A method for describing the uncertainty in geophysical models

Riccardo Barzaghi

DICA – Politecnico di Milano
E-Mail: riccardo.barzaghi@polimi.it

Abstract

Geophysical models are more and more complex and can describe, in a very careful way, the geodynamical processes. These models depend on several geometrical and physical parameters, which characterize the models behaviour. These parameters are not errorless since they are known with some uncertainties. Thus, model predictions are affected by these parameter uncertainties. Usually, the model errors are not taken into account and are not propagated to the estimated quantities. This can cause improper hypothesis testing when geodetic data and model predictions are compared. In this paper, a method is presented which allows considering the model errors through the definition of the spatial covariance function of the model. A simulation is set up to prove the method feasibility.

1 Introduction

The availability of precise observations is nowadays rapidly increasing. This demands for a more careful comparison between models (geometrical, physical and so on) and observations in order to have a better validation of these models. In geodesy, GNSS precise measurements are an invaluable tool for improving the geophysical models which are used to analyse the geodynamic of the crust at different spatial and time scales. The present day distribution of permanent GNSS stations allows defining in some details the actual crustal deformation either at global and continental level (Drewes and Heidbach, 2012). Also, in particular cases, these analyses can be performed at the level of a single fault (or of a fault system) (Riva et al., 2007). These observations, further implemented with non-permanent GNSS campaigns, give precise estimates of the velocities of the observed stations, based on a daily coordinates repeatability of $1-2$ mm. Similarly, considering another geodetic example, radaraltimetry data can estimate the sea surface heights with respect to the ellipsoid at centimetre level precision. These observations, coupled with a geoid estimate of the same precision, allow defining the Dynamic Ocean Topography (DOT), which is functionally related to the geostrophic currents (Rummel, 1993). Also in this case, the comparison between this kind of data and the oceanographic circulation models allows their refinements.

Commonly, the comparison between data and model predictions is performed considering the observation error of the data without considering any model error. These comparisons are usually carried out using the Chi-square random variable (Mood et al., 1983) in the hypothesis that the discrepancies between observed values and model predictions divided for their standard deviations are independent standard normal random variables.

The observed values x_{obs} are supposed to be a sample from a normal random variable having mean value x_{mod}. Thus, one can write

$$z_{obs} = \frac{x_{obs} - x_{mod}}{\sigma_{obs}}, \qquad (1.1)$$
$$(\sigma_{obs} = \text{standard deviation of } x_{obs})$$

Erschienen bei KIT Scientific Publishing
Schriftenreihe des Studiengangs
Geodäsie und Geoinformatik 2018,1

DOI Einzelbeitrag:
10.5445/KSP/1000080207

Festschrift zur Verabschiedung von
Prof. Dr.-Ing. Dr. h.c. Bernhard Heck
(Schw)Ehre, wem (Schw)Ehre gebührt

DOI Festschrift:
10.5445/KSP/1000080324

where z is a standard normal random variable. Since we usually consider more than one observation, we can define the Chi-square random variable as

$$\chi^2_n = \sum_{k=1}^{n} \frac{\left(x^k_{obs} - x^k_{mod}\right)^2}{\sigma^2_{obs}}, \quad (n = \text{number of observations}) \tag{1.2}$$

which can be used to test the coherence between observed values and model predictions.

Following this approach, no error model is considered. In case the model error can be estimated, the more correct Chi-square quantity could be defined as

$$\chi^2 = (x_{obs} - x_{mod})^t \left(C_{obs} + C_{mod}\right)^{-1} (x_{obs} - x_{mod}), \tag{1.3}$$

which implies that by hypothesis we assume that the model predictions and the observations have the same mean. C_{obs} can be straightforwardly derived from least squares, which are usually applied for adjusting the data. On the other hand, the C_{mod} cannot be easily estimated.

In case the model outcomes x_{mod} are assumed to be linearly depended by a set of parameters x_{par}, i. e., it holds that

$$x_{mod} = A x_{par} \tag{1.4}$$

one can apply the covariance propagation law (Sansò, 2000) which gives C_{mod} as

$$C_{mod} = A\, C_{par} A^t. \tag{1.5}$$

The C_{par} matrix can be estimated based on, e. g., physical information on the model parameters.

As an example, if one is considering a geophysical model which depends on parameters such as crustal density and viscoelasticity, suitable mean and range values of these quantities can be defined, based on geophysical assumptions. This will allow a proper definition of the C_{par}. However, in many cases, this cannot be done. In most of the cases, the model is not described in the explicit form (1.4), being it a multistep complex procedure. Thus, for most of the models, the direct formula (1.5) cannot be applied. An alternative way to estimate C_{mod} is to define the covariance function of x_{mod}, which in turn can be used to compute C_{mod}. Furthermore, by following this approach, one can derive also information on the covariance structure of the model signal, such as its spatial correlation. This information can be used to design in a proper way geodetic monitoring networks (as an example, one can consider to compare a geophysical model predicting crustal deformations in a geodynamical active area and the benchmark distribution of a GNSS permanent network designed for monitoring this physical phenomenon). In the next paragraph, the numerical procedure, which allows estimating the model covariance function based on assumptions on the model parameters variability, will be described.

2 The model covariance function estimate

We assume that a linear (or linearized) relationship holds between the model signal and the model parameters

$$x_{mod} = L(\xi, x_{par}), \tag{2.1}$$

where ξ are other possible parameters defining the model, such as the point position where to estimate the model value x_{mod}: these parameters will not be considered in the error propagation.

We further assume that x_{par} is a normal random variable having m independent components. Based on some feasible assumptions, the values of the mean μ_{par} and of the variances $\left(\sigma^k_{par}\right)^2$ $(k = 1, \ldots, m)$ of the parameters can be defined. By sampling the x^0_{par} values by means of a random number generator, the corresponding x^0_{mod} values can be obtained and then used to estimate the covariance function of x_{mod}.

In order to give an example of the devised procedure, let us assume that x_{mod} is a two components signal in the plane. Based on r field estimates computed following this approach, the two auto-covariances and the cross-covariance between the two components of x_{mod} can be derived as

$$C_{lk}(d_n) = \frac{1}{r} \sum_{s=1}^{r} \left[\frac{1}{N} \sum_{i=1}^{N} \delta(x_{mod})^i_{l,r} \frac{1}{N_j} \sum_{j=1}^{N_j} \delta(x_{mod})^j_{k,r} \right] \tag{2.2}$$

with

$$\delta(x_{mod})^i_{k,r} = \left(x^0_{mod}\right)^i_{k,r} - (\bar{x}_{mod})^i_k, \tag{2.3}$$

$$(\bar{x}_{mod})^i_k = \frac{1}{r} \sum_{s=1}^{r} \left(x^0_{mod}\right)^i_{k,s}. \tag{2.4}$$

The index $i = 1, \ldots, N$ runs on the points P_i in the plane where the signal is estimated, the index $j = 1, \ldots, N_j$ runs on all the N_j points P_j having distance d_{ij} from P_i such that $d_n \leq d_{ij} < d_n + \Delta d$ with a proper given value Δd and (l,k) label the two components of x_{mod} ($l = 1,2; k = 1,2$).

In case $l = k$, the auto-covariances of the two model components are estimated, while for $l \neq k$ the cross-covariances between the two components are derived.

These empirical estimates must be then interpolated with suitable model covariances, i. e., with positive definite functions (Barzaghi and Sansò, 1984).

The auto-covariance and the cross-covariance functions of x_{mod} can be in turn used for estimating the C_{mod} that can be used in the testing the model predictions versus the observed data.

The devised procedure is of Bayesian kind since the parameters x_{par} are considered as random variables. Furthermore, assuming that x_{par} is normally distributed allows defining in a proper way the formula (1.3), if we assume that also x_{obs} is normally distributed and independent from x_{mod}. In this hypothesis, being x_{mod} linearly dependent on x_{par}, one can prove that x_{mod} is normally distributed too (Mood et al., 1983), with $\mu_{mod} = L(\xi, \mu_{par})$ and covariance C_{mod}.

Finally, it has to be mentioned that this procedure, developed here for a two-dimensional process, can be generalized to multi-dimensional process.

3 A simulated test

In order to test the feasibility of the proposed approach, a simulation has been devised. Starting from a given covariance function, a covariance matrix on a given set of points in the plane and the signal on these points having this covariance structure can be computed (Barzaghi et al., 1992).

This procedure has been carried out starting from the following covariance function

$$C(d) = A J_0(\alpha d)\,, \tag{3.1}$$

where J_0 is the zero order Bessel function (Watson, 1948) and d is the standard Cartesian distance in the plane.

The values $A = 50$ and $\alpha = 0,15$ km^{-1} have been set and the related covariance matrix C has been computed over points in the plane placed on a regular square grid having a grid step of 5 km and linear dimension of 100 km (the covariance matrix has thus dimension $n = 400$).

The simulated signal in the plane is then estimated according to the formula

$$s(P) = T^t \upsilon \tag{3.2}$$

with

$$C = T^t T \tag{3.3}$$

following the Cholesky decomposition method (Benciolini and Mussio, 1984). υ is a sample from a random variable having

$$E(\upsilon) = 0\,, \qquad C_{\upsilon\upsilon} = I\,. \tag{3.4}$$

As it can be easily proved, the covariance matrix of the simulated signal (3.2) is exactly equal to C since

$$\begin{aligned} C_{ss} = E\left[ss^t\right] = E\left[\left(T^t\upsilon\right)\left(T^t\upsilon\right)^t\right] = \\ = T^t E\left[\upsilon\upsilon^t\right] T = T^t C_{\upsilon\upsilon} T = T^t T = C\,. \end{aligned} \tag{3.5}$$

This procedure has been repeated 100 times, based on different υ samples, so that 100 different realization of the signal s have been computed. Given the simulated values, the formula (2.2) has been applied: in this case $r = 100$ and $l = k = 1$.

The values of the empirical covariance and the model covariance (3.1), using the values of A and α previously defined, are presented in Figure 3.1.

As one can see, the estimated empirical values are remarkably close to the model function. Thus, this simulation proves that formula (2.2) can give a correct estimate of the signal covariance.

Furthermore, in the presented simulation, every single realization has, by definition, the given covariance structure that can be estimated according to the approach described in Mussio (1984). In Figure 3.2, the empirical covariances estimated with the signals s_{10}, s_{50} and s_{90}, selected among the 100 computed simulations of the signal, and the model (3.1) are plotted.

Contrary to the previous case, the empirical covariance values are not so close to the model covariance, particularly for large d values. This is quite obvious since formula (2.2) is the mean of the empirical co-

variances of the different signal realizations. Thus, the values in (2.2) are more reliable than those computed on a single realization.

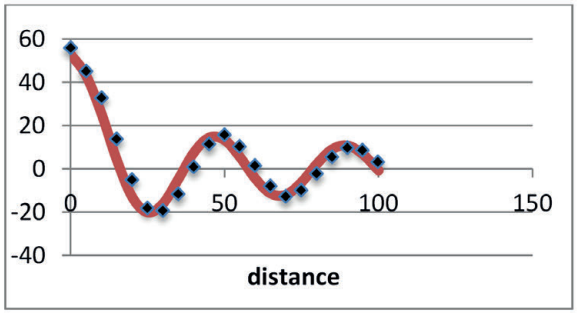

Figure 3.1: The empirical covariance (formula (2.2) - black dots) and the model covariance (solid red line).

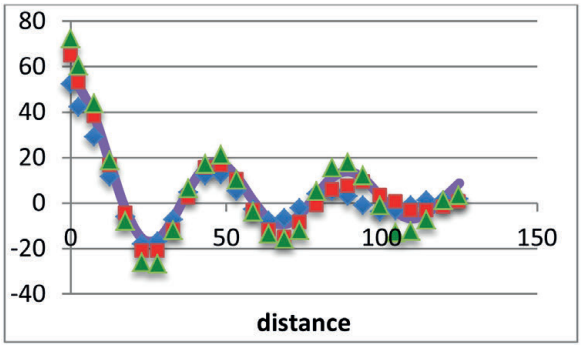

Figure 3.2: The empirical covariances of the three signals and the model function (solid line).

4 Conclusions

The proposed procedure can be efficiently applied for the estimation of the covariance function of the signal implied by an arbitrary model, which depends on a given set of parameters. If each parameter is considered as a normal random variable, having mean and variance that can be derived from physical assumptions on the model, one can estimate the covariance function of the signal derived from the model. In turn, this allows having a testing procedure between model and observed values that takes into account not only the covariance of the observation but also the model covariance. In a recent application of this method to a geophysical model describing the crustal deformation in the Calabrian Arc region, interesting results were obtained that allowed a clearer definition of this geodynamical process (Barzaghi et al., 2014). In the future, the same procedure will be applied to other geodynamical areas in the Italian region and to oceanographic circulation models in the comparison with altimetry data.

References

Barzaghi, R., Gandino, A., Sansò, F., and Zenucchini, C. (1992): The collocation approach to the inversion of gravity data. *Geophysical Prospecting* 40(4):429–451. DOI: 10.1111/j.1365-2478.1992.tb00535.x.

Barzaghi, R., Marotta, A. M., Splendore, R., De Gaetani, C., and Borghi, A. (2014): Statistical assessment of predictive modelling uncertainty: a geophysical case study. *Geophysical Journal International* 197(1):22–32. DOI: 10.1093/gji/ggt510.

Barzaghi, R. and Sansò, F. (1984): La collocazione in geodesia fisica. *Bollettino di Geodesia e Scienze Affini* 43(2):117–152.

Benciolini, B. and Mussio, L. (1984): Algoritmi "esatti" per la soluzione di grandi sistemi. Ricerche di Geodesia, Topografia e Fotogrammetria, vol. 4. CLUP, Milano.

Drewes, H. and Heidbach, O. (2012): The 2009 Horizontal Velocity Field for South America and the Caribbean. In: *Geodesy for Planet Earth: Proceedings of the 2009 IAG Symposium, Buenos Aires, Argentina, 31 August 31 - 4 September 2009*. Ed. by S. Kenyon, M. C. Pacino, and U. Marti. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 657–664. DOI: 10.1007/978-3-642-20338-1_81.

Mood, A. M., Graybill Franklin, A., and Boes, D. C. (1983): Introduction to the theory of statistics. McGraw-Hill: series in probability and statistics, p. 577. ISBN: 0-07-042864-6.

Mussio, L. (1984): Il metodo della collocazione minimi quadrati e le sue applicazioni per l'analisi statistica dei risultati delle compensazioni. Ricerche di Geodesia, Topografia e Fotogrammetria, vol. 4. CLUP, Milano.

Riva, R. E. M., Borghi, A., Aoudia, A., Barzaghi, R., Sabadini, R., and Panza, G. F. (2007): Viscoelastic relaxation and long-lasting after-slip following the 1997 Umbria-Marche (Central Italy) earthquakes. *Geophysical Journal International* 169(2):534–546. DOI: 10.1111/j.1365-246X.2007.03315.x.

Rummel, R. (1993): Principle of Satellite Altimetry and Elimination of Radial Orbit Errors. In: *Satellite Altimetry in Geodesy and Oceanography (Lecture Notes in Earth Sciences)*. Ed. by R. Rummel and F. Sansò. Vol. 50. Berlin, Heidelberg: Springer-Verlag, pp. 190–241. DOI: 10.1007/BFb0117924.

Sansò, F. (2000): Elementi di teoria della probabilità. Quaderni di trattamento statistico dei dati, Città Studi Edizioni, Milano.

Watson, G. N. (1948): A treatise on the theory of Bessel functions. Cambridge University Press.