direction $\boldsymbol{t}_{n+1}$ for the descent in such a way that the gradient of $C$ in the direction $\boldsymbol{t}_n$ remains 0 (at first order), as we go along the direction $\boldsymbol{t}_{n+1}$, this is expressed by the condition:

$$\mathbf{t}_{\mathrm{n}}^{\mathrm{T}} \nabla C(\boldsymbol{w}_{n+1} + \eta \boldsymbol{t}_{n+1}) = 0, \qquad (35)$$

where $\eta$, as said, is the learning rate that tells us how much we are moving along the direction $\boldsymbol{t}_{n+1}$. Using the equation (26) opportunely adapted, we can rewrite:

$$\nabla C(\boldsymbol{w}_{n+1} + \eta \boldsymbol{t}_{n+1}) \approx \nabla C(\boldsymbol{w}_{n+1}) + \boldsymbol{H}\eta \boldsymbol{t}_{n+1}. \qquad (36)$$

Then (35) becomes:

$$\mathbf{t}_{\mathrm{n}}^{\mathrm{T}} \nabla C(\boldsymbol{w}_{n+1}) + \eta \mathbf{t}_{\mathrm{n}}^{\mathrm{T}} \boldsymbol{H} \boldsymbol{t}_{n+1} = 0. \qquad (37)$$

As $\boldsymbol{t}_n \perp \nabla C(\boldsymbol{w}_{n+1})$, we have the condition for the conjugate direction:

$$\mathbf{t}_{\mathrm{n}}^{\mathrm{T}} \boldsymbol{H} \boldsymbol{t}_{n+1} = 0, \qquad (38)$$

where, indeed, $\boldsymbol{t}_{n+1}$ is said to be conjugate to $\boldsymbol{t}_n$. The next step is to estimate the momentum parameter $\mu$, now we show three different methods.

To obtain the Hestenes-Stiefel method, we define:

$$\boldsymbol{b}_n \equiv \nabla C(\boldsymbol{w}_n). \qquad (39)$$

Making use of the equation (34) rewritten for $\boldsymbol{t}_{n+1}$ – i.e. $\boldsymbol{t}_{n+1} = -\nabla C(\boldsymbol{w}_{n+1}) + \mu \boldsymbol{t}_n$ into the condition (38):

$$\mathbf{t}_{\mathrm{n}}^{\mathrm{T}} \boldsymbol{H}(-\boldsymbol{b}_{n+1} + \mu \boldsymbol{t}_n) = 0. \qquad (40)$$

From which we have:

$$\mu \mathbf{t}_n^{\mathrm{T}} \boldsymbol{H} \boldsymbol{t}_n = \mathbf{t}_n^{\mathrm{T}} \boldsymbol{H} \boldsymbol{b}_{n+1} = \boldsymbol{b}_{n+1}^T \boldsymbol{H} \boldsymbol{t}_n, \qquad (41)$$

as $\boldsymbol{H} = \boldsymbol{H}^T$. Equation (41) is the relation that permits to obtain the value of the momentum $\mu$. The last step consists in the approximate calculation of the Hessian matrix, cause its exact determination could be computationally expensive. Using the equation (26) ignoring higher order terms, the definition (39) and the equation (30b), we have:

$$\boldsymbol{b}_{n+1} - \boldsymbol{b}_n = \boldsymbol{H}(\boldsymbol{w}_{n+1} - \boldsymbol{w}_n) = \eta \boldsymbol{H} \boldsymbol{t}_n. \qquad (42)$$

Using this relation into (41), we finally obtain the estimation for the momentum, named Hestenes-Stiefel method:

$$\mu = \frac{\boldsymbol{b}_{n+1}^T (\boldsymbol{b}_{n+1} - \boldsymbol{b}_n)}{\mathbf{t}_n^{\mathrm{T}} (\boldsymbol{b}_{n+1} - \boldsymbol{b}_n)}. \qquad (43)$$

Now two other methods are shown, without derivation (further detail in [43]). In the Polak-Ribiere method [49] and [50], the momentum is given by:

$$\mu = \frac{\boldsymbol{b}_{n+1}^{T}(\boldsymbol{b}_{n+1} - \boldsymbol{b}_n)}{\mathbf{b}_n^{\mathrm{T}}\boldsymbol{b}_n}. \qquad (44)$$

Finally, the Fletcher and Reeves method [48]:

$$\mu = \frac{\boldsymbol{b}_{n+1}^{T}\boldsymbol{b}_{n+1}}{\mathbf{b}_n^{\mathrm{T}}\boldsymbol{b}_n}. \qquad (45)$$

In all these approximate derivations the higher order terms in the expansion of the gradient of $C$ have been neglected, so the several methods differ from each other. Generally, with (44) better performances are achieved.

To conclude, the conjugate gradient method explanation, we must find the optimal value for the learning rate $\eta$ along the search direction $\boldsymbol{t}_n$. This is done by minimizing the error function $C(\boldsymbol{w}_n + \eta\boldsymbol{t}_n)$ in function of $\eta$ itself, along such path. For simplicity we now write only $C(\eta)$ in order to underline the dependence on $\eta$. Generally an iterative algorithm is adopted (Figure 12), as step 1) three points $a$, $b$, $c$ are fixed along the search direction, in such a way that $J(a) < J(b)$ and $J(b) < J(c)$. As $C$ is a continuous function a minimum exists in the interval $(a, c)$; 2) a parabolic curve is fitted on these three points, the correspondent to the minimum $d$ of the parabola is fixed on $C$; 3) three points among $a$, $b$, $c$ and $d$ with the minimum value of $C$ are fixed and the procedure is iterated until convergence is reached.

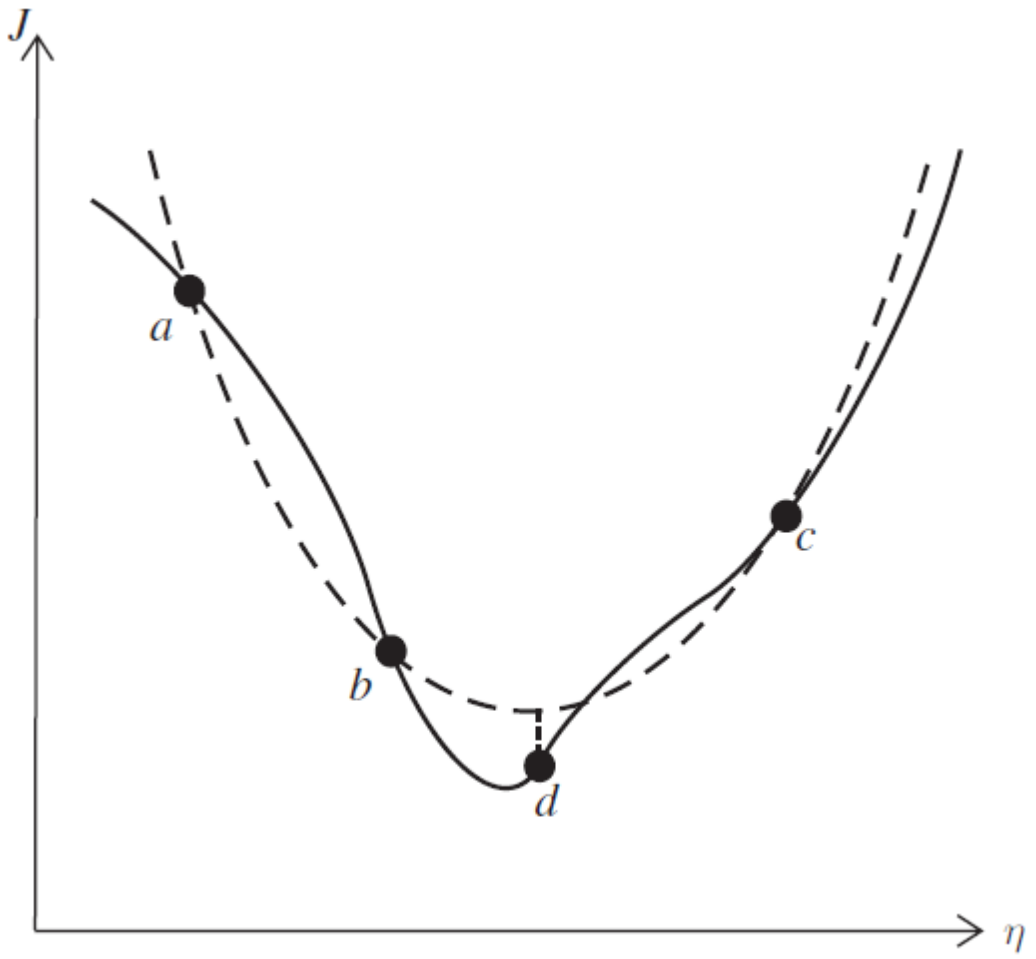


**Figure 12 Example of the line search algorithm for the optimal value of the learning rate. Detailed description of the procedure is in the text. Figure from Hsieh [43].**

### 3.3.3 Quasi-Newtonian methods

At the beginning of this section, with the equation (28), we have introduced the Newton method. Thanks to the definition (39) and by defining the inverse of the Hessian matrix as $\boldsymbol{B}_n$ we can rewrite (28) in a more compact form:

$$\boldsymbol{w}_{n+1} = \boldsymbol{w}_n - \boldsymbol{B}_n \boldsymbol{b}_n. \qquad (46)$$

The derivation of this formulae assumes that the terms above the quadratic in the Taylor series of the error function $C$ can be ignored. This is certainly true when we are in the nearby of a minimum, but in general it is not true. So, in order to obtain generalization capabilities, the so-called learning rate $\eta_n$ is added:

$$\boldsymbol{w}_{n+1} = \boldsymbol{w}_n - \eta_n \boldsymbol{B}_n \boldsymbol{b}_n. \qquad (46)$$

Also in this more accurate form, the Newton method is not so useful because at each iteration it requires the estimation of the inverse of the Hessian matrix. So quasi-Newtonian methods have been developed using a simpler estimation for $\boldsymbol{B}_n$. For example, also the above-mentioned gradient descent and conjugate gradient methods are related to quasi-Newtonian methods, in fact if we replace $\boldsymbol{B}_n$ with the identity matrix $\boldsymbol{I}$ we obtain the gradient descent. While the quasi-Newtonian methods preserve also the conjugate condition (38) of the conjugate gradient method.

Here, two quasi-Newtonian methods are presented, for more details about their derivation we refer to the respective references. The first method is the Davidon-Fletcher-Powell (DFP) [51], [52]. It is an iterative procedure, for $k = 0$ it starts by placing a random value for the initial weights $\boldsymbol{w}_0$ and a random symmetric positive definite matrix as $\boldsymbol{B}_0$. Then the iterative procedure starts:

1. We assume that $\boldsymbol{t}_n = -\boldsymbol{B}_n \boldsymbol{b}_n$.
2. The error function $C(\boldsymbol{w}_n + \eta_n \boldsymbol{t}_n)$ is minimized respect to not negative $\eta_n$. Then we can compute updated weights $\boldsymbol{w}_{n+1}$, $\boldsymbol{p}_n \equiv \eta_n \boldsymbol{t}_n$ and $\boldsymbol{b}_{n+1}$.
3. Setting $\boldsymbol{b}_{n+1} - \boldsymbol{b}_n = \boldsymbol{q}_n$, we also have the updating for the approximate Hessian matrix:

$$\boldsymbol{B}_{n+1} = \boldsymbol{B}_n + \frac{\boldsymbol{p}_n \boldsymbol{p}_n^T}{\boldsymbol{p}_n^T \boldsymbol{q}_n} - \frac{\boldsymbol{B}_n \boldsymbol{q}_n \boldsymbol{q}_n^T \boldsymbol{B}_n}{\boldsymbol{q}_n^T \boldsymbol{B}_n \boldsymbol{q}_n}. \qquad (47)$$

4. We update $k$ and return to point 1. (if convergence is not reached).

Again, we note that if we choose as $\boldsymbol{B}_0$ the identity matrix, the DFP method coincide with the conjugate gradient.

The most widespread quasi-Newton method is the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method [53], [54], [55] and [56]. The update for the estimator of the inverse of the Hessian matrix is:

$$\boldsymbol{B}_{n+1}^{BFGS} = \boldsymbol{B}_{n+1}^{DFP} + \boldsymbol{v}_n \boldsymbol{v}_n^T, \qquad (48)$$

$\boldsymbol{B}_{n+1}^{DFP}$ is given by (47) and $\boldsymbol{v}_n$ is defined as:

$$\boldsymbol{v}_n = (\boldsymbol{q}_n^T \boldsymbol{B}_n \boldsymbol{q}_n)^{1/2} \left( \frac{\boldsymbol{p}_n}{\boldsymbol{p}_n^T \boldsymbol{q}_n} - \frac{\boldsymbol{G}_n \boldsymbol{q}_n}{\boldsymbol{q}_n^T \boldsymbol{G}_n \boldsymbol{q}_n} \right). \qquad (49)$$

In BFGS method the conjugate direction property (38) is again preserved.

We have seen that both the conjugate and the quasi-Newtonian methods do not make use of the explicit inverse Hessian matrix. Furthermore, the second one tries to approximate the inverse Hessian matrix as well as preserving the conjugate direction properties. So quasi-Newton methods are a further step respect to the conjugate gradient. The use of an approximate Hessian matrix brings to faster convergence, furthermore the line search for the estimation of the learning rate $\eta$ is not more necessary. However, the major drawback is the large storage necessary for the matrix $\boldsymbol{B}_n$, which has dimensions $N_w \times N_w$ where $N_w$ is the total number of weights and bias. So, the quasi-Newtonian methods require a memory of order $O({N_w}^2)$ instead of a capability of $O(N_w)$ for the conjugate. In order to reduce memory requirements, a method was proposed by Shanno [57].

In any case, for problems related to small dataset (as is the case for the arguments faced in this thesis) the quasi-Newtonian methods in general, and the BFGS in particular, find a wide application.

### 3.3.4 Nonlinear least squares methods

All the optimization methods illustrated up to now are of general applicability – i.e. they are not related to the explicit form of the error function. In many cases the error function consists in a sum of squares, as for example the Mean Squared Error. In general terms:

$$C = \frac{1}{2}\sum_{k=1}^{N}(\varepsilon^k)^2 = \frac{1}{2}\|\boldsymbol{\varepsilon}\|^2, \qquad (50)$$

here, $\varepsilon^k$ is the error associated to the $kth$ observation, $\boldsymbol{\varepsilon}$ is the vector of all these errors, $N$ is the number of observations. Now, two optimization methods related to this type of error functions will be showed.

At the $nth$ step we have the weights $\boldsymbol{w}_n$ and we "move" towards $\boldsymbol{w}_{n+1}$. We can write the Taylor expansion for $\boldsymbol{\varepsilon}$ in function of the weights:

$$\boldsymbol{\varepsilon}(\boldsymbol{w}_{n+1}) = \boldsymbol{\varepsilon}(\boldsymbol{w}_n) + \boldsymbol{J}_n(\boldsymbol{w}_{n+1} - \boldsymbol{w}_n), \qquad (51)$$

here, $\boldsymbol{J}_n$ is the Jacobian matrix calculated at the $nth$ step. Explicitly it has elements:

$$J_{ki} = \frac{\partial \varepsilon^k}{\partial w_i}, \qquad (52)$$

so, $\boldsymbol{J}$ has dimension $N \times L$ - i.e. (number of observations) × (number of weights/bias). Using (51) into equation (50) the error function at the $nth$ step can be approximated as:

$$\begin{aligned}\frac{1}{2}\|\boldsymbol{\varepsilon}(\boldsymbol{w}_{n+1})\|^2 &= \frac{1}{2}\|\boldsymbol{\varepsilon}(\boldsymbol{w}_n) + \boldsymbol{J}_n(\boldsymbol{w}_{n+1} - \boldsymbol{w}_n)\|^2 \\ &= \frac{1}{2}\|\boldsymbol{\varepsilon}(\boldsymbol{w}_n)\|^2 + \boldsymbol{\varepsilon}^T(\boldsymbol{w}_n)\boldsymbol{J}_{\boldsymbol{n}}(\boldsymbol{w}_{n+1} - \boldsymbol{w}_n) \\ &+ \frac{1}{2}(\boldsymbol{w}_{n+1} - \boldsymbol{w}_n)^T\boldsymbol{J}_n^T\boldsymbol{J}_n(\boldsymbol{w}_{n+1} - \boldsymbol{w}_n). \qquad (53)\end{aligned}$$

In order to find the minimum of $C$ in function of $\boldsymbol{w}_{n+1}$ and so, the weights themselves, we differentiate the right hand side of (53) respect to $\boldsymbol{w}_{n+1}$ and set the result equal to zero:

$$\boldsymbol{J}_n^T\boldsymbol{\varepsilon}(\boldsymbol{w}_n) + \boldsymbol{J}_n^T\boldsymbol{J}_n(\boldsymbol{w}_{n+1} - \boldsymbol{w}_n) = 0, \qquad (54)$$

from which we obtain:

$$\boldsymbol{w}_{n+1} = \boldsymbol{w}_n - (\boldsymbol{J}_n^T\boldsymbol{J}_n)^{-1}\boldsymbol{J}_n^T\boldsymbol{\varepsilon}(\boldsymbol{w}_n), \qquad (55)$$

this is the well-known Gauss-Newton method. We show how this last equation is related to the Newton method (28). In fact, for the error functions given by the sum of squares as (50), we can write the gradient:

$$\nabla C = \boldsymbol{J}^T\boldsymbol{\varepsilon}(\boldsymbol{w}). \qquad (56)$$

The Hessian matrix is composed by the elements:

$$\boldsymbol{H}_{ij} = \frac{\partial^2 C}{\partial w_i \partial w_j} = \sum_n \left(\frac{\partial \varepsilon^n}{\partial w_i}\frac{\partial \varepsilon^n}{\partial w_j} + \varepsilon^n \frac{\partial^2 \varepsilon^n}{\partial w_i \partial w_j}\right). \qquad (57)$$

If $C$ is a linear function of the weights the second term in the summation of (57) is equal to zero. If $C$ is not a linear function of $\boldsymbol{w}$ we can however neglect this term so, we obtain an approximation for the Hessian matrix in terms of the Jacobian:

$$\boldsymbol{H} = \boldsymbol{J}^T\boldsymbol{J}. \qquad (58)$$

It is clear that (55) approximates the Newton method (28) with ∇C given by (56) and the Hessian matrix given by (58).

To resume, by the equation (55) the weights are updated up to convergence to the minimum. The biggest problem of the Gauss-Newton method is that the step size could become too large and

so the assumption done in the equation (51) can fall. So, an evolution of this method is proposed, the so-called Levenberg-Marquardt method [58], [59]. The idea is to add a penalty term to the approximated version of the error function:

$$\frac{1}{2}\|\boldsymbol{\varepsilon}(\boldsymbol{w}_{n+1})\|^2 = \frac{1}{2}\|\boldsymbol{\varepsilon}(\boldsymbol{w}_n) + \boldsymbol{J}_n(\boldsymbol{w}_{n+1} - \boldsymbol{w}_n)\|^2 + \lambda\|\boldsymbol{w}_{k+1} - \boldsymbol{w}_k\|^2. \qquad (59)$$

The last term penalizes larger steps, in fact with larger $\lambda$ we obtain a smaller size for the step. If we minimize (59) respect to $\boldsymbol{w}_{n+1}$, as done for the derivation of (55), we have a new expression for the weights update formulae:

$$\boldsymbol{w}_{n+1} = \boldsymbol{w}_n - (\boldsymbol{J}_n^T\boldsymbol{J}_n + \lambda\boldsymbol{I})^{-1}\boldsymbol{J}_n^T\boldsymbol{\varepsilon}(\boldsymbol{w}_n), \qquad (60)$$

$\boldsymbol{I}$ is the identity matrix. For small $\lambda$ values we come back to the Gauss-Newton formulae, while for large $\lambda$ we have the gradient descent method. So, the Levenberg-Marquardt method takes the best from both these methods. It preserves the convergence velocity of the Gauss-Newton but also the robustness of the gradient descent when we are far from a minimum.

To conclude this sub-section, we note that the Jacobian matrix has dimension $N \times L$, where $L$, the total number of weights and bias, can be higher than $N$, the number of observations of our dataset. So, we need for even higher storage capabilities than that necessary for the Hessian matrix, - i.e. for quasi-Newton methods.

## 3.4 On the use of available data

In this section we want to focus on the aspect of how to reserve data for validation purposes. In the last part of section 3.2 we introduced the early-stopping method. In order to limit the problem of overfitting the available data for the training of the NN model are divided into two parts, one (bigger) for the training and the other for the so-called validation. As the number of iterations increases, the error function calculated on the training data will be a monotonically decreasing function, while if we calculate the error function on the validation data, we obtain a function that exhibits a minimum in correspondence to a certain epoch. This minimum gives the ideal number of iterations at which stop the NN training in order to avoid overfitting. In fact, when the error function starts to rise on the validation set means that our NN model was losing its generalization capabilities on data not involved in the training. So, in next subsection we see how to calculate the best fraction of data to dedicate for validation; then in 3.4.2 we illustrate more refined methods in order to use at best the available data for training purposes.

### 3.4.1 Choice of the fraction of validation data

Considering a dataset composed of $N$ observations, which fraction of data we must use as validation data? Intuitively, if we reserve too much data for validation, we will have a poorer model, while if we reserve too few data, we will fall into overfitting. So, an ideal fraction should exist. Amari et al. [60] introduced a theoretical derivation for this ideal fraction. For what has been said, a fraction $fN$ of data is used for validation, where $f$ represents this ideal fraction. Consequently, $(1-f)N$ represents the fraction of the dataset to be used for training. The ideal fraction can be defined in two ways depending on the number of the weights and bias of the NN. Named $N_w$ this number, if $N < 30N_w$, the ideal value for $f$ is:

$$f_{id} = \frac{\sqrt{2N_w - 1} - 1}{2(N_w - 1)}, \qquad (61)$$

and if $N_w \gg 1$ we have:

$$f_{id} \approx \frac{1}{\sqrt{2N_w}}, \qquad (62)$$

For example, if $N_w = 200$, from (62) we have $f_{id} = 0.05$, only a fraction of 5% of data has to been dedicated for validation.

The second case, for $N > 30N_w$, it may be shown [60] that a small difference exists between using $f_{id}$ or not using validation data at all – i.e. the entire dataset is dedicated to training. So, when we have a number of data much higher than the number of parameters, we can also avoid the early stopping procedure, as overfitting is no more a problem. But in the case of relatively small dataset, as is the case for many physical applications, we must adopt the early stopping procedure. Furthermore, there is the need to use at best the available data, so in the next subsection we will explain the *cross-validation* technique. A limit case of this technique is the so-called *leave-one-out* cross validation, used for this thesis.

### 3.4.2 Cross-validation

For many physical applications data are not plentiful, in particular for meteorology and climatology, reserving part of a dataset for validation can leave us with a small amount of data for the training of the NN model, furthermore, keeping fixed the validation set can bring to a bad representation of the variability of data. These problems are not present when there are plentiful data. In order to

optimize the use of the available data for the NN model construction, the best approach is by the cross-validation technique and related variants.

The idea consists in divide the data record in $K$ segments of (approximately) equal length - the so-called K-fold cross-validation. One segment is used for validation while the remaining $K-1$ for training. Iterating on all the segments permits to all the data to be used for validation improving the generalization capabilities of the NN. At this point one can chose the best model based on the minimum error on the validation segment. Obviously, the use of the ensemble technique – i.e. perform multiple runs for each segment with different initial weights - is yet recommended.

To fix the ideas, suppose a data record of 100 elements, and divide it in 10 segments. So, years 1-10, 11-20, 21-30, and so on, are used once at a time as validation set. The first iteration makes use of data 1-10 for validation, while data 11-100 are used as training set. The second iteration takes data 11-20 for validation and 1-10 and 21-100 as training. This is repeated sliding on all the segments up to use data 91-100 for validation and 1-90 for training at the tenth iteration. One can also increase the number of segments up to arrive at 100 segments, in this case the 100-fold cross-validation will coincide with the so-called *leave-one-out* cross-validation, where a single observation will be the validation set iteratively.

Care must be taken for time series data. In this case we could have autocorrelation inside the series. For example, considering a daily time series, suppose an autocorrelation time of 15 days. So if we apply the leave-one-out technique our NN will not have good generalization capabilities as in the training phase we will make use of data strictly related to our validation element so, the use of cross-validation can bring to an underestimation of the validation error. General prescription for these cases is to take a validation set equal or longer than the autocorrelation time.

A successive step is done when we consider the error to be associated to the NN model. In fact, the error calculated on the validation set is not a good estimation of the model error, because these data are however involved in the model determination. The model error must be estimated on independent data, not used nor for validation nor for training. Thus, a further division of the dataset is necessary in training, validation and *test* set, on the last we can determine the model error. We can consider this procedure as a “double” cross-validation. Considering the dataset of 100 observations, we could have a 10-fold *cross-test*. Data 1-10 are used for testing and 11-100 for training and validation. So, we must divide this last in training and validation, for example we use 11-20 for validation and 21-100 for training. This procedure is iterated with a double loop, the first in order to use all data as test set and the second (inner) in order to use at each “test iteration” all the

remaining data as validation set. Again, an ensemble strategy is recommended. In such a way, the forecast model error can be calculated over the entire dataset and it will be the real, reliable, error of our NN model. At limit we arrive at the *leave-one-out cross-testing*, in which each single data of our dataset is used as test set while the remaining are used for training and validation. This last will be the technique used for our model, as we describe in the following section.

## 3.5 Model description

In this work several open topics of the atmospheric sciences have been studied, obviously each with its proper NN model, but despite some small differences, the various models have a common structure. In this section the general structure of the NN model is furnished, in the chapters dedicated to the analysis of the various topics any variation respect to this "basic" description will be specified. Thanks to the introduction on NN modelling supplied in the previous sections of this chapter the following explanation should result quite simple. This NN model is built in such a way that it can perform realistic – i.e. not affected by overfitting – multiple nonlinear regression laws. Furthermore, the results can be considered reliable – i.e. not affected by peculiarities of the model or of the data themselves. Both these aspects will be clearer when the various applications will be described in the next chapters. This model was gradually developed in the course of the years, starting from the middle of '90 [61], passing by the first years of the new century [23] and [13], up to its description with a dedicated paper [17]. Respect to [17], here the model has been further developed and refined in relation to the different problems faced.

The model adopted is a multi-layer perceptron feedforward NN with backpropagation training, it is the most "popular" architecture for NN regression applications in geophysics. We use a single hidden layer cause, as said, also in this relatively simple case we can approximate any continuous function. Furthermore, keeping the complexity of the NN limited helps us to avoid overfitting. As we consider regression problems each model has a single output - i.e. one single output neuron. In summary, the NN model in use is schematically shown in the previous Figure 7, each problem will have a different number of input and/or hidden neurons depending on the problem itself. The transfer function for the hidden layer is a hyperbolic tangent, any different choice for each NN model will be specified in the following. The output transfer function is a linear one.

As far as the optimization algorithm is concerned, empirical tests allow us to choose from time to time the best one for each problem. For example, for the attribution problem of Chapter 3 the Levenberg-Marquardt algorithm is used (see section 3.3.4) while for the Atlantic Multidecadal

Oscillation problem (Chapter 4) the Broyden-Fletcher-Goldfarb-Shanno (BFGS, see section 3.3.3) is used. The optimization algorithm in use is specified from time to time in the following chapters.

Regardless of these important "technical aspects", we want to focus on the use of the available data for training, validation and test. As stated, here we handle limited length datasets, so the NN algorithm is developed considering this aspect. The probability to income into overfitting is reduced by taking a low number of hidden neurons and by adopting a leave-one-out training-validation-test procedure – already described into section 3.4. To better fix the ideas we consider a 100 years length series of data (for example mean annual temperature). The procedure can be described by the following steps:

1. Each year, once at a time, is used as test set starting from the first one.
2. Of the remaining 99 data, about 10% is randomly selected and used as validation set – i.e. more or less 10 years of data. The initial value of the weights and bias of the NN are fixed (pseudo-randomly) with the Nguyen-Widrow algorithm [46], this is done in order to build an ensemble of models.
3. The adopted backpropagation algorithm is applied on the training set, at each iteration the performance of the NN constructed using this training set is evaluated on the validation set. The procedure is iterated until the value of the error on the validation set starts to rise (the already described early stopping approach). Furthermore, in order to have even better performances with limited overfitting, we insert two further control conditions:
   3.1. we consider the so-obtained NN as a good one if the mean error calculated on the training set is lesser than a fixed threshold (dependent on the problem) and
   3.2. we consider the so-obtained NN as a good one if the number of epochs of training is higher than a fixed threshold – at least 100 epochs.

   The first condition is related to the need of obtain a good performance, the second one to avoid falling into a local minimum. In fact, we could have convergence with a low number of epochs of training but in this case the probability to get trapped into a secondary minimum is very high. If even one of this two conditions are not satisfied, we return at the point 2. and the training procedure is repeated. This can be repeated up to 100 times, in the case none of the two conditions is even satisfied (but it is an unlikely situation) the NN between the 100 with the minimum error on the training set is chosen as the best one.

4. Once the weights and bias of the NN are fixed thanks to the training-validation procedure, the NN error on the test set is calculated. It is an error calculated on data not involved in the NN

determination – the test data is completely unknown to the NN! – it is used for the mean squared error calculation. Now we return to the point 2 and the procedure is repeated 20 times in order to construct the first ensemble of values for the first year of data – i.e. 20 different NN models for each test data.

5. Finally, we return at the point 1. and the next year is selected as test set.

This complex procedure permits to construct an ensemble of NN models that give us an estimation of each single data of the target. The algorithm is constructed in such a way that the NN models ignore the value of the test set, giving us a strong guarantee of model generalization capabilities. So, the reconstruction of the target data can be considered as a real forecast of the target itself. Once the target estimation is obtained, all the statistics can be calculated.

In parallel to the NN training also a multilinear regression law is calculated. The linear law is obtained using data involved in training and validation of the NN. To fix the ideas if we have 100 data that law is calculated on 99 data, leaving one test data for the calculation of the associated error. In this way a parallel comparison between NN and linear regression is possible using the same data. Furthermore, an advantage is given to the linear regression, in fact it makes use of more data respect to the NN cause now we have not need for validation data. In the next Chapters we can see that despite this, NN performs always better than linear regression strengthening the goodness of our choice. Furthermore, we see also that the comparison between non-linear and linear methods permits to deduce very interesting information on the dynamic involved in the topics addressed.

Now we are at the end of this chapter. Here an explanation of NN modelling has been supplied, for regression problems and relatively small datasets. The final focus of this last section has been on the model used for this research activity. It will permit us to introduce next Chapters in a simple and immediate way.

# 4. Attribution of the last 150 years temperature variations

As defined by the IPCC [1], attribution is *the process of evaluating the relative contributions of multiple causal factors to a change or event*. Here, we are interested in the attribution of the last 150 years temperature variations. Up to now the studies on recent global warming are mostly faced with GCMs.

As said in Chapter 2, the climate system is strongly dependent on the boundary conditions: if we modify one of them a variation of the state of the system is expected. Using GCMs to simulate the complexity of the climate system permits also to perform experiments. In order to investigate the last 150 years, the idea is to set some of the boundary conditions – also called forcing – to their pre-industrial values and use these "artificial" data in GCMs simulations. The results of these experiments – also performed with different models – clearly indicate that anthropogenic forcing (in particular GHGs) are the responsible for the recent global warming [1]. The GCMs used by the scientific community have some differences, but also pertain to the same dynamical approach and come from a common ancestor, so the robustness of the result itself is subject of debate [5], [62]. Robustness is achieved when we have a common result from independent means of investigation and GCMs do not seem so independent from each other. So, the idea is to use a completely different approach – i.e. by neural network analysis. In recent years several works making use of data-driven models as neural network and Granger causality have been proposed [9], [10], [11], [12] but a comparison that follow the same rationale of the GCMs is lacking.

In this framework here we present the results obtained by the application of a neural network model of the type described in Section 3.5, these results have already been accepted by the scientific community [16]. Analysing the last 160 years, first we perform the reconstruction of the temperature trend making use of some predictors – i.e. the fundamental drivers of the climate system – after, different attribution experiments are performed. Our findings show that recent warming may be attributed basically to anthropogenic forcing (see next sections for more details), while the Sun seems to influence the period 1910-1975. The use of indices of natural variability permits to reproduce also the interannual variability. The application of our NN model permits to corroborate the previous knowledge given by GCMs and permits to obtain further details about the role of the external forcing and on the internal variability of climate system.

## 4.1 Data and method

We focus on annual data – since the middle of the 19th century – of mean global temperature, of radiative forcing (RF) of several drivers for the climate system and of data about indices of natural variability. In order to perform sensitivity tests, two alternative scenarios for data about anthropogenic radiative forcing are used. A list of all the data used is shown in Table 1: in the next sections we will show how these data have been used, here we provide a list and the source of them.

| | **List of variables** |
|---|---|
| **1** | Global annual mean temperature – T |
| **2** | Radiative forcing of greenhouse gases - RFGHG |
| **3** | Radiative forcing of black carbon – RFBC |
| **4** | Radiative forcing of anthropogenic sulphates – RFSOX |
| **5** | Radiative forcing of solar activity – RFSOLAR |
| **6** | Radiative forcing of volcanoes – RFVOL |
| **7** | Total RF of anthropogenic activities – RFANTH = RFGHG+RFBC+RFSOX |
| **8** | Total natural RF – RFNAT = RFSOLAR+RFVOL |
| **9** | Southern Oscillation Index – SOI |
| **10** | Pacific Decadal Oscillation – PDO |
| **11** | Atlantic Multidecadal Oscillation – AMO |

**Table 1 List of the potential predictors for our NN analysis**

For global mean temperature we consider land and marine temperature anomalies since 1850 of the HadCRUT4 database [63], from the Met Office Hadley Centre, freely available on the web.

As far as radiative forcing data are concerned, we use the freely available dataset collected at http://www.sterndavivi.com/datasite.html [12]. In particular, the data about GHGs are taken from the NASA/GISS website and the calculation of RF are performed using classical formulae developed in scientific literature [64], [65]. (See Figure 13)

Data relative to the emissions of sulphates are available only till 2011 [66], [67]. The calculation of related radiative forcing is based on slight modification of previous studies [12], [68], [69]. Data about radiative forcing of black carbon come from the RCP8.5 scenario [70] (See Figure 13).

Solar irradiance is obtained by an index available at https://data.giss.nasa.gov/modelforce/solar.irradiance/ [71]. The related RF is calculate following standard formula [65]. Furthermore, a synthetic series (RFSOLSTAT) was built for attribution test. It is built with a first-order Fourier series based on the first observed 65 years (See Figure 14). The observed series shows a transition toward a high energy regime at the beginning of the 20$^{th}$ century, so our idea is to consider a low energy regime for attribution purposes (see next section).

Volcanic radiative forcing is calculated by the optical thickness data [72], available from https://data.giss.nasa.gov/modelforce/strataer/, RFVOL is 27 times the optical thickness [12] (See Figure 15).

The three indices representing the natural variability are the Atlantic Multidecadal Oscillation (AMO) available since 1856 at www.esrl.noaa.gov/psd/data/timeseries/AMO; the Pacific Decadal Oscillation (PDO) available since 1854 at https://www.ncdc.noaa.gov/teleconnections/pdo/ and the Southern Oscillation Index (SOI) available since 1866 at www.cru.uea.ac.uk/cru/data/soi/soi.dat (See Figure 16-17-18).

The data relative to the alternative scenarios for the sensitivity tests (see section 3.3) are that of CMIP5 [73] available at https://data.giss.nasa.gov/modelforce/Fi_Miller_et_al14_upd.txt and that of Hansen et al. [74] available at https://data.giss.nasa.gov/modelforce/Fe_H11_1880-2011.txt. As the first available year for the latter is 1880, it was extended backward to 1850 filling the first years by zero values for the anthropogenic RF.

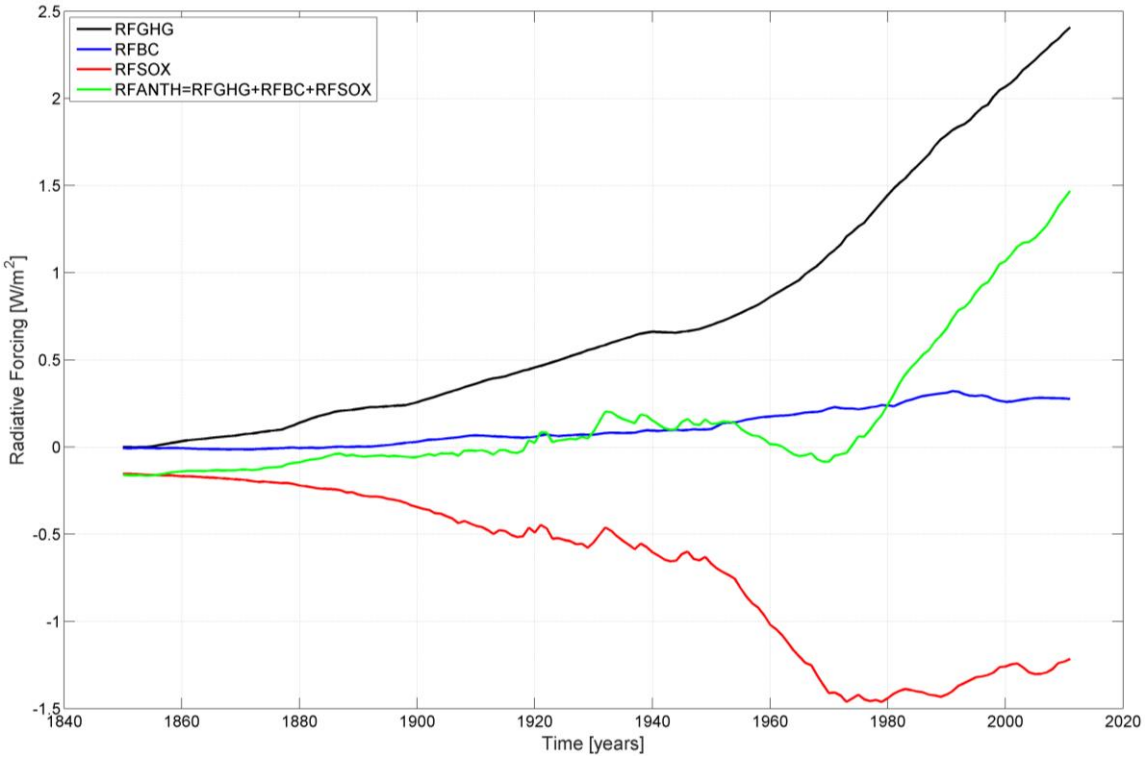


**Figure 13 External forcings used as input for the NN models: anthropogenic forcing anomalies [W/m²].**

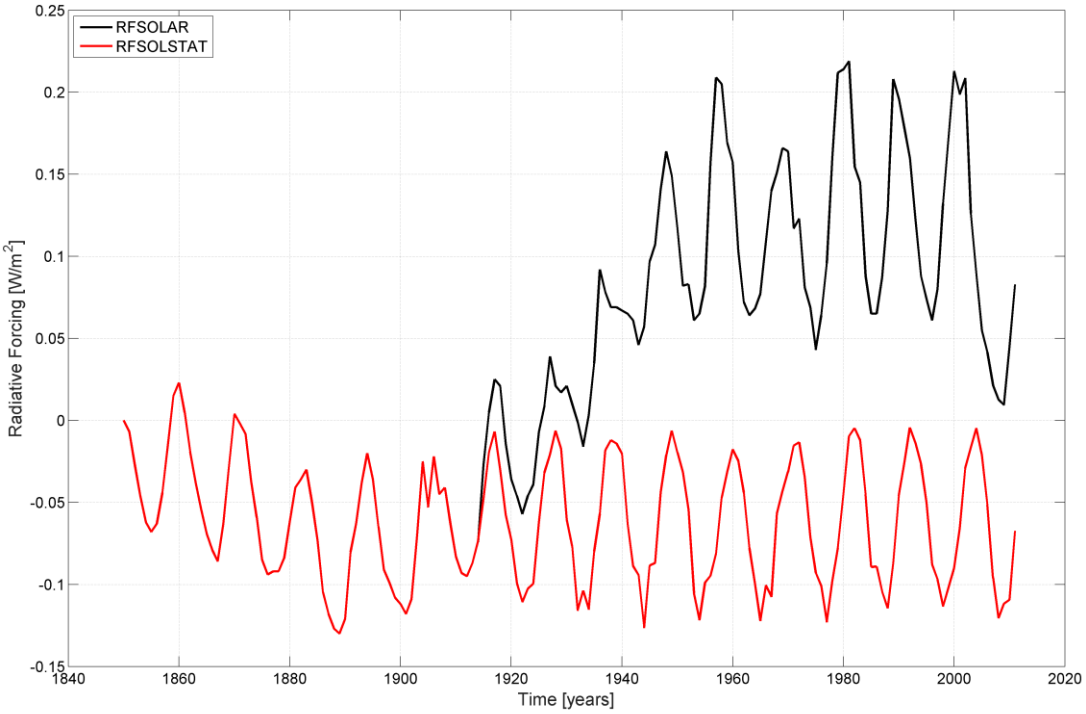


**Figure 14 External forcings used as input for the NN models: observed (black line) and reconstructed (red line) solar radiative forcing anomalies [W/m²]. The latter is synthetic time series under the assumption of stationarity (for attribution tests).**

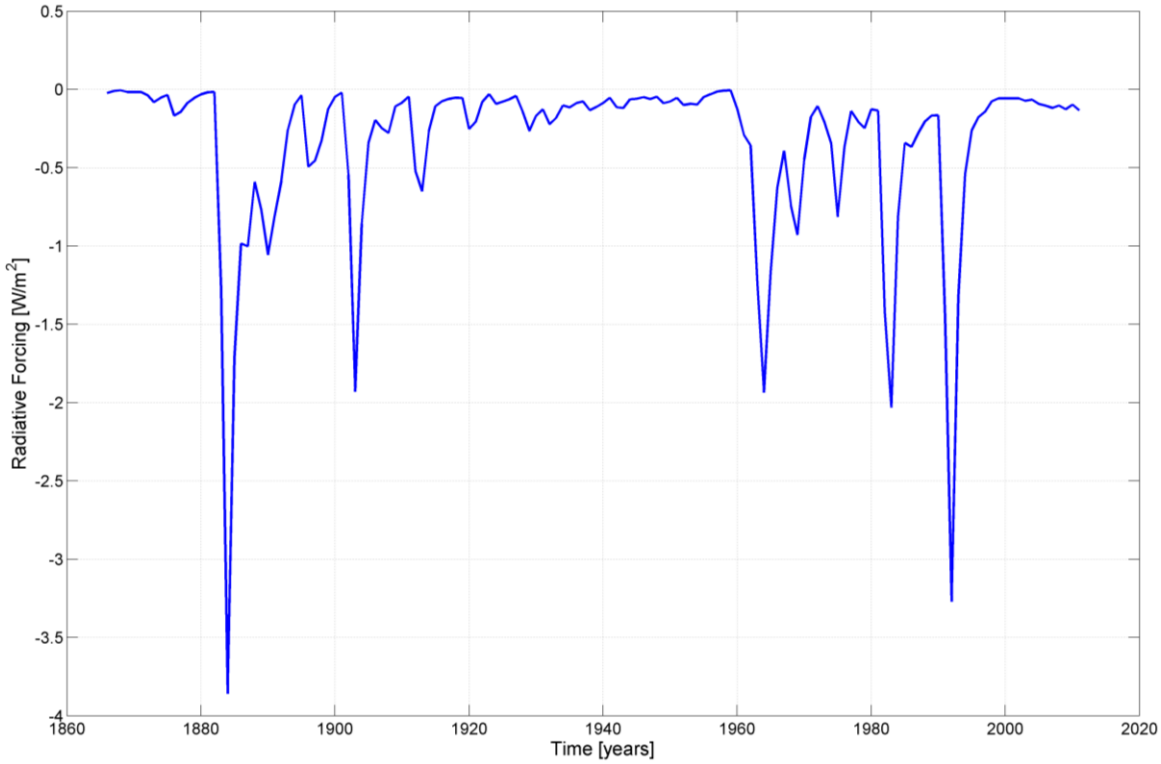


**Figure 15 External forcings used as input for the NN models: volcanic radiative forcing.**

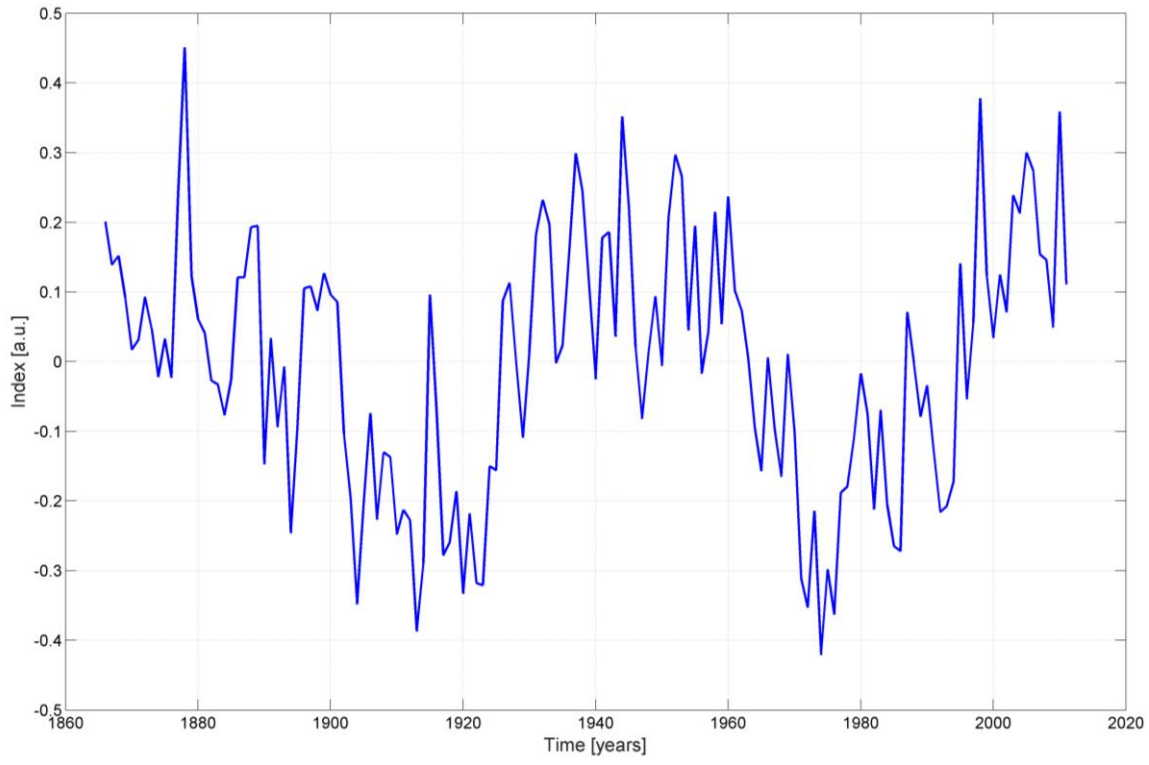


**Figure 16 The Atlantic Multidecadal Oscillation index.**

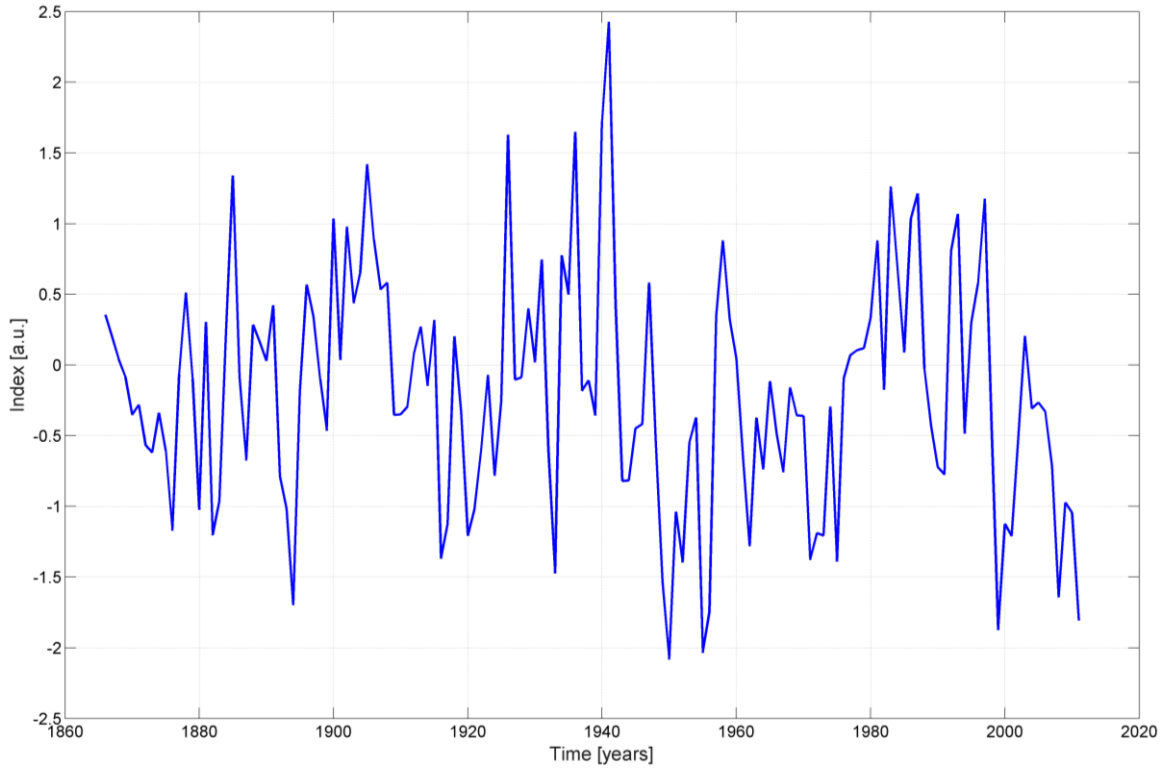


**Figure 17 The Pacific Decadal Oscillation index.**

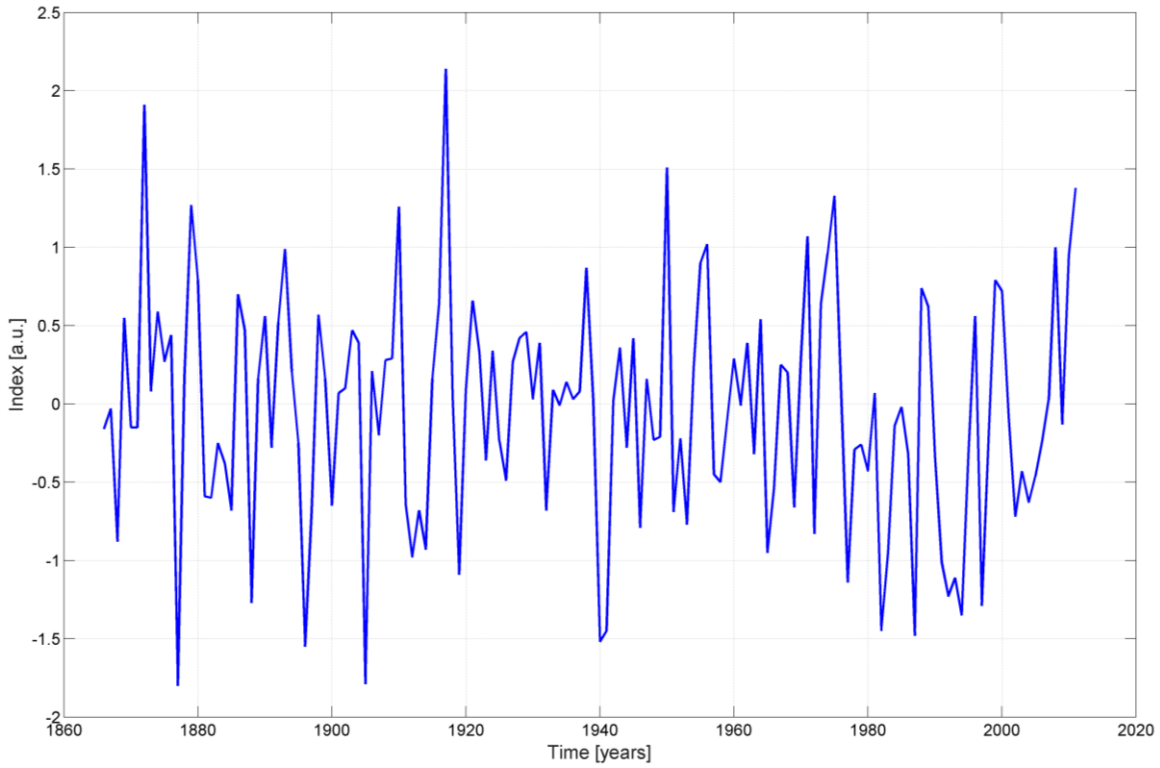


**Figure 18 The Southern Oscillation Index.**

As far as the NN model is concerned, it was well described in the previous Section 3.5. The optimization algorithm is the well-known Levenberg-Marquardt (see Section 3.3.4). We have a variable number of inputs depending on the problem addressed: the exact number will be specified for each application in the next sections. Again, we want to underline that the leave-one-out procedure adopted permits to obtain a real forecast of our temperature dataset, as the test data are unknown to the NN in the training phase. Together with the NN analysis we perform also multi-linear regression (as specified in section 3.5) in order to compare the two approaches.

## 4.2 Temperature reconstruction and attribution tests

Although previous studies have used statistical methods for the reconstruction of the mean global temperature [13], [75], [76], none of them have used an approach comparable to that of the GCMs. In fact, these studies aimed at finding how much variance of the observed temperature could be explained by anthropogenic and natural forcing separately, through the analysis of the performance achieved by NNs that make use the two different kinds of forcing as predictors.

In the world of the dynamical modelling the attribution is performed in two steps, first a validate model is chosen – i.e. able to reproduce the observed trend of mean global temperature – and then it is applied to simulate the temperature supposing that certain forcings remain fixed at their pre-industrial values. So, the idea is to choose a NN model able to reproduce the observed temperature with the observed forcings as predictors and then apply its transfer function – the validated NN model – to new inputs. Following the idea of the GCMs attribution tests, these inputs have values that show no trend since 1850. With this approach one can investigate the roles of the real changes in different forcings on the behaviour of temperature.

### 4.2.1 Temperature reconstruction

We have an ensemble of 20 models – i.e. different neural networks, see again section 3.5 – that supply temperature reconstruction for each year starting from 1850. These NNs are endowed with 3 different inputs – RFANTH, RFSOLAR, RFVOL (see Table 1), 4 hidden neurons and one output. The output must be compared with the observed mean global temperature (target). The number of the hidden neurons is empirically chosen, in any case, it must be kept low in order to maintain the NN small enough to avoid overfitting.

In Figure 19 we can immediately see the good performance of the NNs in reconstructing the observed temperature behaviour. For the ensemble mean we have a correlation $R = 0.913$ and a

Root Mean Squared Error (RMSE) of $0.109K$ (see Table 2 at the end of this subsection for details about all the runs and test performed for this topic). The ensemble mean can well reproduce the trends of the various periods with some exception. Among these the local maxima of 1878 and 1944 are underestimated, while the relative minimum of around 1910 is overestimated. Also, the interannual variability is quite well represented. A curious feature is the high variability visible in the ensembles of 1884 and 1992, both years following the two most intense volcanic eruptions of the considered period. This is understandable if we look at the "spot-like" series of RFVOL, because the reconstructions of these two years generally require big extrapolations respect to the other values of RFVOL.

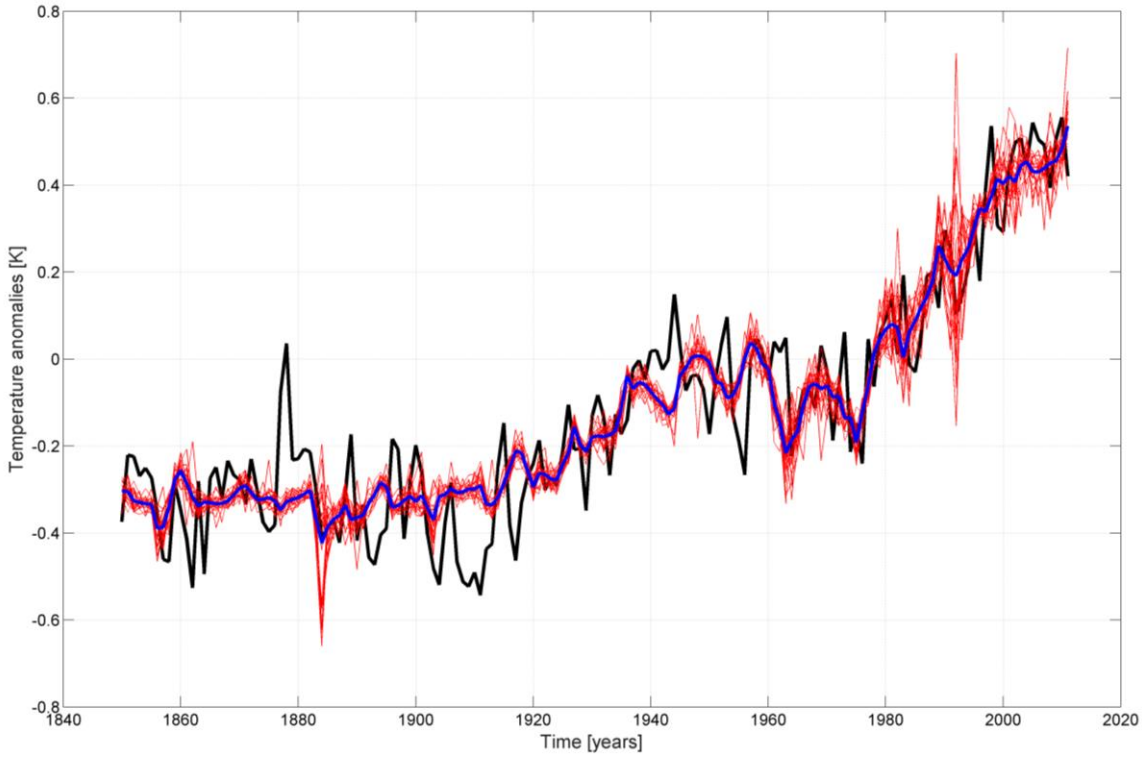


**Figure 19 Reconstruction of global T by NN models. In black the observed mean global temperature; in red the 20 NN ensemble runs; in blue the ensemble mean. As input the observed values of RFANTH, RFSOLAR and RFVOL are used.**

So, these NN models can well reproduce the observed variability of the mean global temperature. The attribution activity may start. Furthermore, we note that even the ensemble mean of these models – that tends to suppress nonlinearities respect to the single ensemble runs – perform better than the multilinear regressions using the same data. So, we are confident that nonlinearities play a role in the temperature series reconstruction.

| NN mode | Validation inputs/Attribution runs | R | RMSE (K) |
|---|---|---|---|
| Reconstruction | RFANTH-RFSOLAR-RFVOL | 0.913 | 0.109 |
| Attribution | Constant RFANTH | 0.645 | 0.247 |
| Attribution | RFSOLSTAT | 0.836 | 0.177 |
| Attribution | Constant RFVOL | 0.909 | 0.113 |
| | | | |
| Reconstruction | Constant RFANTH + residuals from RFGHG, RFBC, RFSOX | 0.905 | 0.113 |
| Attribution | Constant RFANTH + constant RFGHG in residual run | 0.489 | 0.291 |
| Attribution | Constant RFANTH + constant RFGHG, RFBC in residual run | 0.492 | 0.299 |
| Attribution | Constant RFANTH + constant RFSOX in residual run | 0.848 | 0.201 |
| | | | |
| Reconstruction | RFANTH – RFSOLAR – RFVOL + residuals from AMO, SOI, PDO | 0.944 | 0.089 |
| Attribution | Constant AMO in residual run | 0.918 | 0.106 |
| Attribution | Constant SOI in residual run | 0.937 | 0.094 |
| Attribution | Constant PDO in residual run | 0.941 | 0.092 |

**Table 2 List of all the test performed with our NN models. Details about each run is given into the text.**

### 4.2.2 Attribution tests

The first attribution test is the standard one, already mentioned previously in this thesis. We want to see what would been happened if the anthropogenic forcings had remained fixed to their pre-industrial values. In doing so, we use the validated NN models (network with all the weights fixed) and for each years of data we use as input the real observed values for RFSOLAR and RFVOL, while for RFANTH we use the value of 1850 – i.e. we use a constant series of data. In this way we simulate a world in which the natural forcings behave as observed and the human influence

is not considered. The result is shown in Figure 20. It is evident that with a fixed value for the anthropogenic forcings the recent rising of global temperature disappears and it is replaced by a constant trend since 1960.

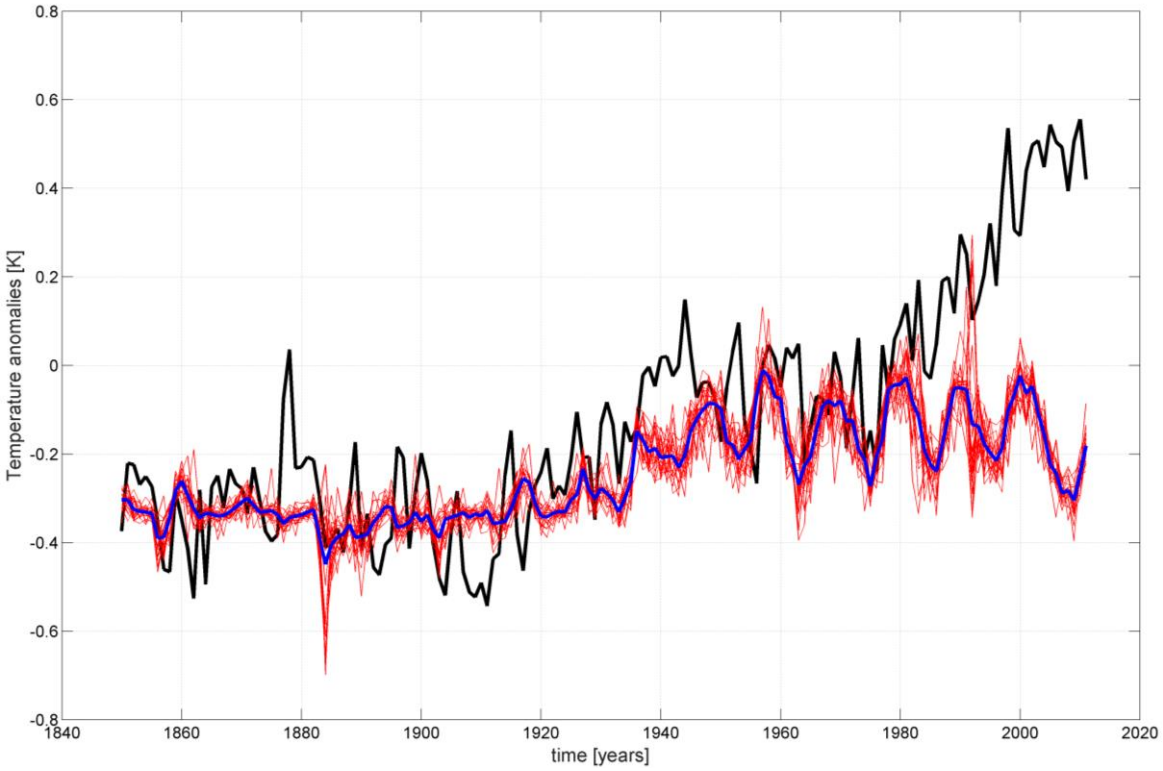


**Figure 20 Reconstruction of mean global temperature with the same validated models of previous Figure 19, but with a constant value for RFANTH (fixed at its preindustrial value in 1850).**

This behaviour is quite like the well-known analogous figure obtained by GCMs (see Figure 10.1(b) of the ref. [77]) with the same conceptual approach. Now, with a completely different method the same result is achieved: anthropogenic forcings had a fundamental role in the recent global warming.

As seen in section 4.1 RFANTH consist in three contributes, the radiative forcing of greenhouse gases (RFGHG), of black carbon (RFBC) and of sulphates (RFSOX) – see again Table 1. Obviously these three components have different magnitudes and so may have a different contribution to the trend of temperature. Furthermore, sulphates have a recognized cooling effect on global temperature behaviour. So, it would be interesting to investigate their single role. One could try to insert these three series as independent predictors in the construction of the NN models and then perform attribution experiments separately for each component. But in doing so we must construct bigger NN structures, increasing the risk to income into overfitting problems. Thus, we

choose a different approach. We consider the series of residuals given by the difference between the ensemble mean of the reconstructed temperature with the observed predictors (Figure 19, blue curve) and that of the attribution experiment with fixed RFANTH (Figure 20, blue curve). This residual time series is the approximation of the influence of total anthropogenic forcings on temperature. The idea consists in try to reproduce the residual series by an ensemble of NNs having the three anthropogenic forcings as predictors and then performing three attribution tests by separately keeping fixed each forcing at its 1850 value. The last step is to add the results of the attribution experiments to the temperature time series of the main attribution experiment (Figure 20, blue curve), in doing so we are able to explore the contribution of the three anthropogenic forcings since 1850 up to now (Figure 21).

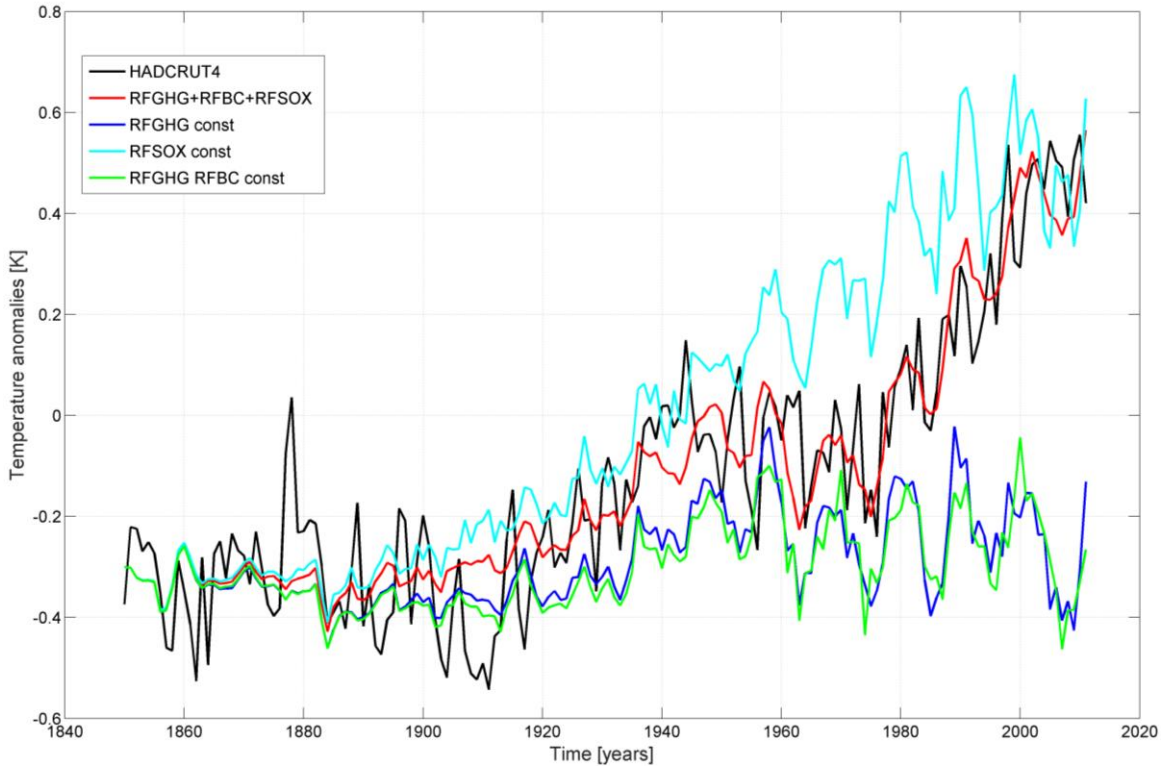


**Figure 21 The role of the distinct anthropogenic forcings discerned by the attribution experiment described into the text. Black line is the observed temperature; red line is the reconstructed T by all inputs with their real values; blue, light blue and green lines are the attribution runs with each forcing (one at a time) kept fixed.**

Thanks to Figure 21 we can appreciate the fundamental role of GHGs in reconstructing the mean global temperature behaviour, in fact with their value fixed at 1850 the temperature would have remained quite constant during the last part of 20$^{th}$ century. As far as concern BC radiative forcing, its role is marginal, in fact there are no appreciable differences respect to the case with only GHGs constant (see blue and green curves in Figure 21). Another interesting detail is that the curve

obtained with GHGs constant (or with both BC and GHGs) is lower than that of the main attribution test (blue curve of Figure 20) by about 0.1K. This is due to the cooling effect of sulphates, since in the above cited test we consider only RFSOX with its real values. This cooling effect is evident also from the light blue curve in Figure 21, it corresponds to the test with RFSOX kept fixed at 1850 values. Without sulphates the *hiatus* during the period 1945-1975 would have disappeared and the increase of global mean temperature would be monotonic. We can quantify in 0.9K the warming effect of GHGs at the end of the considered period and in about 0.3K the cooling effect of sulphates at the end of the hiatus period.

After discussing the role of the anthropogenic forcings, now we deal with that of natural forcings – i.e. due to sun and volcanoes. In the previous Figure 14 we can see the observed solar radiative forcing of last 150 years (black curve). A peculiar characteristic is the eleven years period oscillation, due to the sunspot cycle, that certainly influences the decadal variability. However, the most important feature of the black curve in Figure 14 is the transition from a low power regime to a high one during the period 1910-1950. Obviously, the question is about the possibility that such transition may have had an influence on the temperature behaviour. We can investigate this issue by NN modelling (for the first time at our knowledge). To understand the impacts due to this transition we create a synthetic time series of the solar radiative forcing – called RFSOLSTAT – which is stationary without the transition towards a high-power regime. RFSOLSTAT is built in terms of a first-order Fourier series based on the first 65 years available of the observed RFSOLAR (red curve in Figure 14).

At this point we use the same consolidated NNs endowed with the three inputs RFANTH, RFSOLAR and RFVOL and to understand the role of the sun we supply these validated ensembles of models with RFSOLSTAT, the new output is shown in Figure 22. With a stationary solar forcing the increasing trend in temperature during the period 1910-1945 is not more caught. Furthermore, the hiatus during the 1945-1975 period is about 0.2K lower than the observed one. The last remarkable result is that – despite to a stationary sun – the strong increase of mean global temperature of the last 40 years anyway occurs.

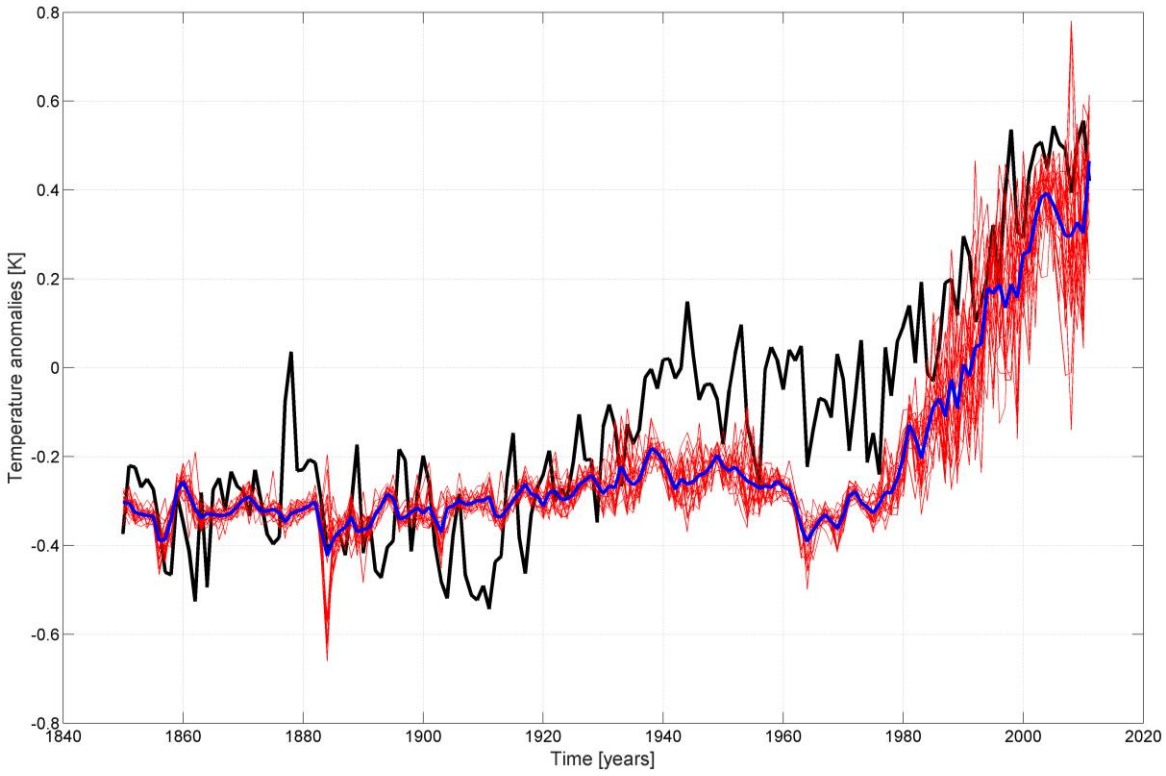


**Figure 22 Results of the attribution experiment for the establishment of the solar influence. The black line is the observed T; the red lines are the ensemble members; the blue line is the ensemble mean. These results are achieved with a stationary solar forcing (RFSOLSTAT).**

The same rationale has been also applied to verify the influence of the radiative forcing due to volcanic eruptions. In Figure 15 the observed value for RFVOL is shown, in order to see what happens without volcanoes we supply the validated NNs with a constant series for RFVOL with zero values – i.e. no eruptions at all in the considered period. The results (not shown) are very similar to that of Figure 19, no appreciable role for volcanic emissions appears if not for interannual variability.

To conclude this part, we underline again the main role for anthropogenic forcing in the observed recent global warming, in particular the warming role of GHGs and the cooling effect of sulphates. Furthermore, our innovative approach permits also to understand more about the role of the Sun. The influence of the switch to a high-power regime appears clear for the rising of temperatures in the middle of the 20$^{th}$ century, in the same way the lack of influence on the recent warming appears clear. In order to have a quantitative view to our results see again Table 2, in which the value of R and of the RMSE helps to understand the goodness of our results. In particular in the attribution tests a high RMSE tell us that the synthetic variable into exam has a stronger influence respect in comparison to a test with a lower RMSE. In the next subsection we see how it is possible to study the role of natural variability – i.e. AMO, SOI, PDO – in reconstructing the interannual variability of the observed temperature series.

### 4.2.3 The role of natural variability

As said in subsection 4.2.1 and as we can see in Table 2, our NN models are able to explain about 83% of the variance ($R^2$) of the global mean temperature through natural and anthropogenic forcings. The remaining unexplained variance is probably due to internal (or natural) variability, so the idea is to investigate the role of three main indices of natural variability – i.e. the Atlantic Multidecadal Oscillation, the Pacific Decadal Oscillation and the Southern Oscillation Index. Again, we could try to reconstruct the temperature with a NN endowed with both natural and anthropogenic forcings and with these three indices, but, as previously mentioned, such a network could be too much large in comparison to the dimension of our dataset. So, we follow an approach similar to that for the attribution assessment: we consider the series of the residuals of the observed (black curve of Figure 19) minus the reconstructed temperature (blue curve of Figure 19) and try to reconstruct it by a NN with AMO, PDO and SOI as predictors. The results are satisfactory (Figure 23). Quantitatively if the ensemble mean of the reconstructed residuals is added to that of the reconstructed temperature (Figure 19, blue curve) we obtain $R = 0.944$ and $RMSE = 0.089$, so now the explained variance is about 89%. In particular (see Figure 24), the improvement respect to the main test in catching the interannual variability is evident in all the period and in the period 1945-1975.

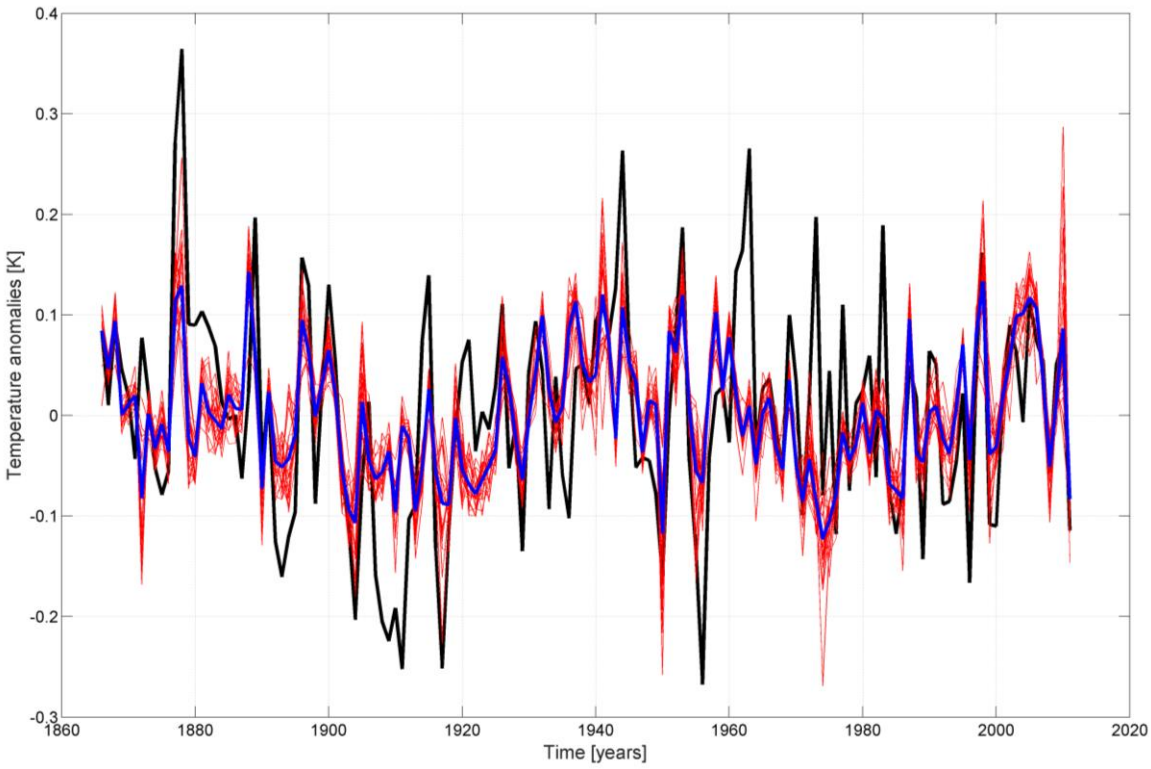


**Figure 23 Results about the role of natural indices on the residual variability not caught by the external (natural and anthropogenic) forcings.**

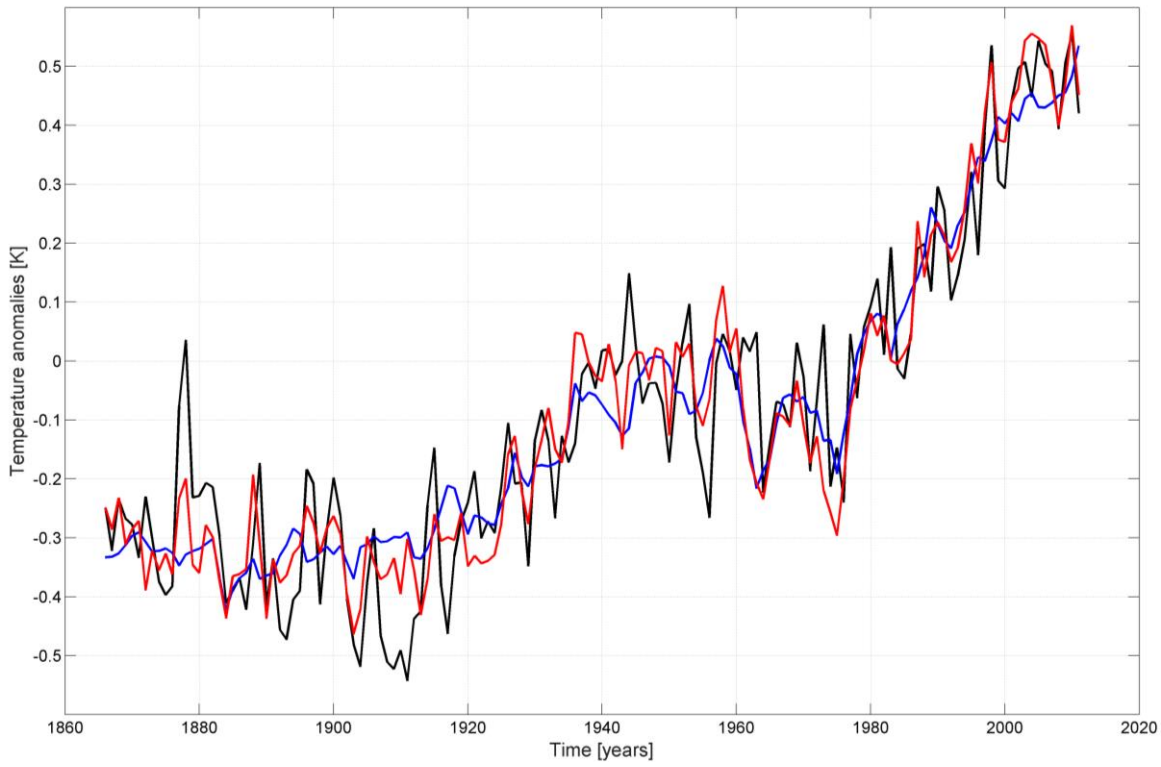


**Figure 24 Reconstruction of the mean global temperature using also data about natural circulation indices. Observed temperature (black line); ensemble mean from NN models with external forcings only (natural and anthropogenic) as input (blue line); ensemble mean when the results of the NN model with the indices of natural variability is added to the previous results (red curve). Note that because data for SOI are available since 1866 this reconstruction starts by this year.**

The role of the single indices is investigated by the well-established approach. In fact, we can apply the attribution strategy by keeping at a fixed value one predictor at a time. So, the validated models with the observed values for AMO, SOI and PDO are used with real values for two of them and a constant value for the remaining index. The results of the three tests are presented only in a quantitative way (see again Table 2 and see supplementary Figure S4b-d in [16]). It is evident the prominent role of AMO in catching both interannual and decadal variability of the temperature, this is consistent with results existing in literature [78], [79], [80] attesting the role of AMO as the main "natural" driver for global mean temperature.

With this approach we disregard the possible influences and relationships among the changes in the external forcings and these indices of natural variability. This could be considered assuming forcings and indices at the same level – i.e. by putting them as inputs of the same NNs. But, as said, in this way we must build larger NNs with consequent overfitting problems. However, due the great influence of AMO, we decide to further investigate only its role, by adding it as predictor in our NN models together with the external forcings. Using a structure 4-4-1 the

temperature reconstruction benefits from the inclusion of AMO as input, reaching similar performances of the so-called residual strategy ($R = 0.943$, $RMSE = 0.089$). A notable result is now achieved when we perform the attribution by keeping fixed AMO at its mean value. In Figure 25 is evident that multidecadal variability is not caught well, in particular in the periods 1910-1945 and in the hiatus of 1945-1975. About the latter, AMO probably have a strong relationship with sulphates, this possible relationship will be further investigated in Chapter 5.

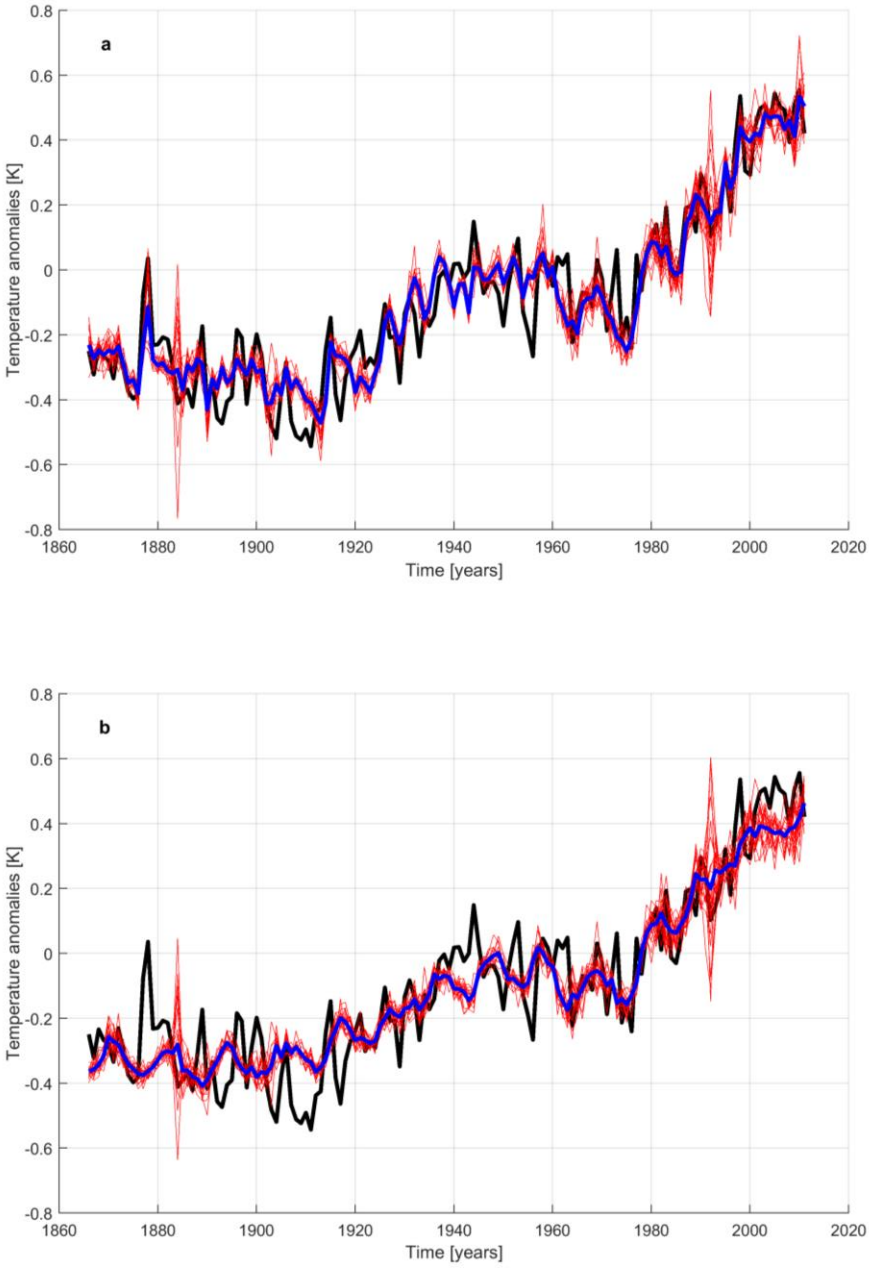


**Figure 25 Reconstruction of mean global temperature with anthropogenic and natural forcings and AMO as predictors. Black line is the observed temperature; in red the ensemble members; in blue the ensemble mean. (a) with observed values of RFANTH, RFSOLAR, AMO as inputs. (b) attribution test with AMO fixed at its mean value.**

## 4.3 Sensitivity tests

In the previous section, data from reliable sources are used as anthropogenic forcing factors. Nevertheless, notable uncertainties are still present in these data, in particular as far as sulphates and black carbon are concerned. Thus, we perform some sensitivity tests with different datasets (see section 3.1) to analyse if our main results are robust. The use of different datasets can be considered as the test of a range of uncertainties for the predictors themselves.

We consider the results of NN ensembles when the networks are endowed with two different time series for the anthropogenic radiative forcing. We make use of the anthropogenic forcing factors used in the Coupled Model Intercomparison Project 5 – CMIP5 [72] – and in a work by Hansen et al. [74] (hereafter RFANTHCMIP5 and RFANTHANSEN respectively). With reference to Figure 26 we can consider the first as an upper limit for the anthropogenic forcing while the second is more similar to that used in this work. In both cases the cooling role of sulphates is clearly less important, especially during the 1945-1975 period.

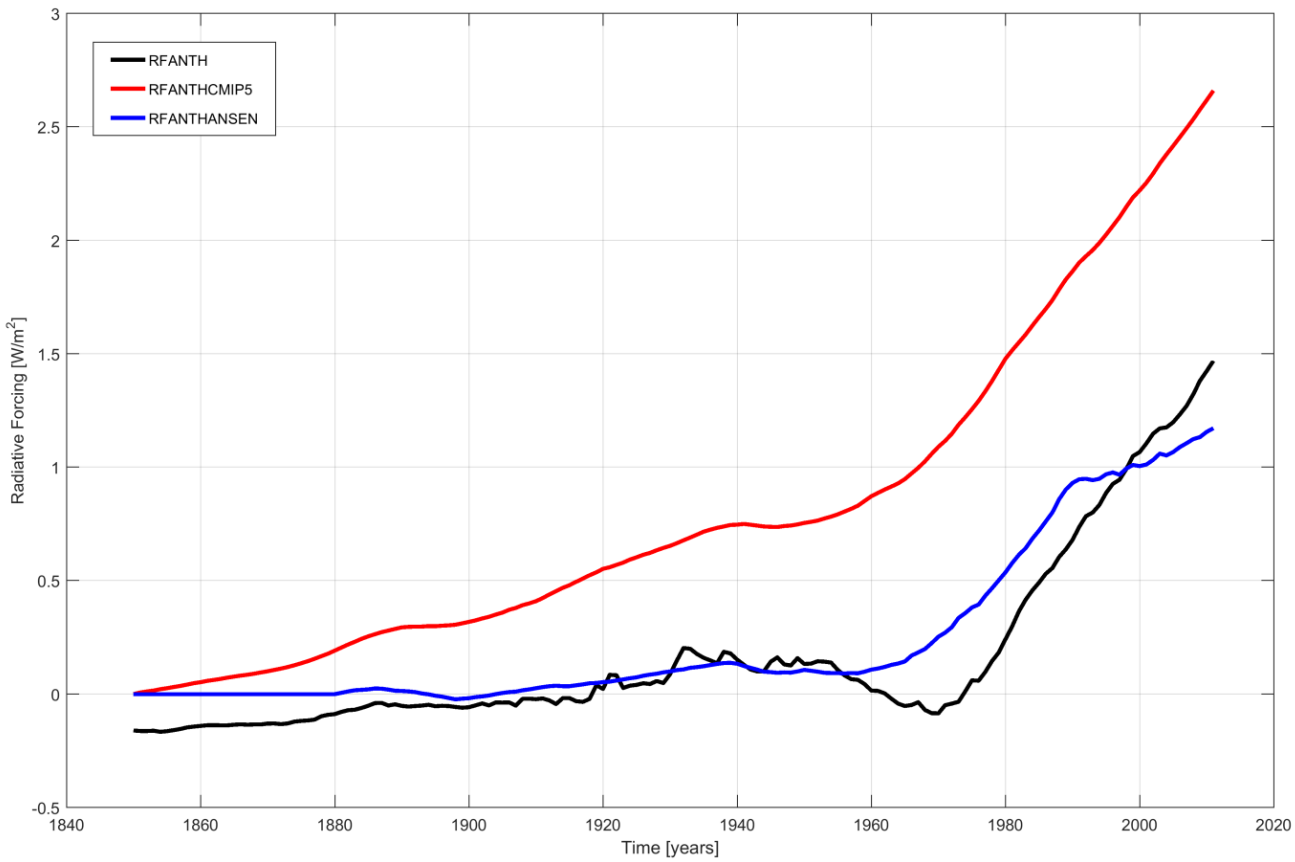


**Figure 26 Anthropogenic forcings used in the NN sensitivity tests**

With the usual NN structure 3-4-1 we perform the sensitivity tests using as inputs RFANTHCMIP5, RFSOLAR, RFVOL and RFANTHANSEN, RFSOLAR, RFVOL respectively. In Figure 27 and 28 we show the results of the two tests for the temperature reconstruction and the respective attribution with anthropogenic forcing kept fixed.

With RFANTHCMIP5 as predictor the NNs are still able to reproduce the observed temperature ($R = 0.912$, $RMSE = 0.109$). However, the ensemble mean shows a monotonic rising trend with a bad representation of the 1945-1975 hiatus, in fact only a slowdown of warming can be seen in Figure 27a. This is due to the reduced influence of sulphates in this dataset. The attribution runs show the same trend in recent years as the main test of Figure 20 but with lower values of anomalies. This is because now the model for temperature reconstruction has been trained with higher values of anthropogenic forcings in the last decades and the difference in RFANTHCMIP5 between 1850 and the more recent years is bigger.

With RFANTHANSEN the performance for the reconstruction is satisfying ($R = 0.909$ and $RMSE = 0.111$) and again we have a bad reconstruction for the hiatus period (1945-1975). The attribution test is now very similar to our main test with the "usual" constant trend in the recent decades.

At this point we can make some considerations. Even if with some differences, the main result of our work is confirmed also by the sensitivity tests. The recent global warming appears due fundamentally to the anthropogenic forcings, in fact all the attribution runs with the three datasets show similar results.

The second point concerns the hiatus on the period 1945-1975. We see that with CMIP5 and Hansen datasets we are no more able to well reproduce this period. So, the question is if with the consideration of the natural variability we are still able to reproduce at best the hiatus. So, again we apply the "residual strategy" previously used, when we try to reproduce the residual time series – observed minus simulated temperature – by NNs using the three natural indices as predictors or in an equivalent way using NNs with external forcings and AMO as input to reproduce the observed temperature. The results of the residual strategy show that using the three indices of natural variability we are now able to reconstruct also the hiatus, even with RFANTHCMIP5 and RFANTHANSEN. The main role of AMO is still confirmed, in fact using constant AMO brings to not caught the hiatus at all by both CMIP5 and Hansen dataset.

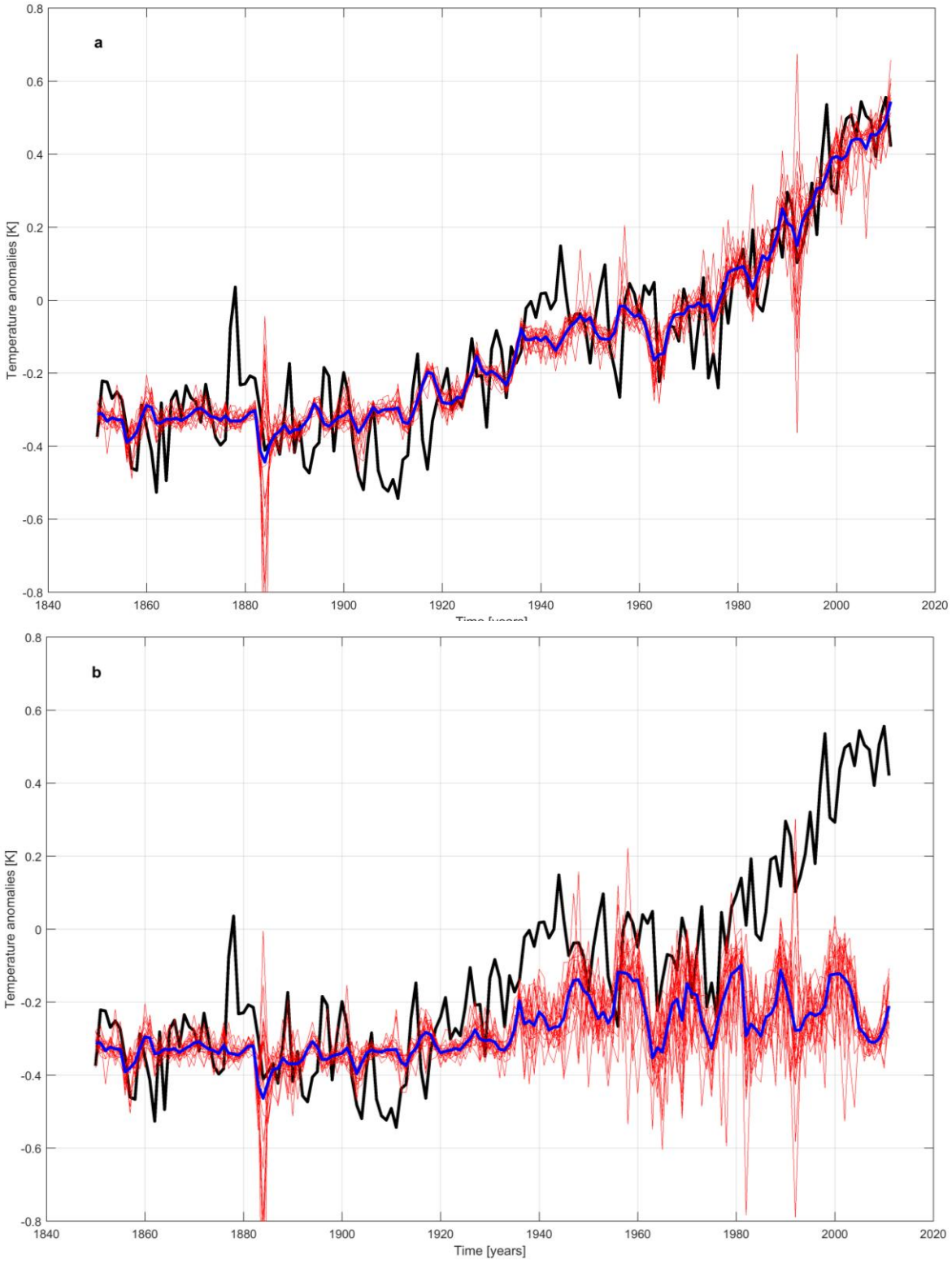


**Figure 27 Reconstruction of mean global temperature by NN models with CMIP5 anthropogenic forcing. In black we have the observed temperature; the red lines are the ensemble member of NNs; in blue we have the ensemble mean. (a) is obtained with the observed values of RFANTHCMIP5, RFSOLAR, RFVOL. (b) is the attribution test, with RFANTHCMIP5 fixed at its preindustrial values of 1850.**

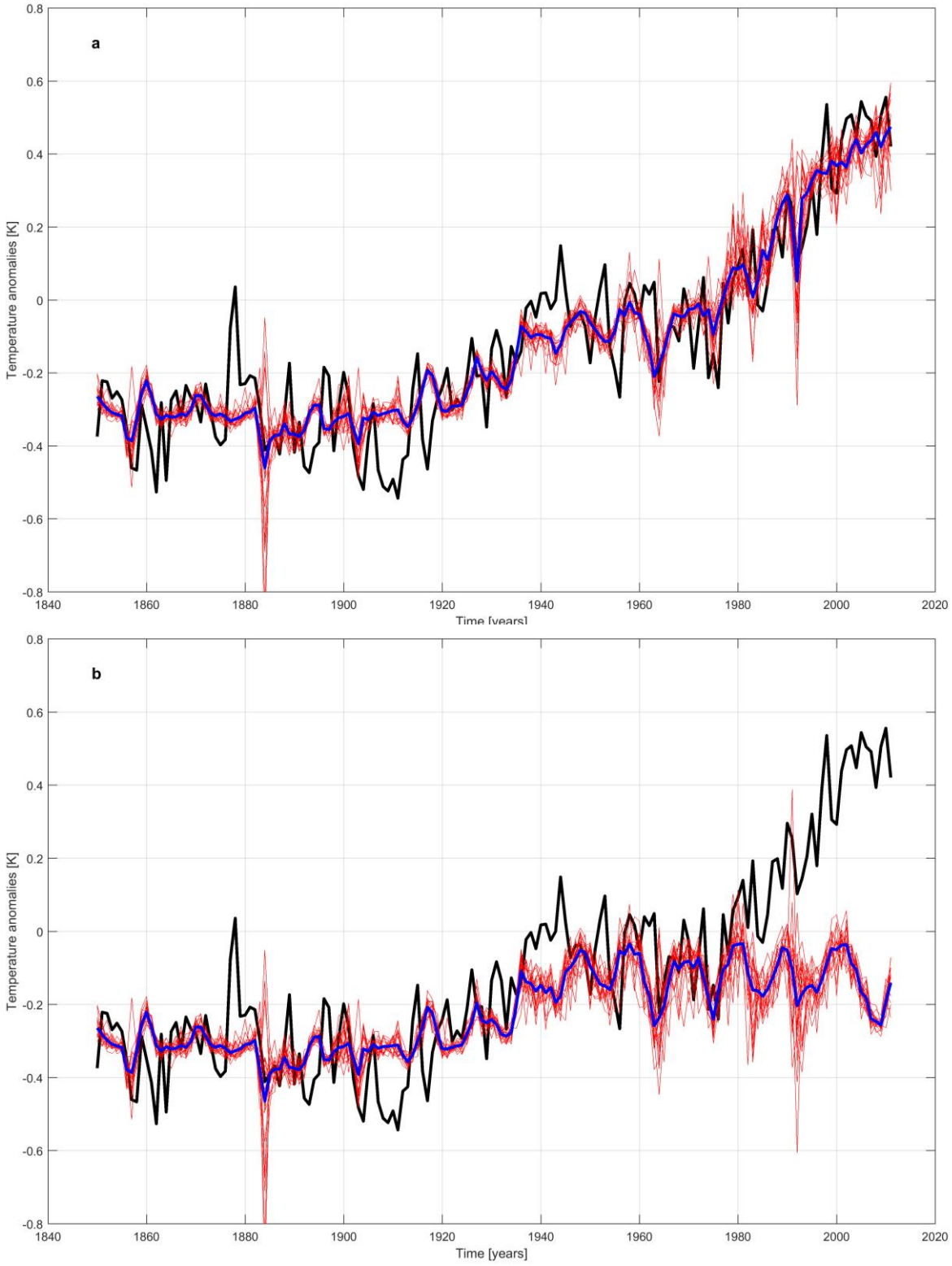


**Figure 28 Reconstruction of mean global temperature by NN models with Hansen anthropogenic forcing. In black we have the observed temperature; the red lines are the ensemble member of NNs; in blue we have the ensemble mean. (a) is obtained with the observed values of RFANTHANSEN, RFSOLAR, RFVOL. (b) is the attribution test, with RFANTHANSEN fixed at its preindustrial values of 1850.**

To conclude this section, we underline that our tests are not completely able to clarify the role of AMO and sulphates in the hiatus period. In fact, the period 1945-1975 would seem to be reconstructed by AMO and sulphates separately, and in particular the contribute of the latter depends on the chosen dataset. So, no conclusion is possible on which driver is the most influential on the 1945-1975 hiatus. In Chapter 5 we further analyse the relationship between AMO and sulphates, as we will see it is a recent hot topic of debate.

## 4.4 Final results and conclusions

To conclude this Chapter about the attribution of the last 150 years temperature trend, we show two other results and we make several interesting observations about our results.

First, the use of a fully nonlinear method permits to investigate about the dependence of the temperature on the forcings. We can explore if the temperature response to high or low forcing is linear or nonlinear. The sensitivity studies of section 4.3 are the framework for this kind of analysis. As we said, in parallel to NN models we perform multilinear regression, comparing the results on both methods in the main reconstruction of temperature behaviour permits to draw some conclusions. We focus our analysis to period in which the anthropogenic forcings and temperature are undoubtedly increasing – i.e. since 1976 – and we compare the performances of both models in this period. In Table 3 we present the results in term of correlation (R).

| Anthropogenic RF | NN R | Multilinear R |
|---|---|---|
| RFANTHCMIP5 | 0.895 | 0.899 |
| RFANTH | 0.901 | 0.888 |
| RFANTHANSEN | 0.884 | 0.833 |

**Table 3 Performance of temperature reconstruction in terms of R of the recent warming period (1976-2011) with the three different datasets used in the sensitivity tests.**

The NNs can reconstruct T better than the multilinear regression when the forcing is relatively low (RFANTH and RFANTHANSEN), while R is similar when the forcing is relatively high (RFANTHCMIP5). This suggest a quasi-linear dependence of temperature on forcings when the latter are strong, while a nonlinear dependence appears with lower forcings. At this level of investigation, we may only suppose that some feedback mechanisms could act with lower forcings, but further research seems necessary on this interesting topic.

Another interesting question, to which we reply in this final discussion, is about the possibility of some temporal lag effect in our results. It is well known that a monthly temporal resolution many lags have been identified between forcing factors and global temperature [81]. On the other hand, at annual time resolution, the indices of natural variability do not seem to show any delayed influence on global temperature [82]. In general, the annual lags are considered only for RFANTH (see again [81]). Thus, we have performed NN tests by using a record (RFANTHLAG10) to be synchronized in our models with RFSOLAR, RFVOL and T related to ten years later. In short, our results show a decreasing in performance with RFANTH delayed, in accordance to recent studies in which a time delay is present in the response of global temperature to emissions while not to concentrations and the related radiative forcings [83].

Finally, we come to the conclusions of this chapter. The present investigation with a completely data driven approach by NN models shows that the attribution results achieved by GCMs about the main role of the anthropogenic forcing are robust and reliable. This, together with our new results about the role of the Sun on sub periods (irrelevant for the trend of the last decades and quite strong for the middle of the $20^{th}$ century) and with the detection of the role of AMO and sulphates on the hiatus period of 1945-1975, gives a more complete picture in attribution studies. A further contribution about the relation between AMO and sulphates will be given in next Chapter 5. Furthermore, the sensitivity study permits to establish the robustness of the NN model itself in reference to the main available anthropogenic datasets.

# 5. The problem of the Atlantic Multidecadal Oscillation, is it real natural variability?

As seen in the previous chapter, the climate system is endowed with complex dynamics, characterized by both internal natural variability and interactions with external (natural and anthropogenic) forcings. We also see that the recent global warming has been mainly driven by external anthropogenic forcings, even if internal variability can have a role in determining temperature values at short time scales (interannual to multidecadal ones). For example, an unforced internal component that varies on multidecadal time scales has been identified in the temperature record [84], the warming and cooling of this component matches that of the Atlantic Multidecadal Oscillation (AMO).

The AMO can be identified as a mode of natural variability occurring in the North Atlantic Ocean, with an estimated period of 60-80 years. It is calculated on the average anomalies of the sea surface temperatures (SST) in the North Atlantic basin, in general over 0-80N, in this way it is possible to define an AMO index (Figure 29). The climate change signal is removed from the AMO index, by detrending SST data at each grid point or detrending the spatially averaged timeseries.

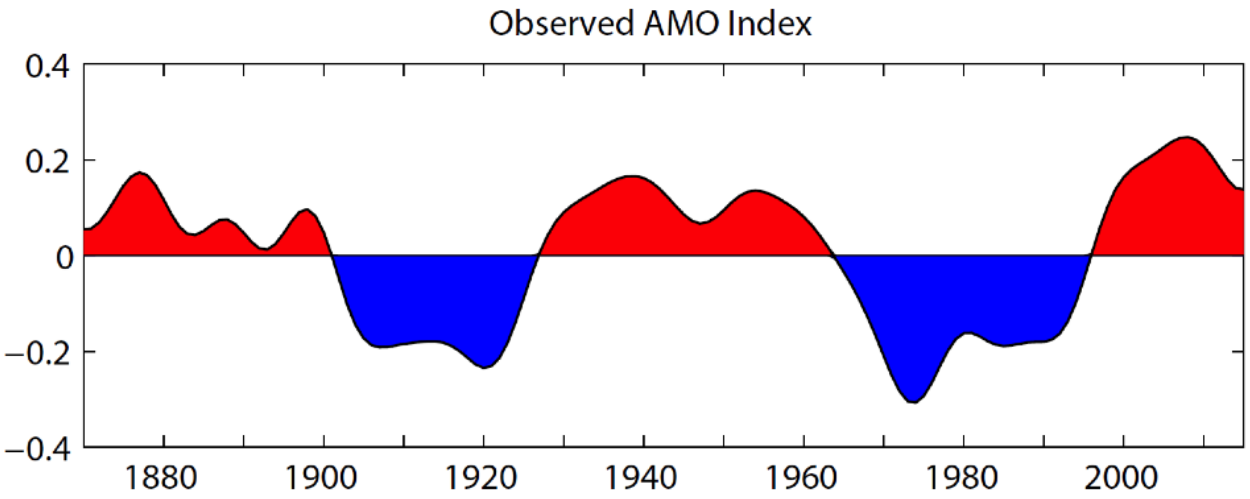


**Figure 29 Observed AMO index, defined as detrended 10-year low-pass filtered annual mean area-averaged SST anomalies over the Norh Atlantic basin (0N-65N, 80W-0E). From NCAR UCAR climate data guide.**

AMO seems to have a role in modulating global temperature, but in recent years a question was raised: may the forcings which recently drive the mean global temperature behaviour also lead the AMO? This dilemma can hopefully be solved by attribution studies of the AMO behaviour

itself. Several studies faced with this problem: [85], [86], [87], [88], [89], [90], [91], [92], [93] and [94], all of them make use of attribution experiments performed by dynamical models. Some of these papers emphasized the role of sulphate aerosols in driving the AMO as for example [86] that underline the role of both anthropogenic and volcanic sulphate aerosols. Other works, as [90], [92] and [93] are also in agreement with the dominant role of external anthropogenic forcings to reproduce the observed AMO trend. On the other hand, other papers claimed a role for ocean dynamics and internal factors of the climate system as the atmospheric circulation [85], [87], [88], [89], [91] and [94], between them we note that in [85] the main drivers are external forcings as the solar radiation and the volcanic aerosols.

We stress again that all these studies are performed with dynamical GCMs. Furthermore, using the NN model described in the previous Chapter, a first attempt of attribution test for establishing the drivers of AMO has already been performed [16]. In this work external forcings (see section 5.1) were used as inputs and, as said, AMO was the target. In [16], the reconstruction of AMO is not extraordinary, so here we develop a new NN model to improve the reconstruction itself and to strengthen the attribution results.

In this framework, here we show the results obtained considering a new version of our NN model. A quasi-Newtonian backpropagation method is used, which is more suitable for small dataset. In this way we can reconstruct better than [16] the trend of AMO. Interesting results are obtained during the attribution tests, if sulphates are fixed at their value of 1866, the oscillating signal of AMO disappears, so claiming a big role of these anthropogenic aerosols for driving the AMO behaviour. While if we keep GHGs constant at their pre-industrial value, only a small oscillation is visible in the simulated AMO signal. We also perform sensitivity tests with different datasets to obtain robust results. Furthermore, a first forecast experiment attempt for AMO up to 2100 is performed, by making use of Representative Concentration Pathway (RCP, see section 5.1) scenarios.

Even if deeper insight into this topic is necessary (especially for the forecast attempt), we can conclude that the AMO signal appears to be modulate by sulphates forcing, with a lower contribution by GHGs. At this stage no role for natural forcings has been detected. For AMO, natural variability seems not to be really natural. In the next sections we show the results of our work.

## 5.1 Data and methods

As said in the introduction of this chapter, the first attempt to reconstruct and make attribution tests for AMO by NN modelling has been performed in [16]. Here we use a modified version of the NN model used in [16]. As far as the dataset of our main test is concerned, we can refer to the previous section 4.1. The target of NN models and MLR is now the Atlantic Multidecadal Oscillation. As predictors we make use of solar radiative forcing, volcanic forcing, sulphates forcing and GHGs forcing: all of them (included AMO) are fully described in section 4.1. In Table 4 we see the predictors and the target for the study of this chapter. An explanation for the rationale behind the choice of the predictors will be furnished in the next section.

Sensitivity tests are performed to get more robust results. The two datasets from Hansen and CMIP5 previously used for mean global temperature attribution (Chapter 4) are again employed, so for the details see again section 4.1.

In section 5.3 we show the results for AMO forecasts attempt. In doing so, we use the Representative Concentration Pathway (RCP) scenarios. A single RCP is a greenhouse gas concentration (not emissions) trajectory (or forecast) starting from 2007 up to centuries, adopted by the Intergovernmental Panel for Climate Change (IPCC) for its fifth Assessment Report (AR5) in 2014. In short, a RCP represents a possible hypothetical future scenario for emissions of gases related to climate change. Here we consider RCP scenarios up to 2100. From the emissions it is possible to calculate the related radiative forcing thanks to standard formulae. We have four different scenarios (Figure 30, http://www.ipcc-data.org/index.html), RCP 2.6, 4.5, 6.0, 8.5, for example RCP 2.6 means that we will have a radiative forcing of 2.6 $W/m^2$ at the year 2100. Data for RCP scenarios are freely available from the IPCC website.

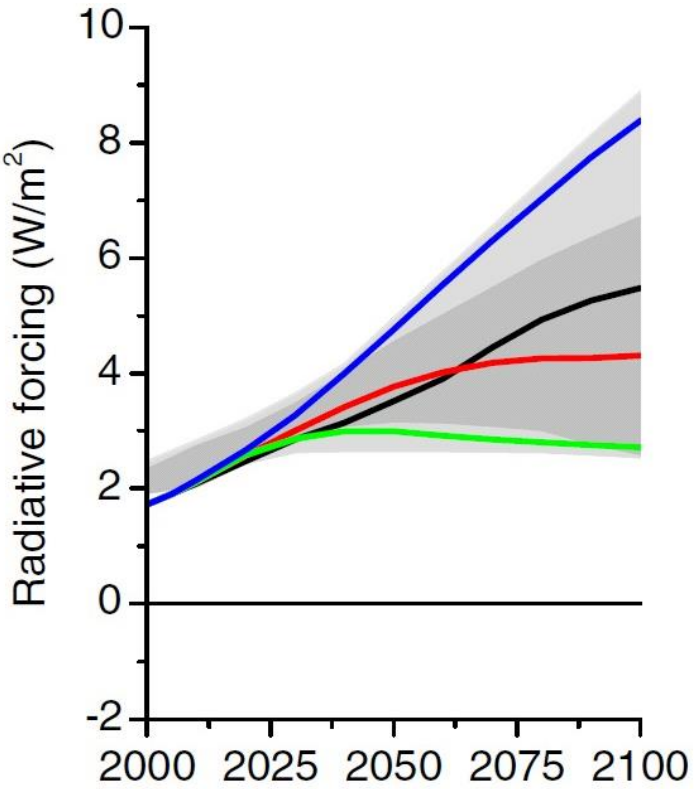


**Figure 30 The four RCP scenarios from 2007 up to 2100. In green the RCP 2.6, the red is RCP 4.5, the black curve represents RCP 6.0 and the blue curve the RCP 8.5.**

Clearly, the RCP 8.5 represents the worst scenario – or the so-called *business as usual* – while the RCP the better one. As explained in section 5.3, the observed part (up to 2006) is used in training and validation phase for NN models, then future projections (scenarios) up to 2100 are used for the forecast of AMO.

| | **List of variables** |
|---|---|
| **1** | Radiative forcing of greenhouse gases and of black carbon – RFWARM |
| **2** | Radiative forcing of anthropogenic sulphates – RFSOX |
| **3** | Radiative forcing of solar activity – RFSOLAR |
| **4** | Radiative forcing of volcanoes – RFVOL |
| **5** | Atlantic Multidecadal Oscillation – AMO |

**Table 4 List of the variables used for AMO reconstruction. In the last row we find the target for our study (AMO).**

The previous results for AMO reconstruction [16], were not fully satisfactory. With the introduction of a different back-propagation algorithm we are now able to improve that results. In detail, our NN model has the usual structure described in depth in section 3.5. The idea is to make use of a quasi- Newtonian back-propagation algorithm, the Broyden-Fletcher-Goldfarb-Shanno (BFGS) is used (see sub-section 3.3.3). As explained in the cited subsection, with this approach more memory capabilities are requested, but for our relatively small dataset the BFGS permits however to have more advantages. So, the usual strategy is adopted, NNs and MLR are applied to our predictors to reconstruct the observed AMO behaviour, then the attribution strategy is applied to investigate which predictor has the main driving role.

## 5.2 Discussion of AMO reconstruction and attribution tests

The guideline of this chapter is trying to reconstruct the observed trend of AMO (Figure 29) by NN modelling and MLR. Furthermore, we also perform attribution activity. The predictors will be: the sum of RFGHG and RFBC (called RFWARM), they are the gases that have a positive radiative forcing value – i.e. a warming effect on the temperature; RFSOX that has a cooling effect on the temperature; RFSOLAR and RFVOL. The rationale of the choice to kept separated the warm and cool parts is related to the different physical impacts that both have on the global temperature, furthermore, with this setting, we can perform attribution and discern on the different contributions.

The main test is the reconstruction of AMO using the above-mentioned predictors. In Table 5 we show the performance of our NN and MLR models in terms of R-squared, in that table we can also see the performance of our models with the different dataset listed in section 5.1. In Figure 31 the results of AMO reconstruction with the Stern dataset are shown, with an R-squared of 0.59 (second row of Table 5), the ensemble mean of NNs can explain about 60% of the observed AMO variance. It is a good result in comparison to the previous version of the NN tool [16]. As said in section 5.1, we make use of CMIP5 and Hansen datasets to give more robustness to our results. In Table 5 we can also see that for the alternative datasets the performances are good. As far as concern MLR the values of R-squared are very low, the comparison with NN suggests that some nonlinear mechanism is at the basis of the relationship between the predictors and AMO. This confirm the goodness of the choice of a NN model.

| Dataset | NN $R^2$ | MLR $R^2$ |
|---|---|---|
| **Stern** | 0.59 | 0.26 |
| **CMIP5** | 0.58 | 0.12 |
| **Hansen** | 0.58 | 0.05 |

**Table 5 Performances of NN and MLR models, for the three dataset used for AMO reconstruction.**

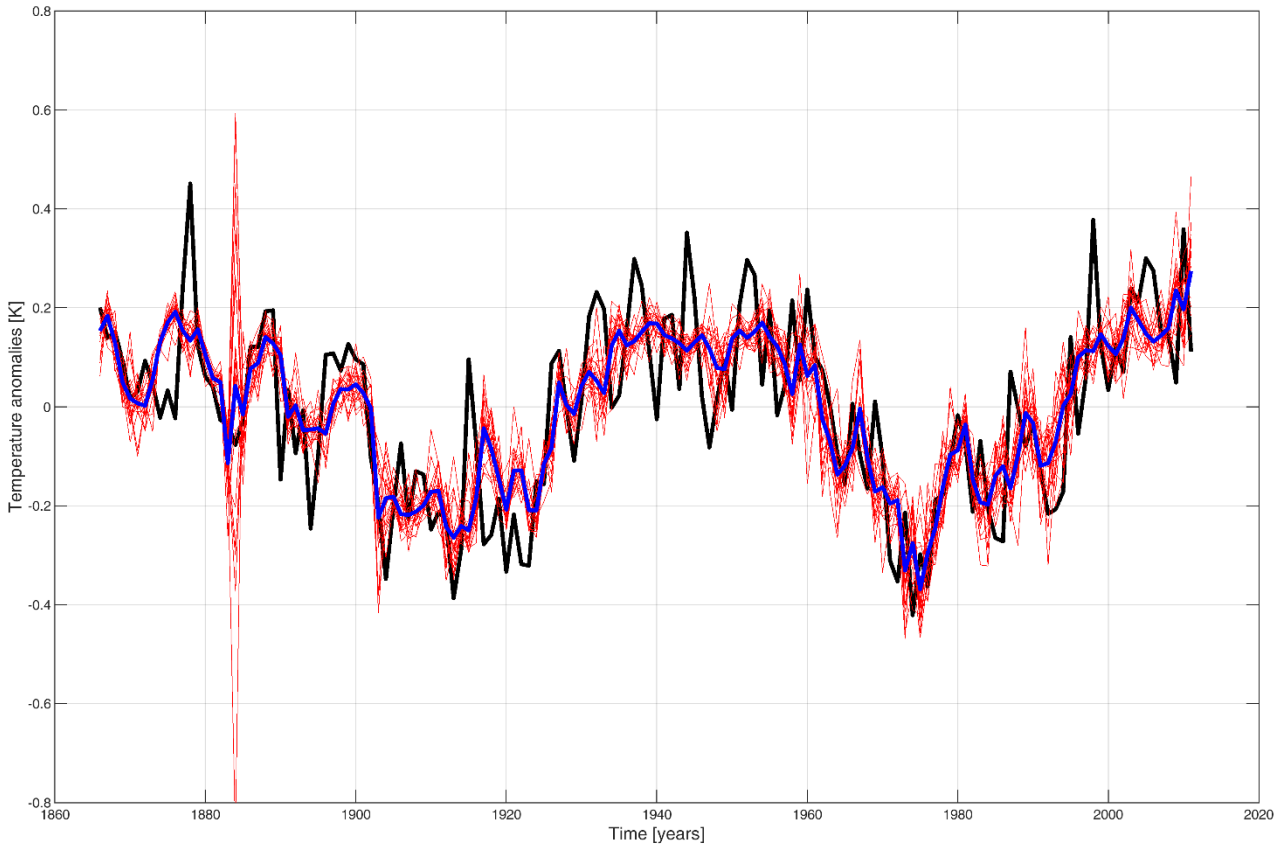


**Figure 31 Reconstruction of the observed AMO behaviour. In black the observed AMO, in red the ensemble members, in blue the ensemble mean. The predictors are RFWARM, RFSOX, RFSOLAR and RFVOL.**

The second step of our work consists in the attribution tests. As done in Chapter 4 for the mean global temperature attribution, we first train NN models by using the observed values of the predictors and then we make projections by suppling the NN with each predictor kept at a fixed value. For RFWARM and RFSOX we kept their values to their pre-industrial ones, for RFSOLAR

we consider the RFSOLSTAT of the Chapter 4, while for RFVOL we have a flat series that correspond to no-volcanic activity at all. In Figure 32 the case of RFSOX constant is shown (only the ensemble mean). It is a very impressive result. If the human cooling activities were not present during last 150 years, the AMO index would seem to show an almost monotonical rising trend, with the characteristically oscillating behaviour that disappears. Obviously, this result must be confirmed by future studies, but at this state of the art it appears very interesting. Natural variability of AMO seems not to be really natural. Going further with our work of attribution, in Figure 33 we show the results with RFWARM kept fixed. Also in this case appears in a surprising way that AMO would be a slightly decreasing trend if RFWARM gases were keep fixed at their pre-industrial values. The warming and cooling effects are confirmed by our tests.

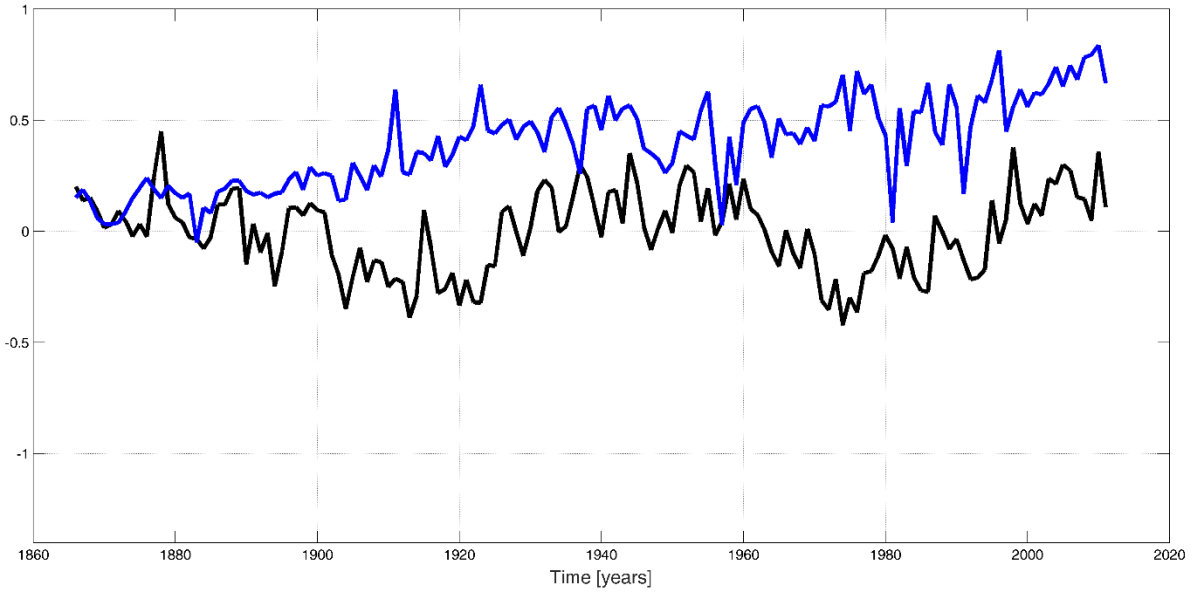


**Figure 32 Result of the attribution test with RFSOX kept fixed at their preindustrial values. In black the observed AMO, in blue the ensemble mean of the attribution test.**

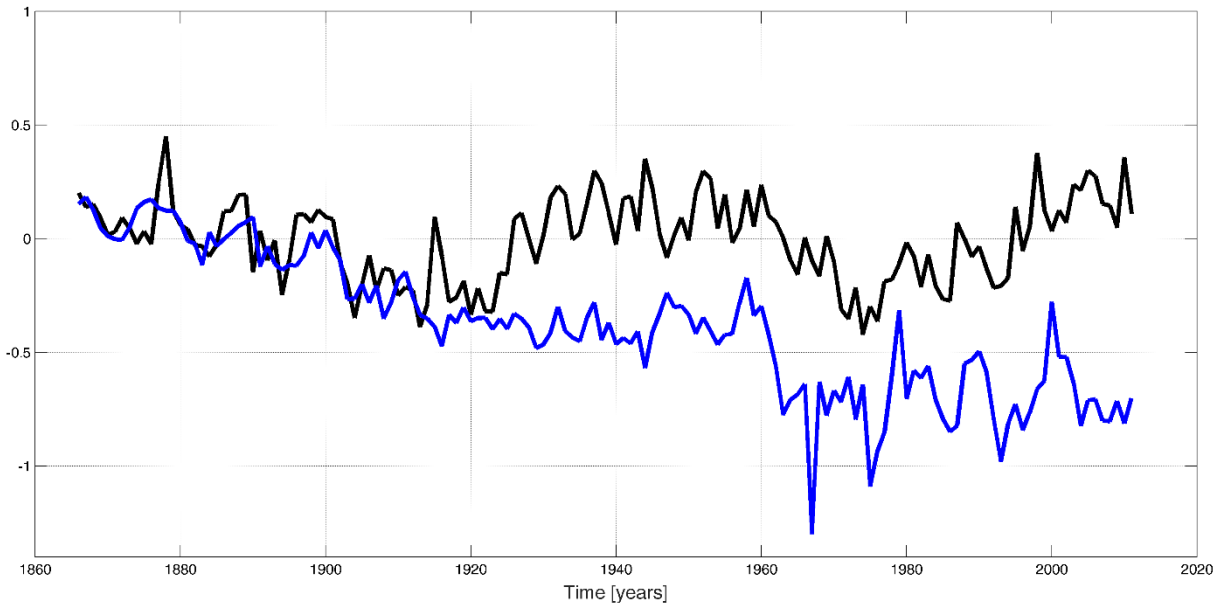


**Figure 33 Result of the attribution test with RFWARM kept fixed at their preindustrial values. In black the observed AMO, in blue the ensemble mean of the attribution test.**

Here we do not show the results of the attribution tests with RFSOLSTAT and RFVOL, because they do not appear significant. Natural variability, at least in terms of solar and volcanic activities, seems not to have a strong role for the reconstruction of the AMO observed behaviour.

## 5.3 Future scenarios

We have tried to perform some forecast experiments with our NN tool. As said in section 5.1, we make use of RCP scenarios. For doing forecasts of AMO up to 2100, we use the observed values of our predictors (available for RCP up to 2006) for the training of NN models and the future projections (from 2007 to 2100) as test data, to forecast AMO. As concern solar activity a constant power regime is considered, while for volcanoes we consider no activity at all. In doing so, we take a precaution. In the phase of the predictors normalization before the NN training, we normalize our data (both the observed and the future RCP data) as a single dataset. This experiment should be seen as a pioneering attempt of AMO forecast, for both the subject and for the method of forecast.

In Figure 34 and 35 we show the results related to the projections with RCP 2.6 and RCP 8.5, they are the extreme scenarios of the IPCC. The first one is the more optimistic, while the second one is the worst (the so-called business as usual). Both figures are consistent with the scenarios. We can note that in both cases the oscillating behaviour of AMO disappears. While for RCP 2.6 we have a quite constant behaviour, for RCP 8.5 we would have a strong rising trend.

We conclude this section by underling the interesting result obtained for AMO future projections – in particular the disappearance of the oscillating behaviour. However, we want to underline that it is a preliminary attempt and most research must be done to obtain more robust results.

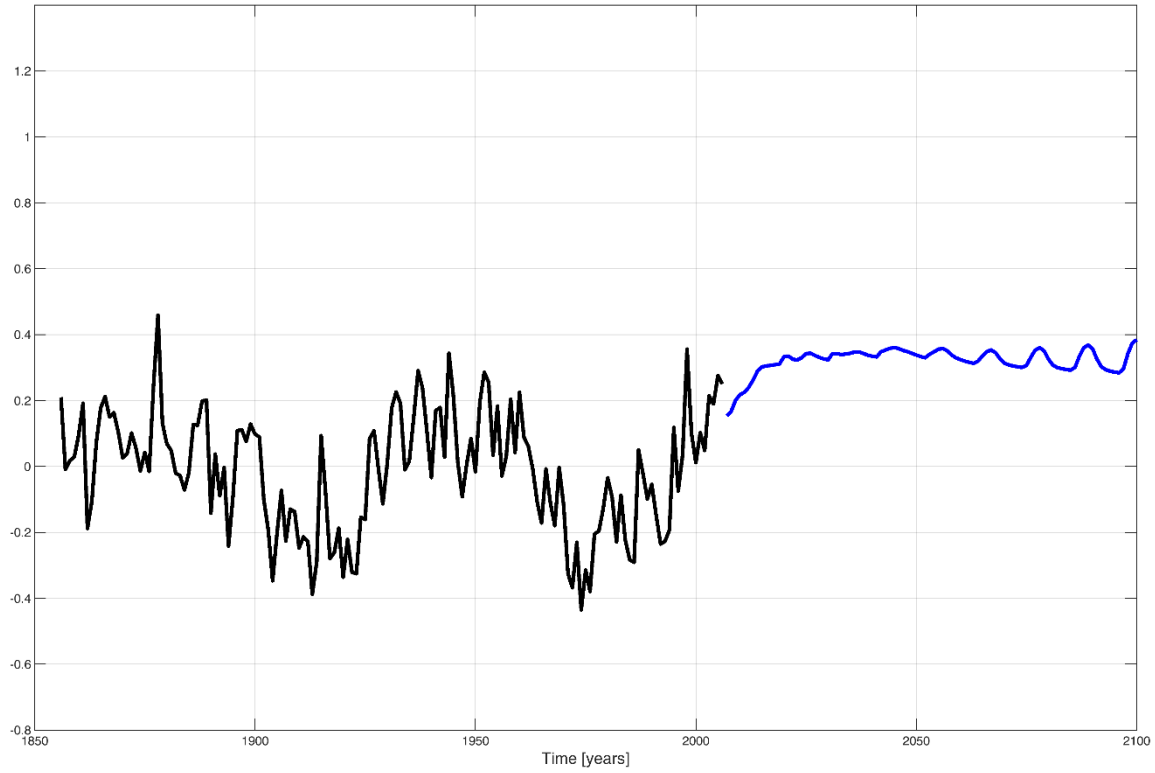


**Figure 34 Future projection for AMO on the basis of RCP 2.6 scenario. In black the observed values, in blue the forecast.**

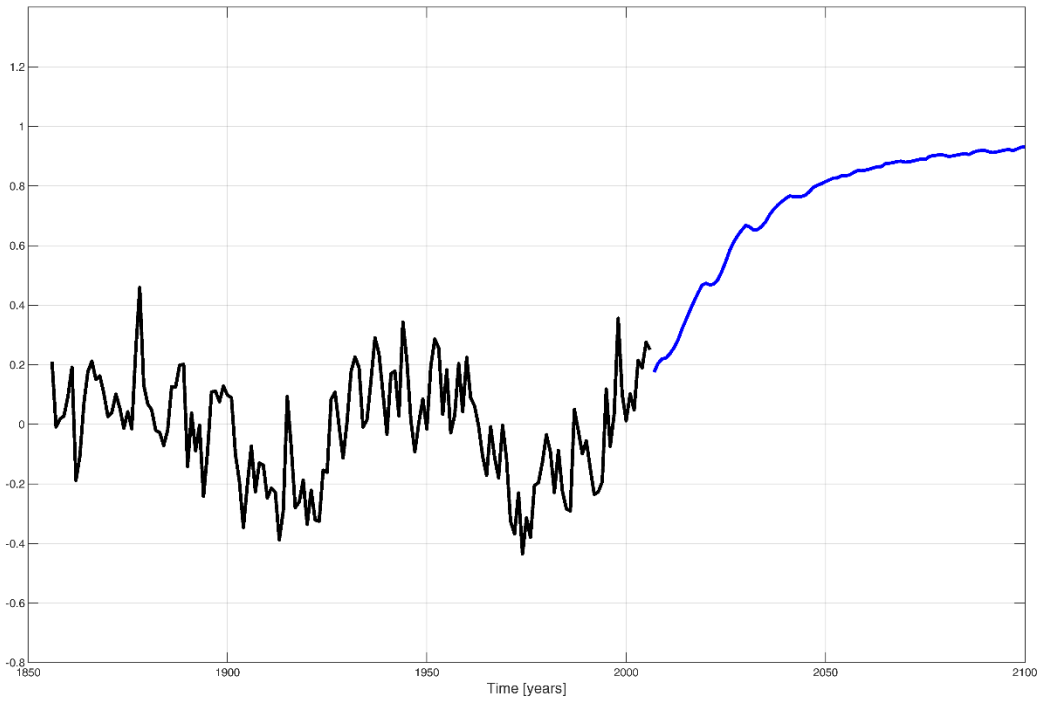


**Figure 35 Future projection for AMO on the basis of RCP 2.6 scenario. In black the observed values, in blue the forecast.**

## 5.4 Conclusions

The debate about the behaviour of the Atlantic Multidecadal Oscillation index is a recent and lively field (see again the references [85] to [94]). As each open debate, we have not yet a clear idea, but with our NN model we supply an independent mean of investigation in comparison to the canonical GCMs.

The main result of this chapter concerns the capabilities of our model in reconstructing the observed AMO trend, with an explained variance of about 60%, the new version of our NN model with the quasi-Newtonian algorithm works better than the previous attempt [16]. Going further with the attribution activity, we find the interesting results about the role of RFWARM – the warming GHGs – and of sulphates. The latter would seem to have a primary driving role for the AMO behaviour, a result that is supported in several GCMs driven experiments. The novel approach here furnished can be considered as a strengthening of the existing results. We again underline the bad performance of the MLR, which means that for the proposed predictors and AMO there are strong nonlinear relationships. In this chapter we also perform a first attempt (at our knowledge) of forecast for AMO. Using the RCP scenarios, we made future projections for AMO up to 2100. At

this state of the art this should be considered as an experimental attempt, but the differences between the different AMO projections seem consistent with the RCP scenario.

Further investigation about the real nature of AMO is necessary. It is important to discern about this topic because it is always being considered as natural variability. If it were due to anthropogenic activities it would be a very impressive result, and also the period of hiatus for global mean temperature – today attributed to the AMO – should be seen in a different way.

# 6. Analysis of meteo-climatic influences on migrations flows in Italy

The possible applications of our NN model is not only related to “pure” meteorological or climatological topics, but also may cover the impacts related to weather and/or climate. The potentiality of NN models in such kind of problems is highlighted in this and in the next chapters.

The influence of climatic changes on migration flows around the world is a topic widely debated in the scientific literature, but investigations are limited to restricted regions and often are more qualitative than quantitative. Nowadays, forced migrations is a critical global challenge ahead of us, in particular is a really hot topic for the European community. Italy can be considered as a natural “bridge” from Africa to Europe, so any advantage in understanding the causes of this phenomenon can help the implementation of policies in all the countries involved in similar phenomena.

Before showing the results of our work, we introduce the topic by a broader point of view. Many causes can be recognized as drivers for the observed migration flows. Obviously the political and socio-economic situation plays a fundamental role, so civil wars, other kinds of conflicts, terrorism, troubles in maintaining subsistence agriculture, loss of yields, etc, may have a strong role in triggering migration flows. In this predominantly sociological framework, environmental causes only recently start to be considered as possible triggering factors [95], [96]. However, evidences for a clear role of climatic changes in triggering and/or amplifying conflicts and/or migrations are appeared in the scientific literature, only in recent years [97], [98], [99], [100], [101], [102], [103], [104] and [105]. Furthermore in [106] we can see how taking climatic changes into consideration brings to a better understanding of forced migrations. Sometimes the climatic role can be quite clear, for example as an initial cause of tension leading to conflict as seen in the Syrian drought that preceded the Syrian crisis (see again [100]). In other cases, the process in more indirect, such as the contribution to land degradation by climatic factors in all the Sahelian belt, and more specifically in the Chad lake region [99]. To resume, even in a context of a complex mixing of causes, many evidences lead us to seriously consider climate change as an important factor in triggering or amplifying migrations. However, some disputes exist, see for example ref. [107] in which a bias in the investigations on direct link between climate and conflicts is showed, where conflicts can be considered the harbingers of migrations.

Furthermore, we want to underline that the phenomenon of migrations toward foreign countries is often the last step of the displacement of a population, for example after internal migrations. Besides, migrations may also be driven by specific economic factors, such as the availability of enough capital to migrate. Unfortunately, despite to these issues, data on migration flows are available only for flows towards the so-called developed countries. So, also in our research we focus on this kind of data.

We analyse data concerning migrations from the Sahelian belt towards Italy relatively to the 15 years (period 1995-2009) before the Syrian crisis and the so-called Arab Spring. This choice is due to the attempt to avoid the major socio-political causes that could overwhelm the direct role of climate. However, a big local crisis was present in the Sahelian countries in the considered period – i.e. the Darfur conflict – that could be a cause of migrations. So, we have also analysed quantitatively the influence on our results by the migrations from Sudan and Chad (the latter was indirectly affected by this conflict).

We apply simple multi linear regression together our NN models with the ensemble strategy previously described. So, in this Chapter we show the results of our research [108] that follows a double step. First of all, we investigate climatic influences on harvest yields in the ten countries of Sahel (Senegal, Gambia, Mauritania, Mali, Burkina Faso, Niger, Nigeria, Chad, Sudan, Eritrea) and then the impact of climatic factors (see details in section 6.1), yields and gross domestic product (GDP) on migration flows from these countries to Italy. At this stage, despite some important explanatory variables are still missing (for example those coming from social science studies, often only qualitative or however uncomplete), with NNs we may investigate linear and nonlinear influences of the above listed predictors on the complex phenomenon of migration flows from Sahel towards Italy. This complexity will be disentangled, at least in our case study.

## 6.1 Data and methods

As far as the migration flows are concerned, we consider data for bilateral – i.e. from the single sahelian country to Italy – flow of people. In detail, we make use of a subset of a more general dataset contained in a work by Cai et al. [101]. We consider, as said, the period 1995-2009, but in some cases the first year is not available, furthermore for Eritrea data for 1995 and 1996 are not available. In any case, the consideration of a subset permits to have a lower percentage of missing data in comparison to the original one [101]. In particular, we will use the *migration rate*, defined for each year as the number of migrants divided by the population of the origin country. We

underline that one is added to all migration flows before the calculation of the rate, in order to avoid issues related with zero-flows.

Data concerning yields are freely available from the World Bank (http://databank.worldbank.org). Yields are measured as kilograms per hectare of harvested land and include wheat, rice, maize, barley, oats, rye, millet, sorghum, buckwheat and mixed grains. As we see in the next section, these data play a double role in our NN models – as predictand in the first step and as predictor in the second one.

Data concerning the gross domestic product – to be precise, the purchasing power parity converted GDP per capita at 2005 constant prices – come from the Penn World Table 7.0 (see ref. [109]).

The considered meteorological data for the sahelian countries are mean annual temperature, total cumulated rainfall and the number of hours with temperature higher than 30°C during the growing season, for the considered period. Temperature and precipitation data are collected from the NASA Modern Era Retrospective Analysis for Research and Applications [110]. These data were originally gridded, so they are aggregated at the single country level and population weighted. Data for the growing season exposure above 30°C are calculated from gridded hourly temperature with the help of the Crop Calendar Dataset (https://nelson.wisc.edu/sage/data-and-models/crop-calendar-dataset/index.php) to determine the growing season for each considered cell, for details see again [101]. The later considered climatic parameter is very important because the growing season is the most delicate phase in the life cycle of cereal harvest.

As mentioned in the introduction of this Chapter, we apply multilinear regression (hereafter MLR) together with our NN model (with an ensemble strategy of 20 members for both). For the regression we make use of the same data used in the training and validation set for NN, so we give an "advantage" (more data for the calculation of the coefficients) to the regression. The NN model was deeply described in section 3.5. In the specific problem here faced, we have about 140 pairs inputs-target for the ten countries, so the application of a model specifically developed for a small dataset appears reasonable.

The NN model (also the linear regression) give as output a "forecast" for the above defined migration rate relative to the considered inputs-target pair. It is achieved with a model obtained using 10% of data for validation and the remaining for training. The Levenberg-Marquardt is the optimization algorithm adopted.

## 6.2 Analysis of linear and nonlinear influences

Our study consists in two steps. In the first one we try to reproduce the observed trend for the cereal yields in the Sahelian countries using climatic factors as predictors. Then, in the second step, we face the problem of migrations.

We perform both MLR and NN analysis and compare the concerning results. We perform also an interesting pruning activity. After the main test, we exclude an input in turn, in this way we can test which input has a dominant role in the climate-yields relation, furthermore we are also able to establish if we are in presence of linear or nonlinear relationships between the analysed variables (we better see this point in the next). The pruning strategy is applied to both the above-mentioned steps.

As far as the first step is concerned, we consider as predictors the mean annual temperature, the annual total precipitation and the number of hours in which the plants are exposed to temperature above 30°C (# of hours with T>30°C) during the growing season. We try to reproduce the observed variance for the cereal yields. This is done with a NN provided with a 3-4-1 topology. The choice of a low number of hidden neurons is consistent with the usual (empirical) strategy to avoid overfitting maintaining good generalization capabilities.

In Figure 36 we show the results of the NN models. Here the red lines represent the single ensemble runs, the blue curve is the ensemble mean and with the black curve the observed yields are shown. We can see an overestimation for Eritrea and Niger, while an underestimation is present for Burkina Faso. The best performance appears to be achieved for Nigeria. In Table 6 quantitative results (in term of R-squared) are presented for the main test and the pruning activity for both NNs and MLR. In the first row we have the results for our main test. The nonlinear method seems to perform better than the MLR, in fact with the NNs we are able to explain approximately half of the observed variance in the data ($R^2$=0.502). The role of each predictor can be discerned by the above-mentioned pruning activity: we perform the NN and MLR runs extracting one input at a time and the performance associated will give us useful indications about the role of the extracted input itself in reproducing the observed harvests.

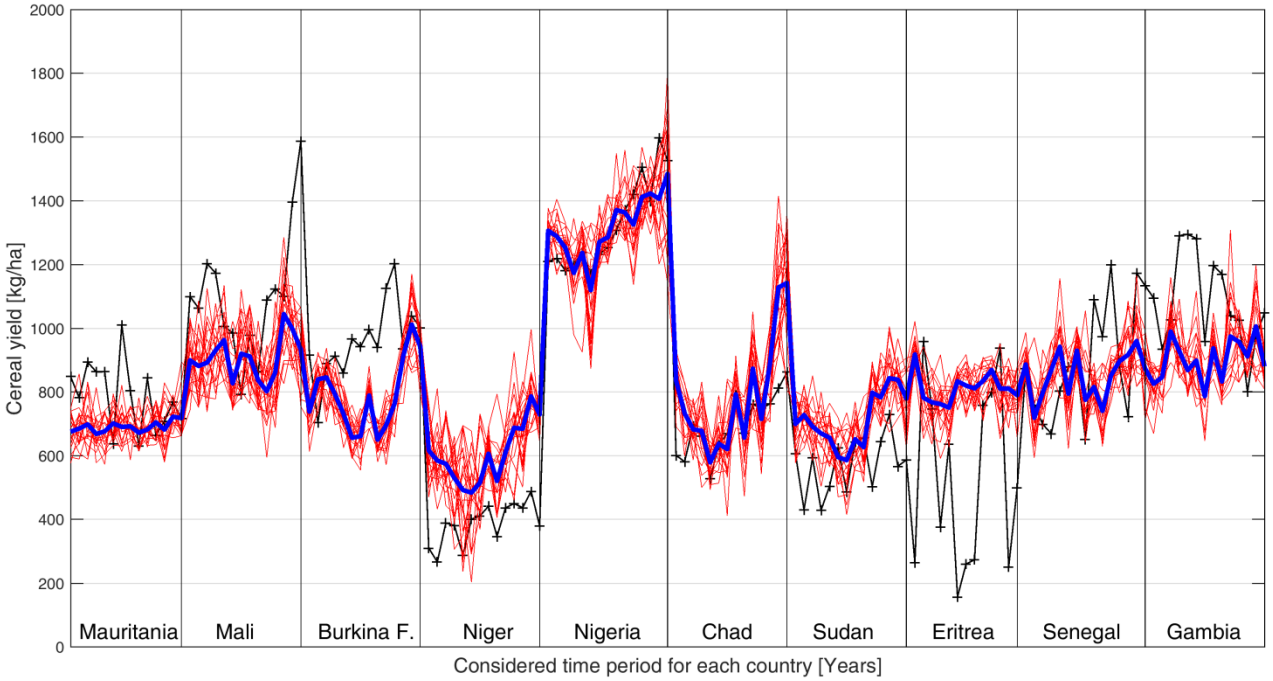


**Figure 36 Cereal yields result for the NNs ensemble. In black the observed yields [kg/ha]; red curves are the ensemble members of the model; blue curve is the ensemble mean.**

| Inputs | Target | NN R-squared | MLR R-squared |
|---|---|---|---|
| Prec-Temp-# hours T>30°C | Yield | 0.502 | 0.405 |
| Prec-Temp | Yield | 0.406 | 0.406 |
| Prec-# hours T>30°C | Yield | 0.516 | 0.407 |
| Temp-# hours T>30°C | Yield | 0.386 | 0.216 |

**Table 6 Performance (R-squared) of yield estimation by NNs and MLR. Results for the main test (first row) and the pruning activity are presented.**

The results related to the pruning activity are also shown in Table 6. Precipitation and the # of hours with T>30°C seem the most influent parameters for the cereal yields. In fact, when these variables are not considered we note a decrease in performance for NN models. While when we extract the annual mean temperature R-squared is quite similar. Another interesting feature that may be seen by the pruning activity is the nonlinear role of the # of hours with T>30°C for the estimation of the yields. In fact, when we extract this predictor (second line in Table 4), we have a decrease in R-squared only for NN models and consequently the performance of NNs and MLR

become now similar. This indicates the possible existence of a nonlinear threshold effect of heat waves on cereals during the growing season. The value of R-squared (about 50% of the variance explained by our NNs) shows that yields are probably driven also by external factors, such as the availability of fertilizers, the availability of water for irrigation or the phenomenon of land abandonment due to external causes.

To face the second step of our research, we must consider that the abundance of yields can be itself a driver for migration flows. Furthermore, in the following we will also consider the role of the Gross Domestic Product per capita as a possible influencing factor.

In doing the second step, we now show the results obtained using NN models with a 4-4-1 topology. In fact, now we have four predictors: the climatic factors already considered and the observed yields data. The target (or predictand) is the natural logarithm of the migration rate from the ten Sahelian countries to Italy.

The results of our NN models are shown in Figure 37. Despite of a clear overestimation of the migration flow from Chad, our results appear quite good. In fact, if we see the first row of Table 7, we see that our NN model shows a high level of explained variance in the observed migration flow ($R^2$=0.775). If we consider that, at this first level of investigation, we are not yet able to consider socio-economic data (for the lack of enough accurate dataset), this result appears quite impressive.

A different measure of performance can be considered, too. We calculate the *Receiver Operating Characteristic* (ROC) diagram, a method generally used for model performance assessment (see [6] for a deeper description of the method). In its simplest form it is a parametric plot of the hit rate (the probability of a good detection or forecast) versus the false alarm rate, as a decision threshold is varied on the range of the estimated quantity (in our case the logarithm of migration rate). A diagonal line means a random estimation and the amount of concavity is a qualitative measure of performance. The area under the curve gives a quantitative (scalar) measure of performance. With an area of 0.5 we have the above-mentioned random estimation, a ROC area=1 gives a perfect estimation.

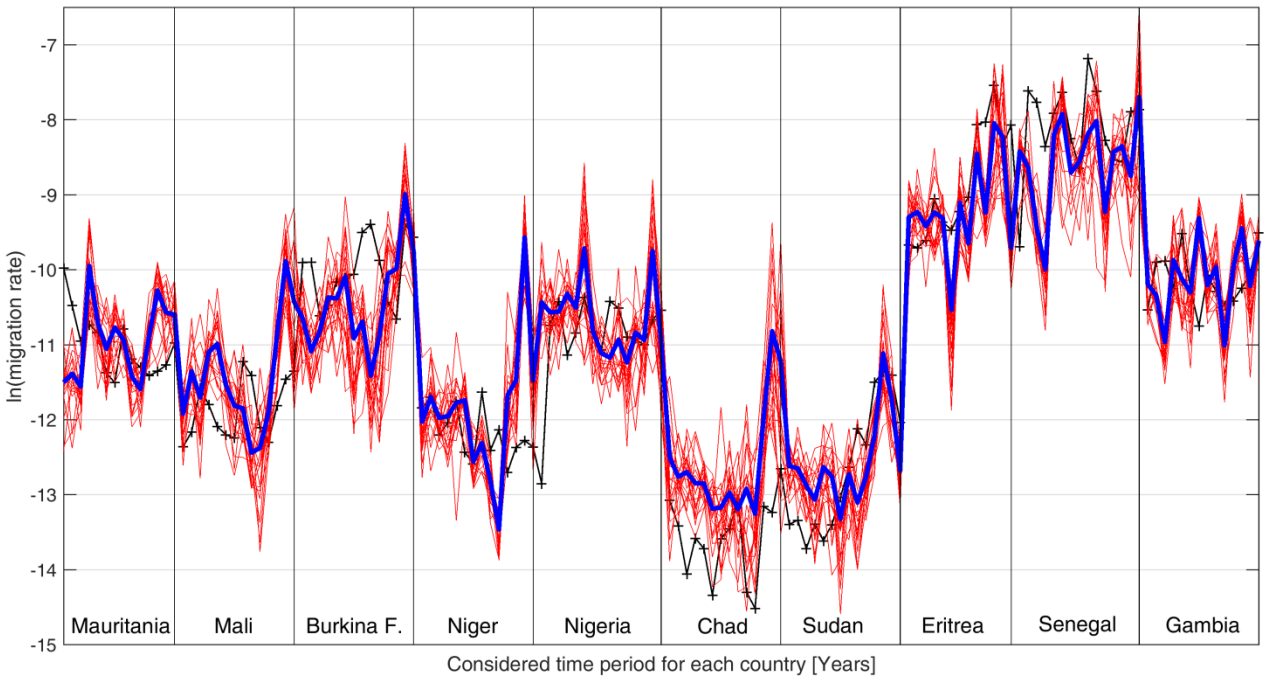


**Figure 37 Neural Network models result of the observed migration rates (black curve) from the Sahelian countries to Italy. In red the ensemble members, in blue the ensemble mean.**

| Inputs | Target | R-squared – NN | R-squared – MLR |
|---|---|---|---|
| Prec-Temp-# hours T>30°C-yields | ln(MigRate) | 0.775 | 0.626 |
| Prec-Temp-# hours T>30°C | ln(MigRate) | 0.671 | 0.611 |
| Prec-Temp-yields | ln(MigRate) | 0.683 | 0.632 |
| Prec-# hours T>30°C-yields | ln(MigRate) | 0.361 | 0.085 |
| Temp-# hours T>30°C-yields | ln(MigRate) | 0.715 | 0.447 |

**Table 7 Performance, in terms of R-squared, of migration rates modeled by NN and MLR models.**

In Figure 38 the ROC approach is applied to the main test of the migration rate estimation (Table 7, first row). On the left we have the NN results and on the right the MLR performance. Our results clearly show an overall better behaviour of the nonlinear approach.

As done for the simulations of yields, we perform a pruning activity (shown in Table 7) from which we can obtain information about the relative importance of each predictor and we can

understand the linear or nonlinear role of each predictor itself. It is clear, from the decay of R-squared (Table 7, fourth row), that temperature is the most influential variable on migration rate. This decrease in performance is less marked when we extract the other variables. Furthermore, we may see that yields and the number of hours with T>30°C have a clear nonlinear role. In fact, in both cases, when we extract these predictors the performance of NNs decreases while that of MLR remains similar.

These findings support the idea that a nonlinear variation in yields and heat waves could have an influence in migration flows. In any case, globally the dominant role on migrations appears due to the annual mean temperature, as R-squared drops for both NN and MLR when temperature is extracted from the inputs (fourth row in Table 7).

The last finding could be seen as a confirmation of previous studies which show that the observed increase of mean temperature and of the occurrence of heat waves directly influence humans (and also animals), by exceeding the thresholds of thermal tolerance, in particular for low-income countries as the Sahelian – see [111] and [112]. Obviously, deeper investigation is necessary with more detailed datasets, but we are confident that this could be the way for bridging the existing gap of knowledge between social and scientific fields.

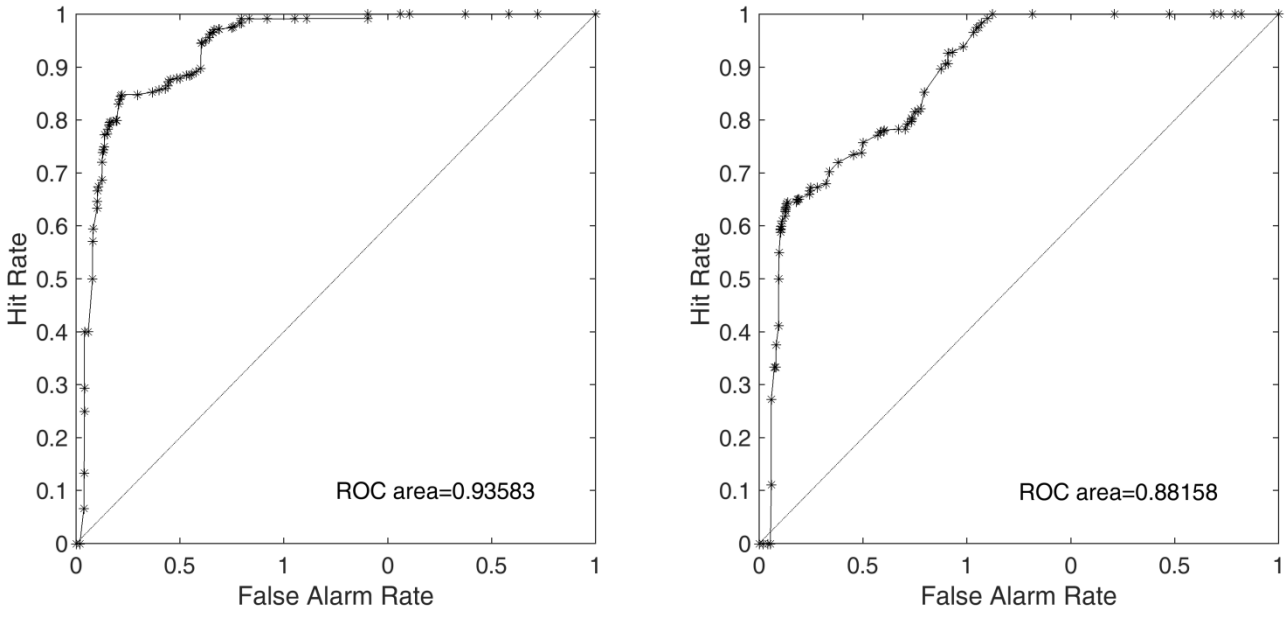


**Figure 38 ROC curve and the related value of the ROC area for the NN model (left) and the MLR model (right). These performances are referred to the main test for the estimation of migration flows (Table 7, first row)**

At the beginning of this chapter, we introduced the theme of the Darfur region conflict as a possible influence factor for migration flows from Sudan and the nearest Chad, the latter strongly involved in migration pressure from Darfur during the conflict itself. Unfortunately, data about the number of deaths, or quantitative values for migrations between these two countries have a low reliability or are totally missing. So, we have performed an analysis about the possible influence of the conflict on our main results, by excluding data after 2002 for Chad and Sudan (the conflict began in 2003). Following this approach, we only see slight changes in the estimation of migration flows for all the considered countries before 2003. Globally we may conclude that the Darfur conflict seems not to have a strong influence on our main results about the importance of climatic factors.

Migration towards foreign countries is influenced also by the availability of enough money to finance such a long and dangerous journey. The lack of more proper data, for example about wealth distribution in the Sahelian countries, brought us to consider a rough proxy for this, the Gross Domestic Product. In detail, GDP was added as a predictor in the 'main test', by using a 5-4-1 NN in order to predict the migration rate. The performance is unchanged, with a value $R^2$=0.762 for NN and $R^2$=0.632 for MLR (compare these results with the first row of Table 7). Because in the Sahelian countries the GDP is almost totally due to agricultural activities, yields and GPD show a similar temporal behaviour. So, considering both of them as predictors adds no new information.

A further test has been performed by replacing GDP for yields in a 4-4-1 NN structure, the related results are $R^2$=0.721 (NN) and $R^2$=0.619 (MLR). It would seem that the migration flows are more related to the agriculture than to the GDP. In an agriculture-based economy as that of the Sahelian countries it seems reasonable.

## 6.3 Conclusions and future perspectives

In this Chapter we have shown the climatic influences on harvest yields in Sahelian countries and on migration flows from these countries to Italy.

The main result is the high variance that may be explained by the NN model ($R^2$=0.775) thanks to climate variables and yields. Considering that other variables – not considered here for the lack of data – are involved in this phenomenon, the result appears very impressive. Our investigation has another peculiarity. The comparison of the performance of our NN model vs. that of a simpler MLR together with the use of a pruning strategy, allows us to achieve several interesting results. In short, agriculture (by harvest yields) may be considered as a link between

climate and migrations, which can improve the latter. Also, the nonlinear influence of heat waves during the growing season has been discussed. Overall, the annual temperature appears to be the dominant factor that influences migrations. So, we may suppose that the overcome of a threshold of thermal tolerance can have a role in such dynamics. It is highly probable that temperature acts in an indirect way also through other mechanisms, but we are limited by our dataset. Finally, we also shown that GDP seems to be not important: it is understandable for the agriculture-based economies of the Sahelian countries.

This is probably the first attempt done to investigate climate-induced migrations in a quantitative way with a non-linear model. We must keep in mind that forced migrations are a complex phenomenon influenced by a myriad of socio-economic factors. Conflicts, poverty, bad sanitary conditions, political instabilities could all be driving factors, but they are all difficult to be quantitatively measured. So, at this stage we are not able to insert socio/economic influences for the lack of the data. In any case, as the lack of agricultural livelihood could be driven by meteorological factors, these last could also bring to political instabilities and conflicts and again to forced migrations [104]. So, we are hopefully that this work can stimulate the interdisciplinary study necessary to improve these interesting relationships, both with similar statistical models that with other methods.

# 7. Meteorological influences on Leishmaniasis spread – a case study over central Italy

The range of applicability of our NN model is very wide. As seen, thanks to our NN tool we can analyse both pure physics of the atmosphere issues and the related impacts. So, in this Chapter we show another interesting application related to the impacts of weather/climate. We discuss the possible influence of meteorological parameters on the observed number of a type of mosquitoes – i.e. the *Sergentomyia minuta* (S.minuta) – that in recent years has been revealed as a driver in the circulation of mammalian leishmaniasis. Recently the detection of this vector – and in the same way that of similar vectors – showed a geographical expansion toward areas previously considered immune. This is a consequence of ecological modifications due to several environmental causes, including meteo-climatic changes. The insect vector sampling is not an easy task. Among the available tools the so-called CDC-light trap [113] is one of the most used. However, as every other method, the number of detections is generally very low, close to the lower threshold detection limit of the instrument. So, to produce reliable information with this kind of data, there is the need for complementary approaches based on data modelling able to explore small data set.

In this framework, we analyse data from trap collections of S. Minuta collected at a specific site in central Italy, and we try to reproduce the observed behaviour from the influence of meteo-climatic factors. As usual, we perform analysis by MLR and NN models. Furthermore, probably for the first time, we can investigate also linear and nonlinear influences of meteo-climatic factors on the abundance of these sandflies. Ultimately, this kind of work is helpful to obtain information about disease transmission risk in a defined area.

## 7.1 Data collection and method

The collection of data has been performed by the Department of Public Health and Infectious Diseases of the Sapienza University of Rome. Sandflies are collected weekly in 2014, 2015 and 2016, approximately on the period July-September. Five CDC light traps are used in a rural area of Monte Porzio Catone (Latium, Central Italy), a site at 30km from Rome. These traps were placed in gardens of farms and private houses, in proximity of animals (preferably dogs). The identification of the species was performed both morphologically [114] and molecularly with the

protocol described in [115]. A total of 609 sandflies were collected during the study, the larger percentage belongs to *Sergentomyia minuta* (97.6%), smaller percentage for *Phlebotomus perniciosus* (1.2%) and *Phlebotomus perfiliewi* (1.2%).

The meteorological data were collected from the database of the Regional Agency for Development and Innovation in Agriculture of the Latium region (ARSIAL). Data of the station of Monte Porzio Catone are used: the location of the station is closed to the trapping site. This weather station collects data about 2-meter air temperature (mean, maximum and minimum), underground temperature at 10 cm and 30 cm, cumulated precipitation, relative humidity (mean, maximum and minimum), total solar radiation, mean pressure, wind (mean direction and speed). On the basis of these data, we calculated also the pseudo-evapotranspiration index (PET) and the standardized precipitation and evapotranspiration index (SPEI). Useful to our analysis are also the characteristics of the sampled summers, they are obtained by means synoptic analysis built on data from the Earth System Research Laboratory (ESRL) of the National Oceanic and Atmospheric Administration (NOAA), freely available at https://www.esrl.noaa.gov/psd/data/composites/day/.

As for the migration topic, we adopt both NN and MLR models. For the Sergentomya, we have a series of inputs-target pairs: each is used as test set in the leave-one-out cross validation. The detailed description of the NN model can be found in section 3.5. We have used the Levenberg-Marquardt algorithm for the training of our NN model. In the following section we see further details about the number and the type of the predictors used in our tests.

## 7.2 Application of NN model and discussion of results

As specified at the beginning of this chapter, the aim of this study is to investigate the role of weather and climate in explaining the observed trend of the data of sandflies abundance and population dynamics in the fixed study area. This method of analysis may be applied to other locations and/or other types of vectors.

First, we note that during the three sampling seasons different climatic conditions occurred (Figure 39). The precipitation patterns on the Latium region, and on our site, are very similar in the considered periods, while strong inter-annual differences in the thermal field are present.

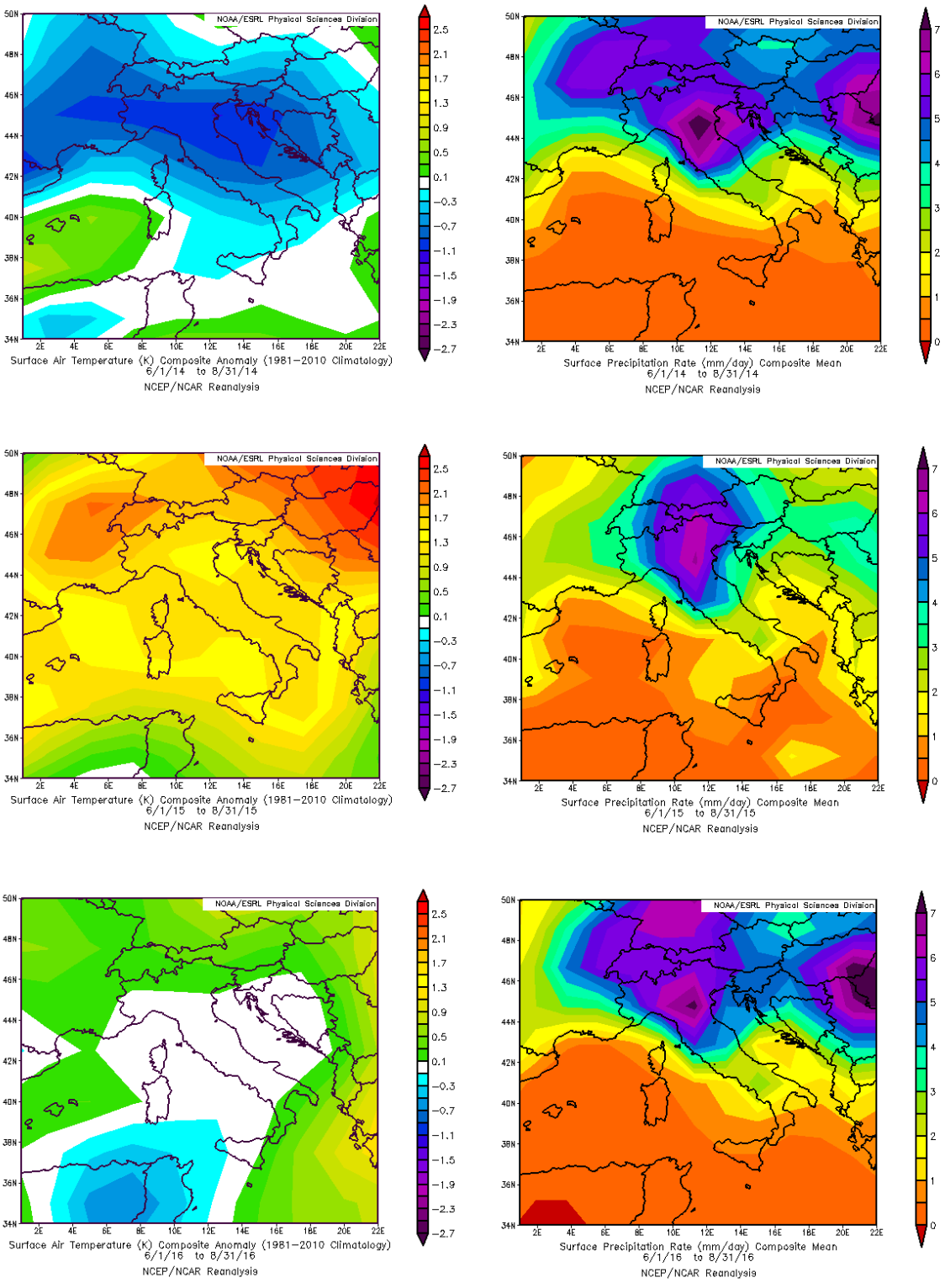


**Figure 39 2-meter surface temperature anomalies (left column) and surface precipitation rate (right column), for the three years considered in this study: 2014 (top); 2015 (center); 2016 (bottom).**

In detail, in reference to the NOAA climatology 1981-2010, 2014 has been characterized by a colder than normal summer season, 2015 by a very hot summer and 2016 by a normal one. In climatological terms, 2014 can be seen as a typical summer of the past climate and 2015 as an anticipation of future climate due to global warming effect.

In general terms, these seasonal climatic conditions may be related to the number of sandflies, but to reconstruct the time series of weekly captures we must consider variables (predictors) more strictly related to the observed fluctuations (predictand). For the latter we consider the number of captures for trap. Meteorological data of the weather station of Monte Porzio Catone are considered in many combinations as inputs of our NNs: here only the more performing cases are showed.

In our tests we found that the abundance of *S.minuta* is strictly related to the mean weather conditions during the oviposition, larval and adult stages. As a matter of fact, the best results in the estimation of the number of captures per trap are obtained for meteorological data averaged on the 45 days before the captures themselves. An explanation of this relation is not well established because data for larval status are not easy to collect. In any case, this range of 45 days here found is in line with literature data about larval development of sandflies in general, which shows a range of three to eleven weeks [116], [117].

The best result is achieved (Figure 40) if we consider as input for our models (NNs and MLR) the values of: 2m mean temperature; mean relative humidity, mean temperature at 10 cm under the ground. As said, all these values are averaged on the previous 45 days with respect to the trapping day. In obtaining this result we use a 3-3-1 NN structure, to avoid overfitting problems. The importance of the selected predictors in describing the population dynamics of similar sandflies species has been already underlined in literature [118], [119].

From Figure 40 we can see that the best performances are achieved with the NNs ensemble. For 2015, NNs are able to reconstruct the bimodal behaviour of the population density, while MLR does not catch at all this peculiarity. We note that also the periods with zero captures are better reconstructed by NN approach. In Table 8 the explained variance in terms of $R^2$ is reported. The ensemble mean of our NN models explains about 63% of the observed variance of the population density. The linear models perform in a poor way.

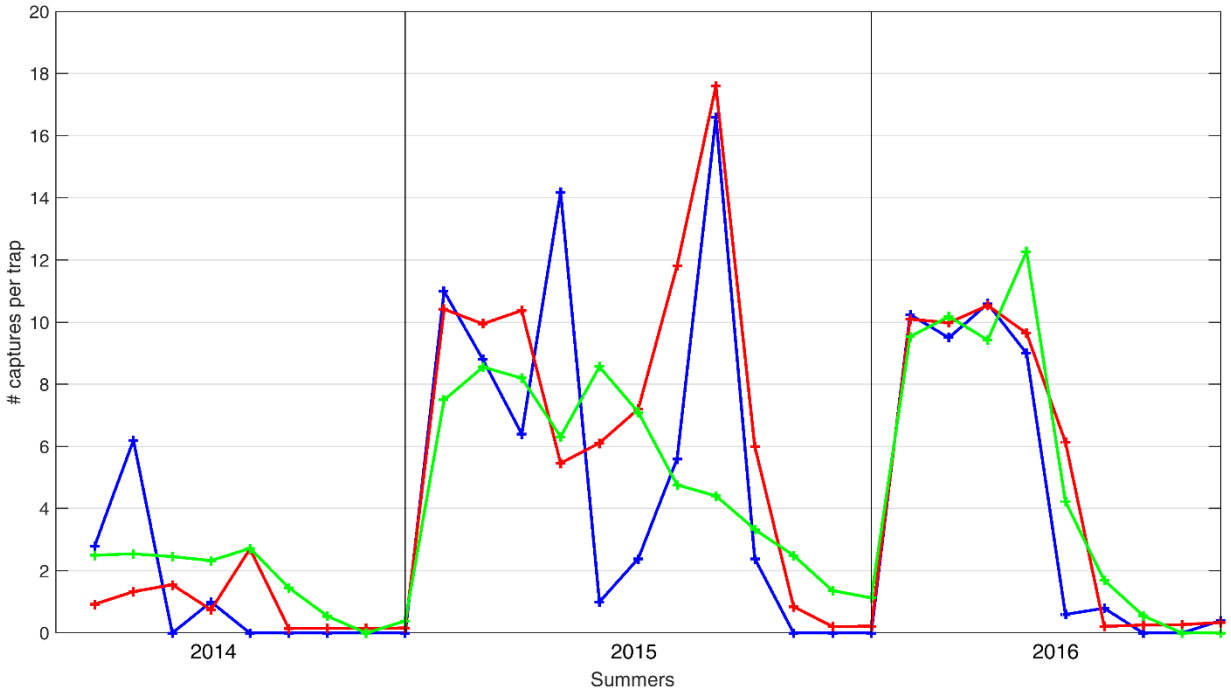


**Figure 40 Estimation of captures density (blue line) by the ensemble mean of our NN models (red line) and by MLR (green line). The predictors are mean temperature, relative humidity and underground temperature below 10 cm averaged in the previous 45 days the date of capture.**

| Predictors (averaged over previous 45 days) | Target | NN ($R^2$) | MLR ($R^2$) |
|---|---|---|---|
| T – humidity – T ground (10 cm) | captures density | 0.632 | 0.479 |
| T – humidity | captures density | 0.482 | 0.462 |
| T – T ground | captures density | 0.383 | 0.329 |
| humidity – T ground (10 cm) | captures density | 0.460 | 0.489 |

**Table 8 Performances (in terms of $R^2$) of captures density estimation by NNs and MLR.**

It is possible to discern the linear and non-linear role of the three predictors thanks to the usual pruning activity. We built models in which we extract a predictor in turn and then their performance is analysed. In Table 8 we see also the results of the pruning activity. When we exclude humidity as input, the performance of NN ensemble mean decreases. So, humidity seems to have a major role in leading captures density trend. The same happens in a lesser extent when air temperature is not considered as predictor and even at a lesser extent when underground temperature at 10 cm is extracted from the input layer.

Furthermore, there is evidence for a clear nonlinear role of T and T ground, because when we do not use one of these inputs, the performance of NNs is like that of the MLR (see again Table 8). This is a clue towards the possibility of nonlinear effects (for example a threshold effect) of heat waves on the population density of sandflies.

## 7.3 Conclusions

The NN model permits to assess the climatological influences on the population dynamics of *S. minuta*, the importance of relative humidity and temperature as major drivers affecting sandflies is also confirmed by our innovative approach. The NN model can explain a wide majority of the observed variance. This best result is achieved by considering temperature, relative humidity and underground temperature at 10 cm, averaged in the previous 45 days respect to the trapping data. This period matches with the larval and growing stage of the life cycle of *S.minuta*, in this sensitive phase of life the sandflies appear to be strongly influenced by environmental conditions so as temperature and relative humidity. The lack of suitable 'external' conditions can bring to difficulties in the growing of the sandflies from the larval to the adult stage, and our model is able to catch the major driver and also to make light about nonlinear influences.

Notable considerations are obtained thanks to the pruning activity, too. The fundamental role of relative humidity as driver is recognized. Besides, we can discern the possible nonlinear role of air temperature, probably due to threshold effects related to heat waves.

Obviously, sandflies dynamics are influenced also by non-climatic factors (the availability of moist area, presence of animals and or humans, and so on), so our results are quite impressive and very promising in view of future consideration of non-climatic factors and also for the investigation about the possibility of spread of diseases related to sandflies presence.

# 8. Conclusions

During last decades, atmospheric sciences acquired a rising interest by the scientific community. Research on this field and on related impacts is generally conducted by GCMs, so the idea behind this research thesis was to supply a different mean of investigation, fast, simple and efficient. Thanks to the NN model specifically developed during this research activity we have given a complementary – or an alternative – mean of investigation. We can consider the attribution results achieved by the GCMs with more confidence, as we obtain very similar results in a completely different way. In fact, the effective independence between the existing GCMs is a heated debate, the consideration of an independent mean of investigation is an instrument for toughen the results about climate change studies.

In the light of these motivations, here we have presented the neural network tool, specifically developed for small dataset. Using the MLR we have also given a constant reference to validate the effective usefulness of our method. Furthermore, as seen in Chapters 6 and 7, the use of MLR was also useful during the so-called pruning activity, thanks to which we can establish linear and/or nonlinear influences of predictors for each target in exam. Thanks to the four applications showed in this work, we have shown the power of this tool, starting from general atmospheric sciences topic as the attribution temperature theme up to the impacts related to climate change as the forced migrations.

In Chapter 4 we have presented the work about the global mean temperature recent behaviour. The NN has been able to reproduce the observed trend of temperature using both natural and anthropogenic forcings. Then, using this validated model, we have performed attribution tests. With the confirmation of a predominant role for GHGs in reproduce the recent global warming we have an impressive result. GHGs the responsible for the recent warming, and it is established with two completely independent means of investigation (GCMs and NN model). This give more robustness to the result itself. In the actual debate on this theme, GCMs are often addicted for the lack of independence (all of them have a common ancestor) so only the NN approach can be really considered as an independent model. Further results were achieved in the attribution tests. The role of the Sun appears important for the warming of the middle of $20^{th}$ century, while for last decades it does not matter. We also establish the role of natural variability in modulating long period

oscillation of temperature (as the hiatus in the ’70 of the 20$^{th}$ century) and in reproducing the interannual variability. Our results are made more robust by the consideration of different dataset (CMIP5, Hansen) which confirm our main discoveries.

Always in the field of climate research, in Chapter 5 we have presented an innovative way to investigate about the widely debated theme of the origin of the Atlantic Multidecadal Oscillation. Thanks to NN models we see that the natural variability of AMO seems to be not really natural. In fact, sulphates seem to have a key role in the recent evolution of AMO. So, we insert into the debate about the origin of the variability of AMO. This topic is obviously object of debate so, in the next years we can expect to apply our model to novel dataset and/or novel predictors.

In Chapters 6 and 7 we have shown the powerful of our model by applying it to impacts related to meteorological and climatological conditions. In Chapter 6 we have presented (for the first time at our knowledge) a quantitative analysis of forced migrations to Italy, a hot topic in the political agenda of most European countries. We focus the analysis to forced migrations from the Sahelian countries (the main basin of migrants) to Italy, due to its role of bridge for migrations from Africa. We found results quite impressive. In fact, by considering meteorological variables and yield harvest data as predictors, we can explain most of the variance present in the migration rate to Italy from the Sahelian countries. Keeping in mind that this kind of problem is influenced by socio-economic variables – whose data are often not available – our result is notable. Furthermore, with the pruning activity and thanks to the consideration of MLR models, we also discern about the linear or nonlinear role of the considered predictors, some threshold mechanism related to heat waves seems to exist. Finally, in Chapter 7 we follow a similar approach for another impact related to meteorological conditions. Using the NN model, we reproduce the observed population density of a vector responsible for the mammalian leishmaniasis, obtaining a good percentage of the variance of the available data. Again, the use of a pruning activity permits to discern about the linear and nonlinear role of the considered predictors. These findings are very promising since in recent years we have seen the spread of similar diseases due to the abnormal diffusion of the vectors responsible for the diseases themselves.

With this work, we have shown the wide field of application of our NN model. It permits to make both novel research and consolidation of pre-existing results. The latter thanks to the statistical origin of the model itself. In the era of data, the availability of a novel tool as our NN model, that is relatively fast, simple to apply, with capabilities of generalization, should be seen as a powerful mean in the wide sector of the atmospheric sciences. A lot of applications could be

implemented in the next years and, as we have seen, they could cover a wide range of items from pure climate issues up to the related impacts. Furthermore, the model can be also able to cover a wide spatial range, from global (Chapter 4) to local (Chapter 7).

# Bibliography

[1] Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.), 2013. *IPCC, 2013: Climate Change 2013: The Physical Science Basis*. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp.

[2] Edwards, P. N., 2011. History of climate modeling. *WIREs Clim Change*, 2: 128–139.

[3] Lorenz, E. N., 1963. Deterministic nonperiodic flow. *Journal of the atmospheric sciences*, 20(2), 130-141.

[4] Hegerl G.C., Zwiers F.W., 2011. Use of models in detection and attribution of climate change. *WIREs Climate Change*, 2:570-591.

[5] Mazzocchi, F. and Pasini, A., 2017. Climate model pluralism beyond dynamical ensembles. WIREs Climate Change. 8.6 (2017): e477.

[6] Wilks DS. 2011. *Statistical Methods in the Atmospheric Sciences*, 3rd edn. Academic Press: New York, NY.

[7] Haupt S.E., Pasini A., Marzban C., 2009. *Artificial Intelligence Methods in the Environmental Sciences*. Springer: Dordrecht.

[8] Granger C.W.J., 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37:424-438.

[9] Attanasio A., Pasini A., Triacca U., 2012. A contribution to attribution of recent global warming by out-of-sample Granger causality analysis. *Atmos Sci Lett*, 13:67–72.

[10] Attanasio A., Pasini A., Triacca U., 2013. Granger causality analyses for climatic attribution. *Atmos Clim Sci*, 3:515–522.

[11] Pasini A, Triacca U, Attanasio A., 2012. Evidence of recent causal decoupling between solar radiation and global temperature. *Environ Res Lett*, 7:034020

[12] Stern D.I., Kaufmann R.K., 2014. Anthropogenic and natural causes of climate change. *WIREs Climate Change*, 122:257-269

[13] Pasini A, Lorè M, Ameli F., 2006. Neural network modelling for the analysis of forcings/temperatures relationships at different scales in the climate system. *Ecol Model*, 191:58–67.

[14] Schönwiese C-D, Walter A, Brinckmann S., 2010. Statistical assessments of anthropogenic and natural global climate forcing: an update. *Meteorol Z*, 19:3–10.

[15] Verdes P.F., 2007. Global warming is driven by anthropogenic emissions: a time series analysis approach. *Phys Rev Lett*, 90:048501.

[16] Pasini, A., Racca, P., Amendola, S., Cartocci, G., and Cassardo, C., 2017. Attribution of recent temperature behaviour reassessed by a neural-network method. *Scientific reports*, 7(1), 17681.

[17] Pasini A., 2015. Artificial neural networks for small dataset analysis. *J Thorac Dis*, 7:953–960.

[18] PielkeSr., R. A. And R. L.Wilby, 2012. Regional climate downscaling: What's the point?, *Eos Trans. AGU*, 93(5), 52.

[19] Benestad RE, Hanssen-Bauer I, Chen D., 2018. *Empirical-Statistical Downscaling*. World Scientific Publishers: Singapore.

[20] Feser F., Rockel B., Von Storch H.,Winterfeldt J, Zahn M., 2011. Regional climate models add value to global model data – A review and selected examples. *Bulletin of American Meteorological Society*. 92(9), 1181-1192.

[21] Amendola, S., Maimone, F., Pasini, A., Ciciulla, F. and Pelino, V., 2017. A neural network ensemble downscaling system (SIBILLA) for seasonal forecasts over Italy: winter case studies. *Met. Apps*, 24: 157–166.

[22] Chadwick, R., Coppola, E., and Giorgi, F., 2011. An artificial neural network technique for downscaling GCM outputs to RCM spatial scale. *Nonlinear Processes in Geophysics*, 18(6).

[23] Pasini, A., Pelino, V., and Potestà, S., 2001. A neural network model for visibility nowcasting from surface observations: Results and sensitivity to physical input variables. *Journal of Geophysical Research: Atmospheres*, 106(D14), 14951-14959.

[24] Marzban, C., 1998. Scalar measures of performance in rare-event situations. *Weather Forecasting*, 13, 753-763.

[25] Marzban, C., 2000. A neural network for tornado diagnosis: Managing local minima. *Neural Comput Appl.*, 9, 133-141.

[26] Marzban, C., and G.F. Stumpf, 1996. A neural network for tornado prediction based on Doppler radar-derived attributes. *J. Appl. Meteorol.*, 35, 617-626.

[27] Marzban, C., and G.F. Stumpf, 1998. A neural network for damaging wind prediction. *Weather Forecasting*, 13, 151-163.

[28] Marzban, C., E.D. Mitchell, and G.J. Stumpf, 1999. On the notion of "best predictors": An application to tornado prediction. *Weather Forecasting*, 14, 1007-1016.

[29] Peixoto J.P., Oort A.H. 1992, *Physics of Climate*, American Institute of Physics. (4)

[30] Taylor, F. W. 2005. *Elementary climate physics*. Oxford Univ. Press.

[31] Goosse H., P.Y. Barriat, W. Lefebvre, M.F. Loutre and V. Zunz, (2008-2010). *Introduction to climate dynamics and climate modelling*. ISBN: 9781107445833

[32] Petit, J. R., Jouzel, J., Raynaud, D., Barkov, N. I., Barnola, J. M., Basile, I., ... & Delmotte, M. 1999. *Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica*. Nature, 399(6735), 429.

[33] Meijer, H. A. J., Smid, H. M., Perez, E., & Keizer, M. G., 1996. *Isotopic characterisation of anthropogenic CO2 emissions using isotopic and radiocarbon analysis*. Physics and Chemistry of the Earth, 21(5-6), 483-487.

[34] Jones PD, Lister DH, Osborn TJ, Harpham C, Salmon M and Morice CP, 2012. *Hemispheric and large-scale land-surface air temperature variations: an extensive revision and an update to 2010*. Journal of Geophysical Research, 117, D05127, doi:10.1029/2011JD017139.

[35] Morice CP, Kennedy JJ, Rayner NA and Jones PD, 2012. *Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: the HadCRUT4 dataset*. Journal of Geophysical Research, 117, D08101, doi:10.1029/2011JD017187

[36] Osborn TJ and Jones PD, 2014. *The CRUTEM4 land-surface air temperature data set: construction, previous versions and dissemination via Google Earth.* Earth System Science Data 6, 61-68 (doi:10.5194/essd-6-61-2014).

[37] Pasini, A., 2005. *From Observations to Simulations: A Conceptual Introduction to Weather and Climate Modelling.* World Scientific Publishers, Singapore.

[38] McCulloch, W. S. and Pitts, W. 1943. *A logical calculus of the ideas immanent in neural nets.* Bulletin of Mathematical Biophysics, 5:115–37.

[39] Rosenblatt, F. 1958. The perceptron: *A probabilistic model for information storage and organization in the brain*. Psychological Review, 65:386–408.

[40] Rosenblatt, F. 1962. *Principles of Neurodynamics*. New York: Spartan.

[41] Widrow, B. and Hoff, M. E. 1960. *Adaptive switching circuits*. In IRE WESCON convention Record, volume 4, pp. 96–104, New York.

[42] LeCun, Y., Kanter, I. and Solla, S. A. 1991. *Second order properties of error surfaces: Learning time and generalization.* In Advances in Neural Information Processing Systems, volume 3, pp. 918–24. Cambridge, MA: MIT Press.

[43] Hsieh W.W., 2009. *Machine learning methods in the environmental sciences*. Cambridge. Cambridge University Press, 364

[44] Rumelhart, D. E., Hinton, G. E. and Williams, R. J. 1986. *Learning internal representations by error propagation*. In Rumelhart, D., McClelland, J. and Group, P. R., eds., Parallel Distributed Processing, volume 1, pp. 318–62. Cambridge, MA: MIT Press.

[45] Werbos, P. J. 1974. *Beyond regression: new tools for prediction and analysis in the behavioural sciences.* Ph.D. thesis, Harvard University.

[46] Nguyen D. and Widrow, B. 1990 *Improving the learning speed of 2-layer neural networks by choosing initial values of the adaptive weights*. Proceedings of the International Joint Conference on Neural Networks, 3:21–26.

[47] Hestenes, M. R. and Stiefel, E. 1952. *Methods of conjugate gradients for solving linear systems*. Journal of Research of the National Bureau of Standards, 49(6):409–36.

[48] Fletcher, R. and Reeves, C. M. 1964. *Function minimization by conjugate gradients*. Computer Journal, 7:149–54.

[49] Polak, E. and Ribiere, G. 1969. *Note sur la convergence de methods de directions conjures*. Revue Francaise d'Informat. et de Recherche Operationnelle, 16:35–43.

[50] Polak, E. 1971. *Computational Methods in Optimization: A Unified Approach*. New York: Academic Press.

[51] Davidon, W. C. 1959. *Variable metric methods for minimization*. A.E.C.Res. and Develop. Report ANL-5990, Argonne National Lab.

[52] Fletcher, R. and Powell, M.J.D. 1963. *A rapidly convergent descent method for minimization*. Computer Journal, 6:163–8.

[53] Broyden, C. G. 1970. *The convergence of a class of double-rank minimization algorithms*. Journal of the Institute of Mathematics and Its Applications, 6:76–90.

[54] Fletcher, R. 1970. *A new approach to variable metric algorithms*. Computer Journal, 13:317–22.

[55] Goldfarb, F. 1970. *A family of variable metric methods derived by variational means*. Mathematics of Computation, 24:23–6.

[56] Shanno, D. F. 1970. *Conditioning of quasi-Newton methods for function minimization*. Mathematics of Computation, 24:647–57.

[57] Shanno, D. F. 1978. *Conjugate-gradient methods with inexact searches*. Mathematics of Operations Research, 3:244–56.

[58] Levenberg, K. 1944. *A method for the solution of certain non-linear problems in least squares.* Quarterly Journal of Applied Mathematics, 2:164–8.

[59] Marquardt, D. 1963. *An algorithm for least-squares estimation of nonlinear parameters*. SIAM Journal on Applied Mathematics, 11:431–41.

[60] Amari, S., Murata, N., Müller, K.-R., Finke, M. and Yang, H. 1996. *Statistical theory of overtraining – is cross validation asymptotically effective?* Advances in Neural Information Processing Systems, 8:176–182.

[61] Pasini, A., Potestà, S. 1995. *Short-range visibility forecast by means of neural-network modelling: a case-study*. Il Nuovo Cimento 18:505-16

[62] Pasini, A., Mazzocchi, F., 2015. *A multi-approach strategy in climate attribution studies: Is it possible to apply a robustness framework?* Environ. Sci. Pol. 50, 191-199

[63] Morice, C.P., Kennedy, J.J., Rayner, N.A., Jones, P.D., 2012. *Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: the HadCRUT4 dataset.* J. Geophys. Res. 117, D08101

[64] Ramaswamy, V. et al. 2001. *Radiative Forcing of Climate Change. In: Climate Change 2001: TheScientific Basis* (ed. Houghton, J. T. et al.), Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 349–416.

[65] Kattenberg, A. et al. 1996. *Climate models–projections of future climate. In: Climate Change 1995: The Science of Climate Change* (ed Houghton, J. T. et al.), Cambridge University Press, Cambridge, 285–357.

[66] Smith, S. J. et al. 2011. *Anthropogenic sulfur dioxide emissions: 1850–2005*. Atmosph. Chem. Phys. 11, 1101–1116.

[67] Klimont, Z., Smith, S. J. & Cofala, J., 2013. *The last decade of global anthropogenic sulfur dioxide: 2000–2011 emissions*. Environ. Res. Lett. 8, 014003

[68] Wigley, T. M. L. & Raper, S. C. B. 1992. *Implications for climate and sea level of revised IPCC emissions scenarios*. Nature 357, 293–300

[69] Boucher, O. & Pham, M. 2002. *History of sulfate aerosol radiative forcings.* Geophys. Res. Lett. 29(9), 22-1–22-4

[70] Meinshausen, M. et al. 2011. *The RCP GHG concentrations and their extension from 1765 to 2300*. Clim. Change 109, 213–241

[71] Lean, J. 2000. *Evolution of the sun's spectral irradiance since the Maunder Minimum*. Geophys. Res. Lett. 27, 2425–2428

[72] Sato, M., Hansen, J. E., McCormick, M. P. & Pollack, J. B. 1993. *Stratospheric aerosol optical depth, 1850–1990*. J. Geophys. Res. 98, 22987–22994

[73] Miller, R. L. et al. 2014. *CMIP5 historical simulations (1850–2012) with GISS ModelE2*. J. Adv. Model. Earth Syst. 6, 441–477

[74] Hansen, J., Sato, M., Kharecha, P. & von Schuckmann, K. 2011. *Earth's energy imbalance and implications*. Atmos. Chem. Phys. 11, 13421–13449.

[75] Schönwiese, C.-D., Walter, A. & Brinckmann, S. 2010. *Statistical assessments of anthropogenic and natural global climate forcing. An update*. Meteorol. Z. 19, 3–10.

[76] Verdes, P. F., 2007. *Global warming is driven by anthropogenic emissions: A time series analysis approach.* Phys. Rev. Lett. 90, 048501

[77] Bindoff, N. L. et al. 2013. *Detection and Attribution of Climate Change: from Global to Regional. In: Climate Change 2013: The Physical Science Basis* (ed. Stocker, T. F. et al.), Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 867–952

[78] DelSole, T., Tippett, M. K. & Shukla, J. 2011. *A significant component of unforced multidecadal variability in the recent acceleration of global warming*. J. Clim. 24, 909–926.

[79] Macias, D., Stips, A. & Garcia-Gorriz, E. 2014. *Application of the singular spectrum analysis technique to study the recent hiatus on the global surface temperature record*. PLoS ONE 9, e107222.

[80] Pasini, A., Triacca, U. & Attanasio, A. 2017 *Evidence for the role of the Atlantic multidecadal oscillation and the ocean heat uptake in hiatus prediction*. Theor. Appl. Clim. 129, 873–880.

[81] Lean, J. L. & Rind, D. H. 2008. *How natural and anthropogenic influences alter global and regional surface temperatures: 1889 to 2006*. Geophys. Res. Lett. 35, L18701

[82] Attanasio, A., Pasini, A. & Triacca, U. 2016. *Has natural variability a lagged influence on global temperature? A multi-horizon Granger causality analysis*. Dynamics and Statistics of the Climate System 1, 1–16.

[83] Ricke, K. L. & Caldeira, K. 2014. *Maximum warming occurs about one decade after a carbon dioxide emission*. Environ. Res. Lett. 9, 124002.

[84] Del Sole, T., Tippett, M.K., Shukla J. (2011), *A Significant Component of Unforced Multidecadal Variability in the Recent Acceleration of Global Warming*. Journal of Climate 24, 909-926.

[85] O.H. Otterå, M. Bentsen, H. Drange, L. Suo (2010), *External forcing as a metronome for* Atlantic multidecadal variability. Nature Geoscience 3, 688-694.

[86] B.B.B. Booth, N.J. Dunstone, P.R. Halloran, T. Andrews,N. Bellouin (2012). *Aerosols implicated as a prime driver of twentieth-century North Atlantic climate variability*. Nature 484, 228-232.

[87] R. Zhang et al. (2013). *Have Aerosols Caused the Observed Atlantic Multidecadal Variability*? Journal of the Atmospheric Sciences 70, 1135-1144.

[88] M.F. Knudsen, B.H. Jacobsen, M.-S. Seidenkrantz, J. Olsen (2014). *Evidence for external forcing of the Atlantic Multidecadal Oscillation since termination of the Little Ice Age*. Nature Communications 5, 3323.

[89] A. Clement, K. Bellomo, L.N. Murphy, M.A. Cane, T. Mauritsen, G. Rädel, B. Stevens (2015). *The Atlantic Multidecadal Oscillation without a role for ocean circulation*. Science 350, 320-324.

[90] A. Bellucci, A. Mariotti, S. Gualdi 2017, *The Role of Forcings in the Twentieth-Century North Atlantic Multidecadal Variability: The 1940–75 North Atlantic Cooling Case Study*. Journal of Climate 30, 7317-7337.

[91] M.A. Cane, A.C. Clement, L.N. Murphy, K. Bellomo (2017), Low-Pass Filtering, Heat Flux, and Atlantic Multidecadal Variability, Journal of Climate 30, 7529-7553.

[92] L.N. Murphy, K. Bellomo, M.A. Cane, A.C. Clement (2017), The role of historical forcings in simulating the observed Atlantic multidecadal oscillation, Geophysical Research Letters 44, 2472-2480.

[93] K. Bellomo, L.N. Murphy, M.A. Cane, A.C. Clement, L.M. Polvani (2018), Historical forcings as main drivers of the Atlantic Multidecadal variability in the CESM large ensemble, Climate Dynamics 50, 3687-3698.

[94] W.M. Kim, S.G. Yeager, G. Danabasoglu (2018), Key Role of Internal Ocean Dynamics in Atlantic Multidecadal Variability During the Last Half Century, Geophysical Research Letters 45, 13,449-13,457.

[95] Black R., Bennett S.R.G., Thomas S., M. and Beddington J. R. 2011. *Migration as adaptation* Nature 478 447–9

[96] Mallick B. and Etzold B. (ed). 2015. *Environment, Migration and Adaptation Evidence and Politics of Climate Change in Bangladesh* (Dhaka: AH Development Publishing House)

[97] Hsiang S.M., Meng K.C. and Cane M.A. 2011. *Civil conflicts are associated with the global climate*. Nature 476 438–41

[98] Werrell C.E., Fermia F. and Slaughter A. (ed) 2013. *The Arab Spring and Climate Change* (Washington DC: Center for Climate and Security)

[99] UNCCD 2014. *Desertification. The invisible frontline*. 2nd edn (Bonn: United Nations Convention to Combat Desertification) https://www. unccd.int/publications/desertification-invisible-frontline-second-edition

[100] Kelley C.P., Mohtadi S., Cane M.A., Seager R. and Kushnir Y. 2015. *Climate change in the fertile crescent and implications of the recent syrian drought*. Proc. Nat. Acad. Sci. (US) 112 3241–6

[101] Cai R., Feng S., Oppenheimer M. and Pytlikova M. 2016. *Climate variability and international migration: the importance of the agricultural linkage*. J. Environ. Econ. Manag. 79 135–51

[102] Carleton T. and Hsiang S. 2016. *Social and economic impacts of climate*. Science 353 6304

[103] Schleussner C.-F., Donges J.F., Donner R.V. and Schellnhuber H.J. 2016. *Armed-conflict risks enhanced by climate-related disasters in ethnically fractionalized countries* Proc. Nat. Acad. Sci. (US) 113 9216–21

[104] Mastrojeni G. and Pasini A. 2017. *Greenhouse effect, war effect. Climate, conflicts, migrations: Italy in frontline* (Milan: Chiarelettere publisher)

[105] Missirian A. and Schlenker W. 2017. *Asylum applications respond to temperature fluctuations* Science 358 1610–4

[106] Castles S. 2003. *Towards a sociology of forced migration and social transformation Sociology* 37 13–34

[107] Adams C., Ide T., Barnett J. and Detges A. 2018. *Sampling bias in climate-conflict research* Nature Clim. Change 8 200–3

[108] Pasini A., Amendola S. 2019. *Linear and nonlinear influences of climatic changes on migration flows: a case study for the 'Mediterranean bridge'.* Environmental Research Communications. 1 011005

[109] Heston A., Summers R. and Aten B. 2011. Penn *World table Version 7.0* (Philadelphia: Center for International Comparisons of Production, Income and Prices at the University of Pennsylvania)

[110] Rienecker M.M., Suarez M.J., Gelaro R., Todling R., Bacmeister J., Liu E., Bosilovich M.G., Schubert S.D., Takacs L. and Kim G.-K. 2011. *MERRA: NASA's modern-era retrospective analysis for research and applications*. J. Clim. 24 3624–48

[111] Herold N., Alexander L., Green D. and Donat M. 2017. *Greater increases in temperature extremes in low versus high income countries* Environ. Res. Lett. 12 034007

[112] Mora C. et al 2017. *Global risk of deadly heat*. Nature Clim. Change 7 501–6

[113] Muñoz, C. et al., 2018. *Investigations of Phlebotomus perniciosus sand flies in rural Spain reveal strongly aggregated and gender-specific spatial distributions and advocate use of light-attraction traps.* Medical and Veterinary Entomology, 32(2), 186–196.

[114] Dantas-Torres, F., Tarallo, V.D. & Otranto, D. (2014) *Morphological keys for the identification of Italian phlebotomine sand flies (Diptera: Psychodidae: Phlebotominae).* Parasites & Vectors, 7, 479.

[115] Latrofa, M.S., Annoscia, G., Dantas-Torres, F., Traversa, D. & Otranto, D., 2012. *Towards a rapid molecular identification of the common phlebotomine sand flies in the Mediterranean region.* Veterinary Parasitology, 184, 267–270.

[116] Lane, R.P. & Crosskey, R.W. (1993) *Medical Insects and Arachnids.* Chapman and Hall, London.

[117] Cruthers, L.R. & Pombi, M. (2019) *Arthropoda, Diptera, Nematocera, in Parasiticide Screening: Volume 1: In Vitro and In Vivo Tests with Relevant Parasite Rearing and Host Infection/Infestation Methods* (ed. by Marchiondo, A.A., Cruthers, L.R. & Fourie J.J.), pp. 21–…. Academic Press, New York.

[118] Ballart, C. et al. (2014) *Importance of individual analysis of environmental and climatic factors affecting the density of Leishmania vectors living in the same geographical area: the*

*example of Phlebotomus ariasi and P. perniciosus in northeast Spain*. Geospatial Health, 8(2), 389–403.

[119] Gómez-Bravo, A., German, A., Abril, M., Scavuzzo, M. & Salomón, O.D. (2017) *Spatial population dynamics and temporal analysis of the distribution of Lutzomyia longipalpis (Lutz & Neiva, 1912) (Diptera: Psychodidae: Phlebotominae) in the city of Clorinda, Formosa, Argentina.* Parasites and Vectors, 10(1), 352.