# Variogram as a tool for assessing the quality of climate models

**Vitor Baccarin Zanetti**

**Maria Luiza Teófilo Gandini**

vitorz@gmail.com

mltgandini@gmail.com

Instituto Tecnológico de Aeronáutica

Praça Marechal Eduardo Gomes, 50 - Vila das Acacias, São José dos Campos - SP, 12228-900, Brazil

**Sin Chan Chou**

**André Lyra**

chou.sinchan@gmail.com

andrelyra1@gmail.com

CPTEC/INPE

Rodovia Dutra, km 39 Cachoeira Paulista - SP, 12630-000, Brazil

**Paulo Ivo Braga de Queiroz**

**Wilson Cabral Souza Junior**

pi@ita.br

wilson@ita.br

Instituto Tecnológico de Aeronáutica

***Abstract.*** *Climate models are very sensitive to spatial resolution. Their skill must always be verified, as they involve several phenomena which take place in different scales. For that reason, some of those phenomena must be adequately parameterized, with appropriate techniques of upscaling. The proposal of this work is to present the variogram as a tool for assessing the* *quality of climate models, based on comparison of model results with different spatial discretization. Results of the ETA Model from INPE are presented in two different levels of discretisation: for resolutions higher than 5 km, to which non-hydrostatic models must be taken into account, and for resolution lower than 8 km, to which hydrostatic models are suited. Variograms for 36 km, 18 km, 4 km, 2 km and 1 km are calculated and their results are discussed, together with other metrics for quality assessment of forecast models. Variograms showed that there is an impact of grid coarseness over these numerical models, which was less noticeable in plots of precipitations for coarser grids.*

***Keywords:*** *Variogram, Geostatistics, Climate model, Model quality assessment*

# 1 INTRODUCTION

Regardless of the discussion on causes of *Global Warming* and related economic issues, climate change has become in fact a consensus in scientific climatology (Cook et al., 2016). Hence, climatic forecasting and numeric climate models are becoming more and more important in both engineering and management sciences.

Initial conditions for climate models are usually defined by a series of procedures called *data assimilation*, which deal with conflicting data from meteorological instrumentation, in order to provide a coherent starting state. Navier-Stokes equations are main sources of calculations for atmospheric models; they are known to be a non-linear set of partial differential equations which are very sensitive to subtle changes in initial conditions. This behavior is kown as one of the the main sources of classical chaos (Kalnay, 2003). These facts make error verification to be a large concern in climate forecast; altough there are textbooks dedicated to this subject, it is still an open subject in climate models (Jolliffe & Stephenson, 2003).

Geostatiscics is not a new subject in hydrology (Bras & Rodríguez-Iturbe, 1993). Although its origin comes from ore mining (Matheron, 1965), geostatistics has found applicability on several subjects in nature sciences. The main purpose of this work is to present applications of variogram technique to assess quality and intrinsic aspects of climate models.

# 2 THE VARIOGRAM

To the point of view of engineering practice, geostatistics provides two important tools: kriging and geostatistical simulations. Kriging is the name of several techniques for data interpolation and average estimation (Armstrong, 1998). Geostatistical simulation is a series of techniques for creation of random functions that have the same geostatistical structure (Lantuéjoul, 1994).

The geostatistical structure is defined by a function called *semivariogram* or *variogram*. Let $Z(x)$ be a random funcion, where $x$ is either a real number or a vector. The variogram is defined as the following funcion of the distance $h$ between any two points:

$$\gamma(h) = \frac{1}{2N} Var\left[Z(x+h) - Z(x)\right], \tag{1}$$

where $Var$ is the statistical variance.

There are some theoretical restrictions to $Z(x)$ in order to ensure the existence of varigrams, like "intrinsic hypothesis" (Armstrong, 1998),

$$\begin{aligned} E\left[Z(x+h) - Z(x)\right] &= 0 \\ Var\left[Z(x+h) - Z(x)\right] &= 2\gamma(h) \end{aligned} \tag{2}$$

which allows existence of variograms even for "pathological cases" of $Z$, where a covariance as a function of distance $h$

$$C(h) = E\left\{\left[Z(x+h) - m\right]\left[Z(x) - m\right]\right\}, \tag{3}$$

cannot be defined, due to large variations of $Z$ at large distances, which avoids the existence of an average value $m$ for $Z(x)$.

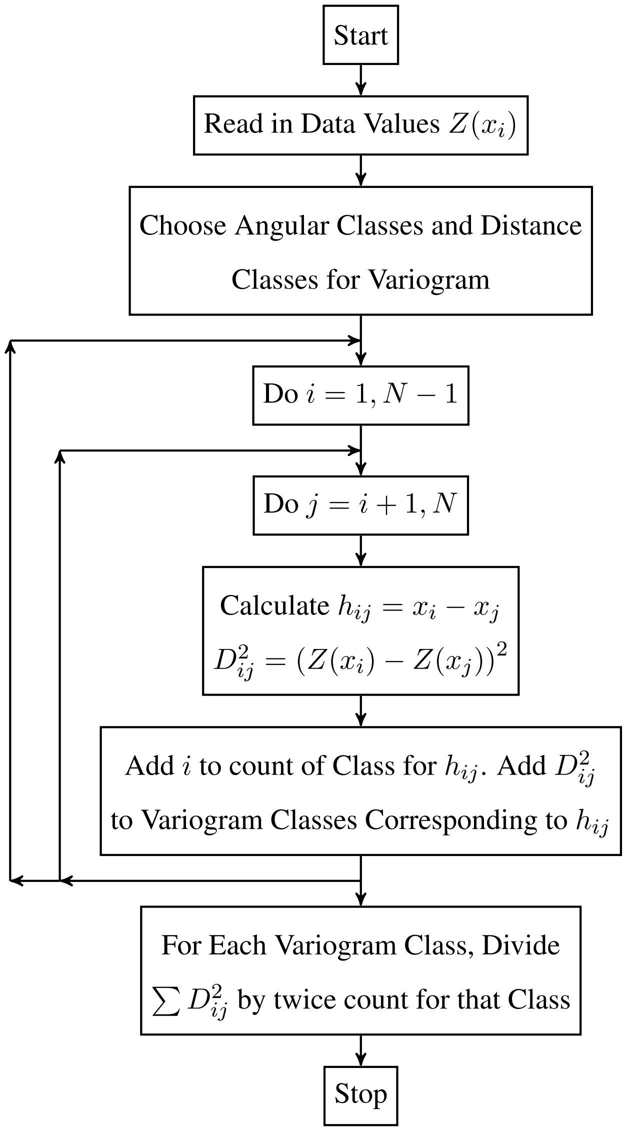



**Figure 1: Flowchart showing how to calculate experimental variograms Armstrong (1998)**

An observed set of data $Z(x_i)$, that is, any variable $Z$ sampled at $N$ locations $x_i$, may be used in order to determine an experimental variogram. Figure 1 shows a flowchart of the procedure for calculation of experimental variograms.

Experimental variograms have been used in hydrology to characterize the main types of precipitation. As an example, Lou (2004) reported attempts to characterize frontal and convective events by certain characteristics of rain variograms, like range, nugget effect and sill. That study was based in both weather radar and a raingauge network located in the Alto-Tietê, São Paulo. Table 2 presents results obtained for spherical variograms. In fact, one must bear in mind limitations of Lou's results, as this is only one case study, with limited number of raingauges. Moreover, radar estimates rely on procedures of remote sensing. Nevertheless, both radar and raingauge data showed defined geostatistical structures like a small nugget effect, finite (non-null) derivative at origin and limits to distance correlations as an estimated sill. All ranges were about 12 km, except for radar data of convective events, which resulted in a range of 6.4 km.

**Table 1: Variogram parameters obtained for precipitations at Alto-Tietê, São Paulo (modified from Lou, 2004)**

| Variogram characteristics | frontal events | | convective events | |
|---|---|---|---|---|
| | radar | gauge | radar | gauge |
| range (pixels)† | 12.4 | 12.4 | 6.4 | 12 |
| nugget effect | 0.2 | 0.058 $(mm/h)^2$ | 0.2 | 2.468 $(mm/h)^2$ |
| sill | 0.8 | 0.13 $(mm/h)^2$ | 0.8 | 4.477 $(mm/h)^2$ |

† A pixel is a 2 km × 2 km square

# 3 THE ETA MODEL

It is a known fact that nowadays global climate models (GCMs) just cannot provide climate data with adequate spatial accuracy for practical use in regional climate forecasting. This kind of problem is important in management of climate impact over urban areas and major crop production, among other real problems. Hence, regional climate models (RCMs) nested to GCMs became an important solution for regional forecasting in recent years.

According to Chou et al. (2014), Eta is a RCM that uses the $\eta$ (eta) vertical coordinate (Mesinger, 1984), which is suitable for steepy mountain areas, due to the fact that the other two coordinates stay horizontal in this scheme. Other important characteristics of Eta are:

- model dynamics in finite volume scheme (Janjić, 1984; Mesinger et al., 2012)
- deep and shallow convection parametrized by a modified Betts-Miller scheme (Betts & Miller, 1986; Janjić, 1994)
- Zhao scheme for cloud microphysics (Zhao et al., 1997)
- NOAH scheme for land-surface processes modeling (Ek et al., 2003)
- radiative transfer modeled by Lacis-Hansen scheme (Lacis & Hansen, 1974) for short waves and Fels-Schwarzkopf scheme for long waves (Fels & Schwarzkopf, 1975)

Like all RCMs, in order to produce forecastings, Eta should run having boundary conditions provided by global climate models. Chou et al. (2014) have run Eta reported and assessed runnings of Eta with the Brazilian Earth System Model version 2.3.1 BESM (Nobre et al., 2013), the Hadley Centre Global Environmental Model HadGEM2-ES (Martin et al., 2011; Cox, 2001) and with the Model for Interdisciplinary Research on Climate (MIROC), version 5 (Watanabe et al., 2010). In all these runnings, Eta was capable of reproduce trends of climatic extreme indicators, like warm nights, heat waves and hot days, although some simulations have underestimated or overestimated precipitations in certain Brazilian regions.

# 4 A CASE STUDY IN CUNHA COUNTY

In order to assess geostatistical characteristics of Eta's numerical results (nested in HadGEM2-ES), a case study was taken from Eta simulations over a region surrounding Cunha county, at Brazilian Southest region. Five runnings with different cell sizes (1 km, 2 km, 4 km, 18 km and

**Table 2: Average and variance of precipitation for the case study with different grids**

| lattice size | 1 km | 2 km | 4 km | 18 km | 36 km |
|---|---|---|---|---|---|
| average | 28.8373 | 27.7387 | 27.3179 | 39.252 | 40.1221 |
| variance | 1066.84 | 683.575 | 582.839 | 377.025 | 212.427 |

36 km lattices) were performed and precipitation at the region from 46°W 24°S to 42°W 22°S was accumulated from Jan 8$^{th}$ 2011 12:00 to Jan 12$^{th}$ 2011 12:00.

Figure 2 shows some plots of these accumulated precipitations. One can notice that the coarse meshes simulations show all the major trends of the 1 km simulation. The spots of 50-100 mm are larger at the north east side of the mesh, while the north west side looses the peak precipitation values. A certain degree of "degeneracy" of the results is in fact expected as a consequence of coarse meshes. This degeneracy is due to loss of information that occurs when average data is used to calculate nonlinear rain parametrization.

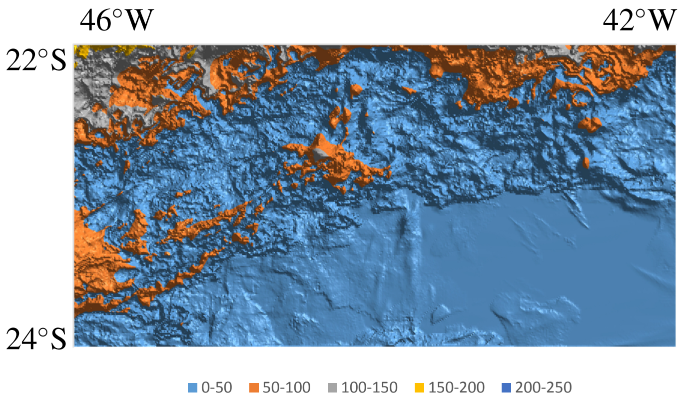
Table 2 shows the average and variance of precipitation for all points calculated. It is clear that for the first three columns (1 km, 2 km and 4 km square cells), no significant change in average is observed, while for 18 and 36 km, precipitation has changed by a large ammount.

As the first three columns show precipitation being averaged in one, four and sixteen cells of 1 km, if they were not correlated, variance was expected to fall at a ratio of 1:4 between two adjacent columns. In other extreme case, if they were completely correlated, no change in variance should be observed. In the case of Eta model, the variance behavior suggests a partial correlation as a function of distance, which evidences a geostatistical structure.

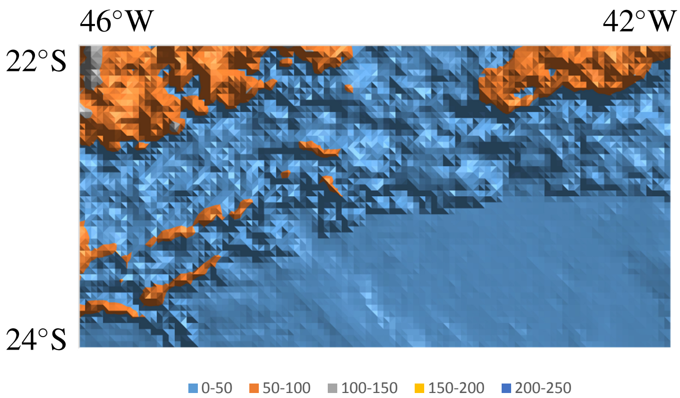
Figure 3 shows the experimental variograms for Eta simulations with different grids. For the finer meshes, it is observed variograms that are linear at origin with small nugget effect, which is in accordance to the bibliography (Lou, 2004). Nevertheless, no sill was observed in any numerical simulations, which preserved a linear trend to more than 200 km. This in fact avoids the use of the covariance concept for precipitation in Eta's numerical simulations.

The variogram also testify any degree "degeneracy" of numerical results for coarser grids as, according to Armstrong (1998), regularized linear variograms should theoretically maintain the same slope (derivative) for distances much larger than the new support (area in which the average of any random function is performed). Nevertheless, regularization due to change of support should lower the entire variogram, causing nugget effect to disapear or even to be negative, for variograms taken at distances larger than support. In fact, this negative nugget effect is observed in the experimental variogram for 36 km grid.

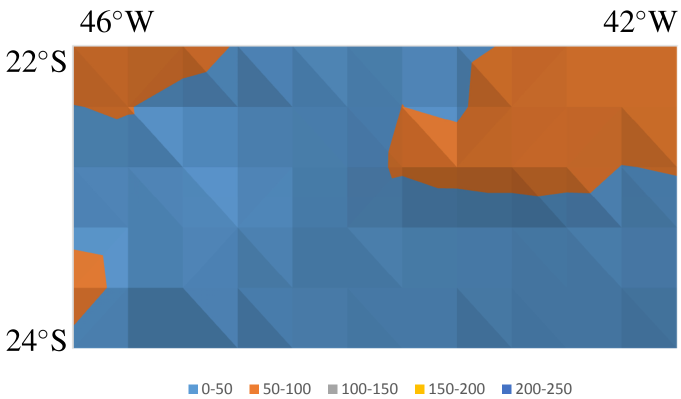
As reference for comparison, Figure 4 shows the experimental variograms to which support was changed by average calculations over the Eta simulations with 1 km square grids. One can notice that all experimental variograms have almost the same slope, as predicted by theory. Figure 5 shows a comparison between results of Eta simulation with a 18 km grid to the simulation with 1 km grid, after a support change for 20 km. Although extremes are sharper for the simulation with change of support, it is easy to notice that most precipitation magnitudes are similar and most regions of intense precipitation are preserved, indicating a certain robustness of the cloud parametrization scheme, which is used for cells larger than 5 km.




**(a) 1 km grid**




**(b) 4 km grid**




**(c) 36 km grid**

**Figure 2: Precipitation taken from Eta nested in HadGEM2-ES, from Jan 8$^{th}$ 2011 12:00 to Jan 12$^{th}$ 2011 12:00**

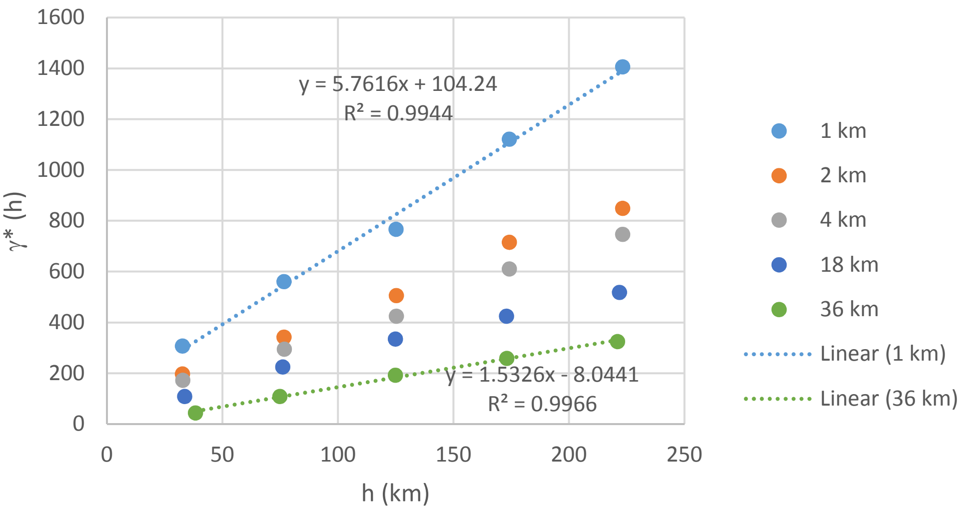



**Figure 3: Experimental variograms for Eta simulations with different grids**

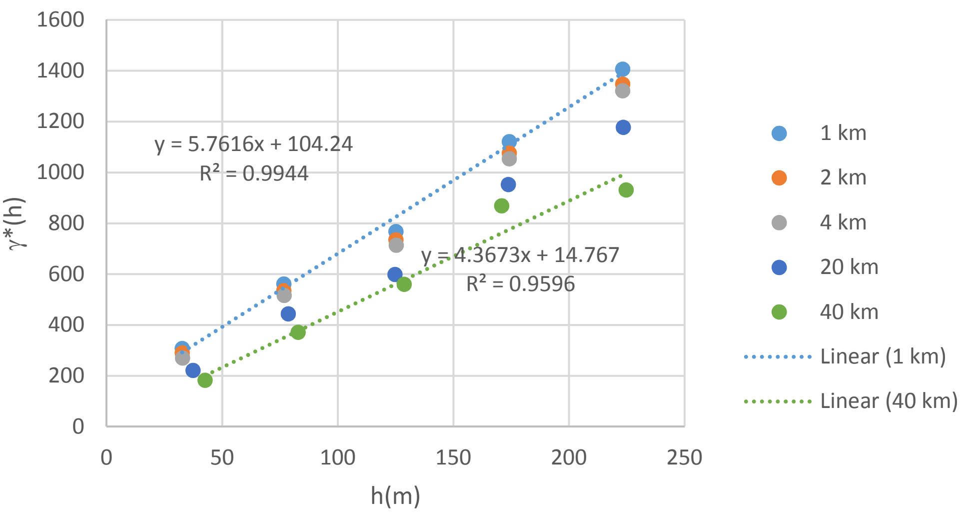
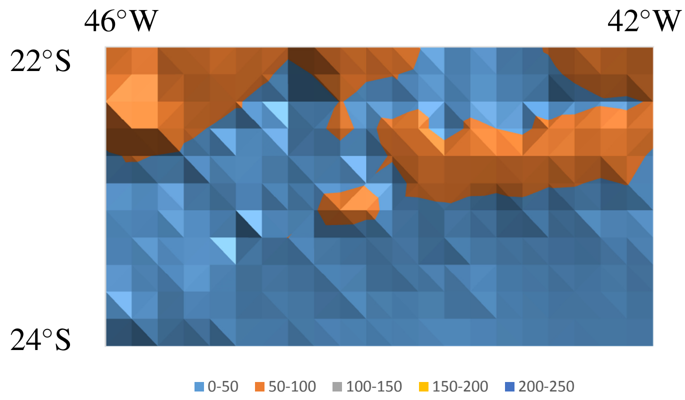
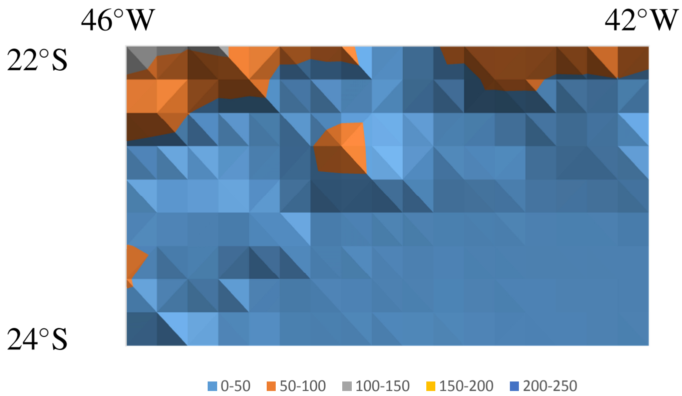



**Figure 4: Experimental variograms for Eta simulations with 1 km grid with change of support**




**(a) 18 km grid (original grid of ETA simulation)**




**(b) Eta simlulation with1 km grid, with change of support (average) for 20 km**

**Figure 5: Precipitation taken from Eta nested in HadGEM2-ES, from Jan 8$^{th}$ 2011 12:00 to Jan 12$^{th}$ 2011 12:00**

## 5 FINAL REMARKS

This work presented an assessment of the sensitivity of Eta model (Chou et al., 2014) to grid refinement. Precipitation data from five simulations with different degrees of mesh refinement was analysed, in order to evaluate correlation at large distances. Variograms showed that there is a in impact of grid coarseness over variograms, which was less noticeable in plots of precipitations for coarser grids. Future works may evaluate specific aspects of Eta model, like different parametrization schemes and cross-variograms (Wackernagel, 2003) with other variables, like wind speed or temperature.

## REFERENCES

Armstrong, M., 1998, *Basic Linear Geostatistics*, Springer Verlag, Berlin.

Betts, A. K. & Miller, M. J., 1986, A new convective adjustment scheme. part ii: Single column tests using GATE wave, BOMEX, ATEX and arctic air-mass data sets, *Quarterly Journal of the Royal Meteorological Society*, vol. 112, pp. 693–709.

Bras, R. L. & Rodríguez-Iturbe, I., 1993, *Random Functions and Hydrology*, Dover Publications, New York.

Chou, S. C., Lyra, A., Mourão, C., Dereczynski, C., Pilotto, I., J. Gomes, J. B., Tavares, P., Silva, A., Rodrigues, D., Campos, D., Chagas, D., Sueiro, G., Siqueira, G., Nobre, P. & Marengo, J., 2014, Evaluation of the eta simulations nested in three global climate models, *American Journal of Climate Change*, vol. 3, n. 5, pp. 438–454.

Cook, J., Oreskes, N., Doran, P. T., Anderegg, W. R. L., Verheggen, B., Maibach, E. W., Carlton, J. S., Lewandowsky, S., Skuce, A. G., Green, S. A., Nuccitelli, D., Jacobs, P., Richardson, M., Winkler, B., Painting, R. & Rice, K., 2016, Consensus on consensus: a synthesis of consensus estimates on human-caused global warming, *Environmental Research Letters*, vol. 11, n. 4, pp. 048002.
**URL:** *http://stacks.iop.org/1748-9326/11/i=4/a=048002*

Cox, P. M., 2001, Description of the “TRIFFID” dynamic global vegetation model. Hadley Centre Technical Note 24.

Ek, M., Mitchell, K. E., Lin, Y., Rogers, E., Grunmann, P., Koren, V., Gayno, G. & Tarpley, J. D., 2003, Implementation of Noah land surface model advances in the national centers for environmental prediction operational mesoscale eta model, *Journal of Geophysical Research*, vol. 108, pp. 1–16.

Fels, S. B. & Schwarzkopf, M. D., 1975, The simplified exchange approximation: A new method for radiative transfer calculations, *Journal of the Atmospheric Sciences*, vol. 32, pp. 1475–1488.

Janjić, Z. I., 1984, Nonlinear advection schemes and energy cascade on semi-staggered grids, *Monthly Weather Review*, vol. 112, pp. 1234–1245.

Janjić, Z. I., 1994, The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes, *Monthly Weather Review*, vol. 122, pp. 927–945.

Jolliffe, I. T. & Stephenson, D. B., 2003, *Forecast verification: a practitioner's guide in atmospheric science*, John Wiley & Sons, Chichester, UK.

Kalnay, E., 2003, *Atmospheric Modeling, Data Assimilation and Predictability*, Cambridge University Press, Cambridge, UK.

Lacis, A. A. & Hansen, J., 1974, A parameterization for the absorption of solar radiation in the earth's atmosphere, *Journal of the Atmospheric Sciences*, vol. 31, pp. 118–133.

Lantuéjoul, C., 1994, Nonconditional simulation of stationary isotropic multigaussian random functions, *in* M. Armstrong & P. Dowd (eds), *Geostatistical Simulations*, Kluwer Academic Publishers, Dordrecht, pp. 147–177.

Lou, A. P. F., 2004, *Modelagem geoestatística aplicada à integração entre dados de postos pluviométricos e radar meteorológico*, Master's thesis, COPPE-UFRJ, Rio de Janeiro.

Martin, G. M., Bellouin, N., Collins, W. J., Culverwell, I. D., Halloran, P. R., Hardiman, S. C., Hinton, T. J., Jones, C. D., McDonald, R. E., McLaren, A. J., O'Connor, F. M., Roberts, M. J., Rodriguez, J. M., Woodward, S., Best, M. J., Brooks, M. E., Brown, A. R., Butchart, N., Dearden, C., Derbyshire, S. H., Dharssi, I., Doutriaux-Boucher, M., Edwards, J. M., Falloon, P. D., Gedney, N., Gray, L. J., Hewitt, H. T., Hobson, M., Huddleston, M. R., Hughes, J., Ineson, S., Ingram, W. J., James, P. M., Johns, T. C., Johnson, C. E., Jones, A., Jones, C. P., Joshi, M. M., Keen, A. B., Liddicoat, S., Lock, A. P., Maidens, A. V., Manners, J. C., Milton, S. F., Rae, J. G. L., Ridley, J. K., Sellar, A., Senior, C. A., Totterdell, I. J., Verhoef, A., Vidale, P. L., & Wiltshire, A., 2011, The HadGEM2 family of met office unified model climate configurations. geoscientific model development, , vol. 4, pp. 723–757.

Matheron, G., 1965, *Les variables régionalisées et leur estimation: une application de la théorie des fonctions aléatoires aux sciences de la nature*, Masson, Paris.

Mesinger, F., 1984, A blocking technique for representation of mountains in atmospheric models, *Rivista di Meteorologia Aeronautica*, vol. 44, pp. 195–202.

Mesinger, F., Chou, S. C., Gomes, J. L., Jovic, D., Bastos, P., Bustamante, J. F., Lazic, L., Lyra, A. A., Morelli, S., Ristic, I. & Veljovic, K., 2012, An upgraded version of the eta model, *Meteorology and Atmospheric Physics*, vol. 116, pp. 63–79.

Nobre, P., Siqueira, L. S. P., de Almeida, R. A. F., Malagutti, M., Giarolla, E., Castelão, G. P., Bottino, . J., Kubota, P., Figueroa, S. N., Costa, M. C., Baptista, M., Irber, L. & Marcondes, G. G., 2013, Climate simulation and change in the brazilian climate model, *Journal of Climate*, vol. 26, pp. 6716–6732.

Wackernagel, H., 2003, *Multivariate Geostatistics – An Introduction with Applications*, 3$^{rd}$ edn, Springer Verlag, Berlin.

Watanabe, M., Suzuki, T., O'ishi, R., Komuro, Y., Watanabe, S., Emori, S., Takemura, T., Chikira, M., Ogura, T., Sekiguchi, M., Takata, K., Yamazaki, D., Yokohata, T., Nozawa, T., Hasumi, H., Tatebe, H. & Kimoto, M., 2010, Improved climate simulation by MIROC5: Mean states, variability, and climate sensitivity, *Journal of Climate*, vol. 23, pp. 6312–6335.

Zhao, Q., Black, T. L. & Baldwin, M. E., 1997, Implementation of the cloud prediction scheme in the eta model at NCEP, *Weather and Forecasting*, vol. 12, pp. 697–712.