Weather Forecasting Considerations

Celso Luis Levada[1], Osvaldo Missiato[2], Antonio Luis Ferrari[3], Miriam de Magalhães Oliveira Levada[1]

[1]*Hermínio Ometto Foundations-Fho/Uniararas/Brazil*
[2]*Faculdades Integradas Einstein De Limeira/Brazil*
[3]*Faculdade De Engenharia E Agrimensura De Pirassununga/Brasil*

Abstract: - **A numerical climate model is the computer program that simulates an atmospheric movement in space and time. A variety of meteorological phenomena can be analyzed and predicted by different types of numerical weather forecasting models. Numerical weather forecast uses mathematical models of the atmosphere and oceans to predict the climate based on current weather conditions. Manipulating the vast datasets and performing the complex calculations necessary for modern numerical weather forecasting requires some of the world's most powerful supercomputers. A more fundamental problem lies in the chaotic nature of the partial differential equations governing the atmosphere. The idea of numerical weather forecasting is to sample the state of air at a given time and use the fluid dynamics and thermodynamics equations to estimate the state of air at some point in the future. It is impossible to solve these equations exactly, and small errors grow with time.**

I. INTRODUCTION

This introduction is based on the articles "Evolution of climate models and weather and climate forecasting" by SAMPAIO e SILVA DIAS(2), An overview of numerical weather prediction models of GAZTELUMEND(1) and History of Numerical Weather Prediction, by FLYNN(3), that make a historical development about weather and climate forecasting. The review begins in the year 1904, when a more explicit analysis for weather forecasting was proposed by the Norwegian Vilhelm Bjerknes, who established a two-year plan steps to make weather forecasts. In the first stage, called diagnosis, the initial state of the atmosphere is determined from observations in weather stations and instrumented balloons. Secondly, a prognostic step originated, in which the laws governing the state of the atmosphere are used to calculate temporal evolution. At that time, there were already a small number of observations of the atmosphere, and some observational programs were started, which allowed a reasonable diagnosis of the state of the atmosphere, at least in part of Europe. Bjerknes proposed a set of equations that represent physical principles of conservation of energy, mass and momentum, as well as diagnostic relationships between pressure, temperature and density. He determined a set of seven independent equations, one for each dependent variable, that describe the temporal evolution of the atmosphere. Bjerknes developed a graphical method for solving the equations, because he could not solve them numerically because of the complexity of the equations. In 1913 Lewis Fry Richardson proposed that the physical principles governing the behavior of the atmosphere, expressed by the system of mathematical equations defined by Bjerknes, could be solved by discretizing the atmosphere in a latitude / longitude mesh and vertical columns, that is, through a finite difference method. The results of applying this method were very bad and unrealistic. In the 1930s, mathematician John von Neumann became interested in flows in turbulent fluids and proposed that advances in hydrodynamics could numerically advance from the solution of complex equations. Von Neumann led the construction of an electronic computer at the Princeton Institute for Advanced Studies, which, among other things, was used for early weather forecasts. Jule Charney, who in the 1940s, based on the analysis of the equations defined by Bjerknes, used the scale analysis technique of meteorological phenomena and reduced the complex model used by Richardson to a single equation. It was with this simplified model that the first weather forecasts were made in the late 1940s. By the early 1950s, the Princeton meteorology group had already completed the necessary mathematical analyzes and designed a numerical algorithm to solve the system of equations a little more complex than that used by Charney and von Neumann. However, each 24-hour integration took about 24 hours to perform at Eniac, meaning that it was not practical, but theoretically useful. The first application with more realistic equations than those used by Charney and Von Neuman was a success. It was the first time that primitive equations were routinely used. The first longer term numerical simulation (about a month) was performed by Phillips (1956), who used the simulation from a zonal flow with small random perturbations and a wave-like disturbance. Phillips found good agreement with observations of weather systems in the atmosphere. Today, computers that are capable of doing something around a quintillion sum and subtraction operations per second are used to forecast weather and climate. Satellite weather observations began to be made in the 1960s but it was not until the late 1970s that the information collected was more accurate. During the 1980s, with the evolution of computers, various weather centers around the world began to make numerical weather forecasts. From that moment on, it was possible to make predictions a few days in advance. In the 1990s, computers were further refined, and mathematical models as well. Nowadays, it is possible to predict the time days in advance of the order of one week with very high hit rates. The nature of modern weather forecasting is not only highly complex, but also highly quantitative. There are three methods used in weather forecasting: Synoptic Weather Forecast, Numerical Methods, and Statistical Methods.

II. SYNOTIC METEOROLOGICAL FORECAST

This is considered to be the most traditional approach to weather forecasting, where the weather map describing the weather at a given time is widely used. We can say that in synoptic weather forecasting there is no scientific basis and little quantification occurs. Generally speaking, the sequence of events on a weather map is interpreted subjectively, depending on the experience and skill of the individual predictor. To get an average view of the climate change pattern, the weather center prepares a series of synoptic maps every day. To regularly prepare synoptic maps there is a huge collection and analysis of observational data obtained from thousands of weather stations. From the careful study of weather charts over many years, certain empirical rules have been formulated. These rules helped the predictor estimate the speed and direction of movement of climate systems. However, the sudden changes in this system make forecasts only valid for a short time, say a few hours or a day.

III. NUMERICAL WEATHER FORECASTING

ARAVÉQUIA & LEAL DE QUADRO(7) , comment that the idea of numerical weather forecasting is to sample the state of air at a given time and use the fluid dynamics and thermodynamics equations to estimate the state of air at some point in the future. A numerical climate model is the computer program that simulates an atmospheric movement in space and time. A variety of meteorological phenomena can be analyzed and predicted by different types of numerical weather forecasting models. Numerical weather forecast uses mathematical models of the atmosphere and oceans to predict the climate based on current weather conditions. Several global and regional forecast models are run in different countries around the world, using current weather observations transmitted from radiosonde, meteorological satellites, and other observing systems as inputs. Mathematical models based on the same physical principles can be used to generate short-term weather forecasts or long-term climate predictions; the latter are widely applied to understand and design climate change. Manipulating the vast datasets and performing the complex calculations necessary for modern numerical weather forecasting requires some of the world's most powerful supercomputers. A more fundamental problem lies in the chaotic nature of the partial differential equations governing the atmosphere. It is impossible to solve these equations exactly, and small errors grow with time. In addition, the partial differential equations used in the model need to be complemented with parametrizations for solar radiation, wet processes, heat exchange, soil, vegetation, surface water and the effects of the terrain. The equations used are nonlinear partial differential equations, impossible to be solved exactly by analytical methods, with the exception of some idealized cases. Therefore, numerical methods obtain approximate solutions. According ARAVÉQUIA & LEAL DE QUADRO(7), different models use different methods of solution: some global models and almost all regional models use finite difference methods for the three spatial dimensions, while other global models and some regional models use spectral methods for horizontal dimensions and finite differences methods. These equations are initialized from the analysis data and the change rates are determined. These rates of change predict the state of the atmosphere in a short time in the future; the time increment for this prediction is called the time step. This future atmospheric state is then used as a starting point for another application of predictive equations to find new rates of change, and these new rates of change predict the atmosphere at an even greater step in the future. This time interval is repeated until the solution reaches the desired forecast time. The length of the chosen time interval within the model is related to the distance between the grid points and is chosen to maintain numerical stability. The equations are solved by the computer at each point for a very short period of time, say 10 minutes. By repetitive calculations for every 10 minutes following, the prediction is obtained by 24, 48 or 72 hours ahead.

IV. STATISTICAL METHODS

Statistical models are mathematical relationships that can show the most probable scenarios based on statistics obtained from historical series of observed data of one or more meteorological variables. Its advantage is to require few computational resources and its disadvantage is the need for long historical series for better performance. Statistical methods usually complement the numerical method. Statistical methods use the previous records of meteorological data based on the assumption that the future will be a repetition of the past tense. The main purpose of studying the meteorological data of the past is to discover the climate aspects that are good indicators of future events. Once you establish these relationships, the correct data can be safely used to predict future conditions. Only the overall time can be predicted in this way. It is particularly useful for designing only one aspect of time at a time. For example, it is of great value to predict the maximum temperature of a day at a given location. The procedure consists of compiling statistical data relating temperature to wind speed and direction, amount of cloudiness, humidity and the specific season of the year. Later, this data is displayed in graphs. These graphs provide an estimate of the maximum temperature of the day from the data of the current conditions. Statistical methods are of great value in long-range weather forecasts. The National Weather Service prepares monthly and weekly weather forecasts. In fact, these are not weather forecasts in the strict sense of the term. They represent estimates or projections of the precipitation and temperatures that can be expected during these periods. These estimates only give an idea if the temperature and precipitation in the region will be above or below normal. Another statistical approach to weather forecasting is called the Analog Method. In this method, an attempt is made to identify in the previous climate records, climatic conditions similar to the current conditions. Once such analogous conditions are located, it is assumed that currently the same sequence of climatic events will follow as shown in previous records.

V. FUNDAMENTALS EQUATIONS

The forecast of the time is basically the simulation made from the equation of Navier Stockes[5,6] with the help of the known initial conditions. Initial conditions may be density, temperature, velocity, and pressure. The Navier Stokes equations are differential equations that describe the flow of fluids. They are partial differential equations that allow determining the velocity and pressure fields in a flow. They were named after Claude Louis Navier and George Gabriel Stokes developed a set of equations that would describe the movement of fluid substances such as liquids and gases. These equations establish that changes in momentum and acceleration of a fluid particle. They represent the result of changes in pressure and dissipative viscous forces acting on the fluid, which originate in molecular interaction. These are one of the most useful sets of equations, since they describe the physics of a large number of phenomena of economic and academic interest, including in several branches of engineering. They are used to model climate, ocean currents, water flows into oceans, estuaries, lakes and rivers, movements of stars in and out of the galaxy, flow around automobile and aircraft aerofils, spread of smoke in fires and models of dispersion. The Navier-Stokes equation for compressible fluid is given by:

$$\rho\left(\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v}\cdot\nabla\mathbf{v}\right) = -\nabla p + \nabla\cdot\mathbf{T} + \mathbf{f}$$

In this equation $\mathbf{v}$ is the flow velocity, ρ is the fluid density, p is the pressure, T is the component of the total stress tensor, which has order two, f represents body forces per unit volume acting on the fluid and ∇ is the del operator. In addition to the Navier Stockes equation, the mass conservation equation is required.

$$\frac{\partial \rho}{\partial t} + \nabla\cdot(\rho\mathbf{v}) = 0$$

The thermodynamic equation[4] that represents the state of the gas, is also part of the system of equations

$$p=\rho RT$$

Additionally we have the energy conservation equation, represented by the first law of thermodynamics, $c_v\, dT + RT\, \nabla.V = Q$, where c_v is the specific heat at constant volume, T is the temperature, R is the constant of gas, and Q is the heating.

Posteriorly we have the second form of the energy equation [4] given by

$$\rho\frac{D\left(e+\frac{V^2}{2}\right)}{Dt} = \frac{\partial}{\partial t}\left[\rho\left(e+\frac{V^2}{2}\right)\right] + \nabla\left[\rho\left(e+\frac{V^2}{2}\right)\vec{V}\right]$$

$$where\ \frac{E_{total}}{m} = e$$

For complex situations, such as a global climate system, solutions to the Navier-Stokes[6,7] equation are often to be found with the help of computers. This is a field of science known as CFD, or Computational Fluid Dynamics.

VI. GRID-POINT MODEL

The figure (1) shows a grid model, where, a geographical area can be neatly divided into a mesh of regularly spaced points called a grid. For the record, these grid points are the locations at which the computer calculates the numerical forecast. The spacing between grid points varies from grid-point model to grid-point model. Some grid-point models have a "coarse" mesh, with large spacing between relatively few grid points. Other grid-point models are "fine" mesh, with a small spacing between relatively many grid points. Whether the mesh of grid points is coarse or fine ideally governs the quality of the forecast. Mimicking a game of three-dimensional chess, there are also meshes of regularly spaced points at specified altitudes, stacking from the ground to the upper reaches of the atmosphere. Thus, a three-dimensional array emerges that covers a great volume of the atmosphere, ready to be filled with weather data at each grid point. The typical horizontal spacing of grid points for operational computer models is typically on the order of ten kilometers. The vertical spacing of grid points varies from tens to hundreds of meters. Once a simulation has begun, the computer calculates the values of moisture, temperature, wind, and so on, at each grid point at a future time, typically a few virtual minutes into the future. To perform this feat, powerful, high-speed supercomputers make trillions to quadrillions of calculations each second. Thereafter, the computer calculates the same parameters for the next forecast time, and so on. For a short-range prediction, this "leapfrog" time scheme typically ends 84 hours into the virtual future, taking on the order of an hour of real time to complete. Even leap frogging just a few virtual minutes into the future is fraught with error because the computer makes calculations for one time and then “leaps” to make calculations a few more virtual minutes later, skipping calculations for intermediate times between the starting and ending points of the leap. Like taking short cuts while solving a complicated algebra problem, skipping steps inevitably leads to errors. Forecasters could, theoretically, make a more accurate computer forecast by reducing the size of the time interval in the leapfrog scheme. However, a smaller time interval would require faster and more powerful computers to support the increased computational load. Also, meteorologists could increase the number of grid points to improve forecasts with the hope that smaller grid spacing could better capture smaller-scale weather phenomena. But such a scheme also demands faster and faster supercomputers. Though technology continues to advance, there is a practical limit to what computers can do, so there will never be a perfect computer forecast (ARAVÉQUIA & LEAL DE QUADRO[7]).

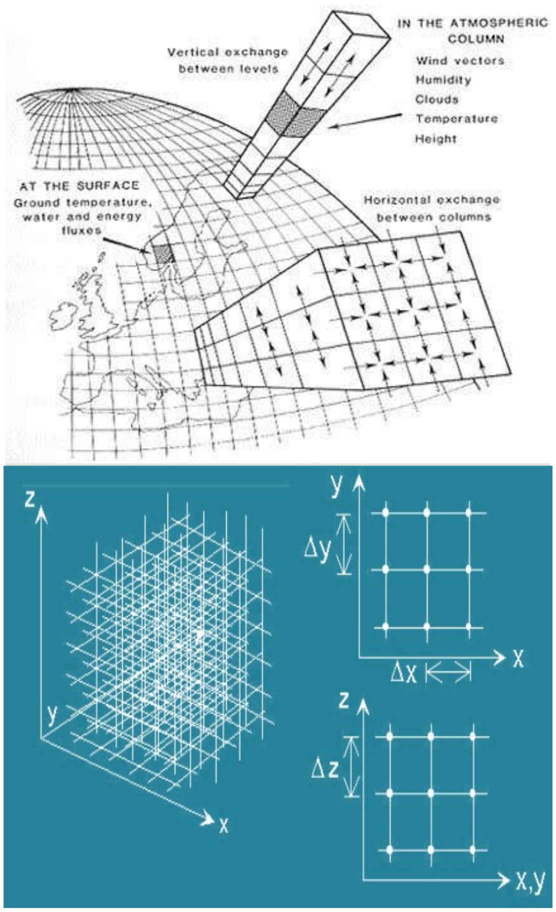

Figura 1- Grid Point Model fonte:www.euve.org

VII. LINEAR REGRESSION

Simple linear regression is the most commonly used statistical technique; It is a way of modeling a relationship between two sets of variables. In general, linear regression is a statistical method that uses between two or more variables so that one variable can be estimated (or predicted) from the other or others. The model is called simple when it involves only two variables and multiple when it has more than two variables. The result is a linear regression equation that can be used to make predictions about the data. WITT[8] mentions that several scientists have applied a simple linear regression method to predict meteorological parameters such as temperature, pressure, wind speed, humidity, cloudiness and precipitation, among others. If the relationship between the two variables is approximately linear, then the data can be summarized by fitting a line through the data. The equation of this line is given by

$$y = ax + b$$

where a is known as the intercept and b is the slope. Intuitively, we want a line that gives small differences between the true weights and those given by the line for the corresponding heights. The standard method for obtaining the best fit line is called least squares which literally mini- mizes the sum of squares of distances from yi to the fit line. In principle this requires drawing possible lines by calculating the sum of squares of distances:

$$S = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad = \sum_{i=1}^{n}\{y_i - (a + bx_i)\}^2$$

and find the values of a and b (equivalently the line) that provide the smallest value of S. It is possible to show that the best line is one such that

$$b = \frac{\sum(y_i - \bar{y})(x_i - \bar{x})}{\sum(x_i - \bar{x})^2} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x^2 - n\bar{x}^2} = \frac{s_{xy}}{s_x^2}$$

$$a = \bar{y} - b\bar{x}.$$

Multiple Linear Regression[8] is a statistical technique that uses several explanatory variables to predict the outcome of a response variable. The objective of multiple linear regression is to model the relationship between explanatory variables and responses. It is common in the scientific field to apply Multiple Linear Regression to develop a model to predict the weather conditions for a given season using locally collected data.

As defined by WONNACOTT (9) regression model in this case is given by: $Y_i = \beta_o + \sum_{j=1}^{k} \beta_j x_{ij} + \varepsilon_i$

com $i = 1, 2, ..., n$

The model presented in the equation is a system of n equations that can be represented matrix by

$$Model: \quad \underline{Y} = X\underline{b} + \underline{\varepsilon}$$

Being that

$$\underline{Y} = \begin{bmatrix} Y_1 \\ . \\ . \\ . \\ Y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ . & . & . & \cdots & . \\ . & . & . & \cdots & . \\ . & . & . & \cdots & . \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix} \quad \underline{b} = \begin{bmatrix} b_0 \\ . \\ . \\ . \\ b_k \end{bmatrix} \quad \underline{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ . \\ . \\ . \\ \varepsilon_n \end{bmatrix}$$

Linear Correction Coefficient[9] or Pearson's Coefficient is a parameter (r) that measures the relationship between two variables within the same metric scale, and serves to determine how strong the relationship exists between known data sets or information. It is given by:

$$r = \frac{n\sum(x_i \cdot y_i) - (\sum x_i)(\sum y_i)}{\sqrt{n\sum x_i^2 - (\sum x_i)^2} \cdot \sqrt{n\sum y_i^2 - (\sum y_i)^2}}$$

$$-1 \le r \le 1$$

We use the term positive correlation when r> 0, and in this case as x grows y also grows, and negative correlation when r <0, and in this case as x grows y decreases (on average). The higher the value of r (positive or negative), the stronger the association. At the extreme, if r = 1 or r = -1 then all points on the scatter plot fall exactly in a straight line. At the other extreme, if r = 0 there is no linear association. Of course, interpretations depend on each particular context.

Note the following examples:

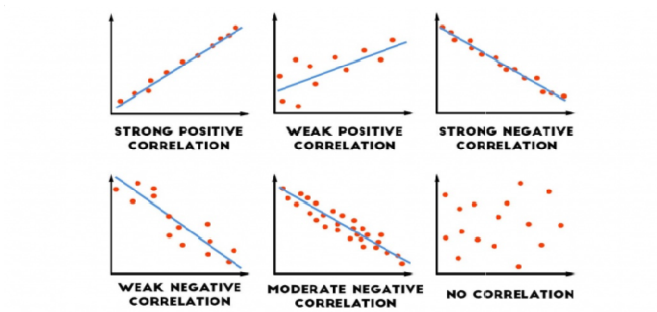

Fig3: Examples of correlation-Source: http://danshiebler.com/2017-06-25-metrics/

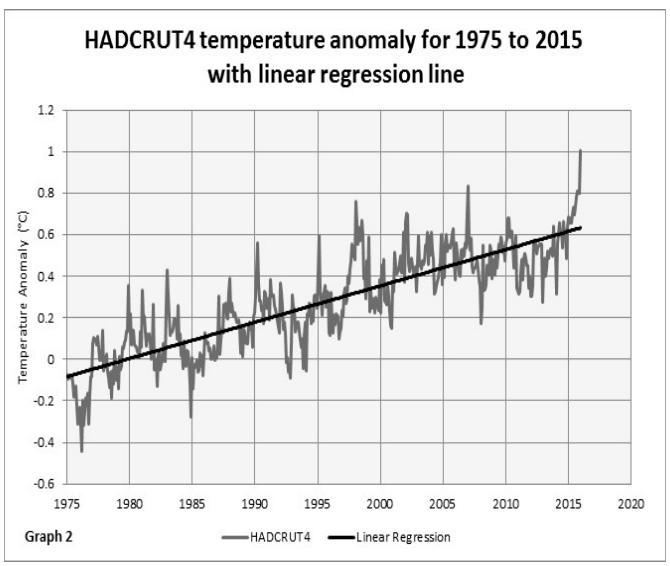

Fig.2 monthly temperature anomaly and linear regression line for 1975 to 2015. Source : https://wattsupwiththat.com

VIII. FINAL CONSIDERATIONS

Modeling and weather forecasting are crucial for both the demand and supply side of the climate derivatives market. On the demand side, to assess the hedging potential against weather surprises and formulate the appropriate hedging strategies, it is necessary to determine how much noise there is for time derivatives to eliminate. This requires modeling and weather forecasting. Supply side, default Approaches to arbitrage free pricing are irrelevant in time derived contexts, and thus the only way to reliably price options is again modeling and predicting the underlying time variable. Interestingly enough, it seems that little thought has been given the crucial question of how best to approach climate modeling and forecasting in the context of climate derivatives demand and supply. The vast majority of existing weather conditions prediction literature has a structural "atmospheric science" and although this approach is certainly best for very short-range forecasting, as results and those of many others it is not obvious that better for longer horizons relevant to climate derivatives, like 12 weeks or 6 months. In addition, density predictions, but not point forecasts, they are of utmost relevance in the context of derivatives. Good distributive forecasting does not necessarily require a structural model, but it does require accurate approximations to the stochastic dynamics.

BIBLIOGRAPHICS REFERENCES

[1]. GAZTELUMENDI, S. - An overview of numerical weather prediction models, Europe Virtual Engineering, www.euve.org, text available in http://www.bcamath.org/documentos_public/archivos/actividades_cientificas/TalkBCAMWorkshoponMultiphysics20100607SG.pdf, access in 30/05/2019

[2]. SAMPAIO, G. and SILVA DIAS,P.L. Evolution of climate models, USP Journal, São Paulo, n. 103 p. 41-54, 2014; available in https://www.revistas.usp.br/revusp/article/download/99179/97655/, access in 30/05/2019.

[3]. FLYNN, C.- A history of Numerical Weather Prediction, text available in https://blog.metservice.com/HistoryNWP, access in 12/06/2019

[4]. ÇENGEL, Y. A.; CIMBALA, J. M. Fluid Mechanics: Fundamentals and Applications. Porto Alegre: MCGRAW-HILL BRASIL, 2007. p. 345-374.

[5]. ORREGO, D.A.F.- Navier-Stokes Equations, University of Sao Paulo Polytechnic School, Brazil 2012, text available in https://www.academia.edu/1531566/Equacoes_de_Navier-Stokes_Navier-Stokes_Equations_, access in 30/06/2019

[6]. HARWOOD, R.S. Atmospheric Dynamics, University of Edinburgh, 2005 available in http://xweb.geos.ed.ac.uk/~rharwood/teaching/phys4/dynamics/index.html_, access in 20/06/2019

[7]. ARAVÉQUIA,J.A. and LEAL DE QUADRO, M.F. - General Aspects of numerical forecast of time and climate, INPE – Brazil, 2003, available in https://www.researchgate.net/publication/256109997, access in 23/05/2019

[8]. WITT, G - Using Data from Climate Science to Teach Introductory Statistics, Journal of Statistics Education Volume 21, Number 1 (2013), available in www.amstat.org/publications/jse/v21n1/witt.pdf, access in 25/05/2019

[9]. WONNACOTT,T.H.;WONNACOTT,R.J. WONNACOTT, T.H., WONNACOTT, R.J. Econometrics. São Paulo: Technical and Scientific Books. 1983