

Geodatabase updating by new measurements; a Bayesian intermezzo

Maria Antonia Brovelli · Fernando Sansò

Received: 12 October 2008 / Accepted: 23 March 2009 / Published online: 6 May 2009
© Società Italiana di Fotogrammetria e Topografia (SIFET) 2009

Abstract The question the paper deals with refers to how it is possible to update existing geodatabases considering both their accuracies and those of the new measurements taken for their updating. Traditionally, maintaining geodatabases (or map bases) has been highly time consuming, costly, and sometimes difficult work, especially in urban and high-density areas. The most common procedure is to globally generate geodatabases every few years by photogrammetric techniques. On the opposite, the possibility of dynamically updating the landscape information from a maintained core spatial database can be considered as an appealing alternative to traditional map revision techniques. A kriging solution, based on the hypothesis that the vector field of the position error on a geodatabase is a homogeneous, isotropic intrinsic random field with constant mean and variogram depending only on the squared distance, known a priori from the relative accuracy of the map, is proposed. The method is a first approach to the problem, as far as at the moment it does not consider constraints to which points on the geodatabase must adapt to. That is the reason why it is presented as an intermezzo.

Keywords Geodatabase · Updating · Bayesian · Kriging

Position of the problem

Spatial databases can be nowadays easily managed by means of geographic information system tools (Rigaux

et al. 2002). The hot point in the first era of the digital revolution was mainly to collect pieces of information. On the opposite now the great challenge relates to provide geographic data of high quality level (and therefore the estimation of such a kind of information) and to make available methods to exploit as much as possible the already available information, correctly modeling the uncertainty intrinsic into spatial data, instead of remaking maps every time from the beginning. Thus data sharing, data conflating (Brovelli and Zamboni 2004, 2006), and updating (Arnold and Wright 2005) must be supported by rigorous approaches having a sound statistical basis. Efforts have been made in this regard by some researchers (Leung et al. 2004). The paper places in this frame, trying to solve the problem at least for some elementary cases.

We consider a geodatabase basically as a collection of points $\{P_i; i=1,2,\dots,N\}$ of known planar coordinates, $\{(x_i, y_i)\}$, topological information, allowing to identify linear or area features and other numerical or thematic information concerning several attributes related to these points and objects. Here, we will be concerned only with the first aspect, namely the set of points $\{P_i\}$ and their coordinates. The question we want to handle is the following: imagine we perform new geodetic measurements involving one part of the points $\{P_i\}$, how do we update their coordinates in view of such an information?

As an example, think that at some points P_i we have done observations with a GPS receiver, connected to a positioning service in a RTK mode, as it is quite usual nowadays.

Of course if it was available the full covariance matrix of the vector of coordinates $\{(x_i, y_i)\}$ and that of the new measurements, the update could be performed rigorously, at least in terms of least squares theory.

This is however not the practical situation and for two reasons. The former is that when the number of points is of

M. A. Brovelli (✉) · F. Sansò
Politecnico di Milano—DIAR,
Polo Regionale di Como, via Valleggio,
11-22100 Como, Italy
e-mail: maria.brovelli@polimi.it

F. Sansò
e-mail: fernando.sanso@polimi.it

the order of 10^4 or more its covariance information is of the order of 10^8 real numbers or more and storing and using such a huge quantity of information is not the easiest thing to do.

The latter reason is even more important and it is that, in practice, this information is usually not available because the production of such coordinates is the result of many steps in which noise propagation is not possible or was not performed. On the opposite logical side there would be the practice of moving, namely updating, only the points which have been involved into the new measurements. This immediately leads to unacceptable results, because it would modify the relative position of objects in an incompatible way, for instance moving a building into another one or in the middle of the road. It is obvious that it is simply impossible to ignore that there is a relation between points so that once one is shifted by new observations, also the others should follow. The question is that part of the points should move like rigid bodies, verifying metric constraints as lengths of sides, rectangular shapes, etc., while in general all points should have a kind of “elastic” relation to one another.

In order to cope with such problems, the authors have already invoked a Bayesian concept (Belussi et al. 2006), introducing first of all geometrical constraints into the prior distribution of the coordinate vector and binding all the points, one to the other, by means of a simple network of pseudo-observations.

The idea bears some merits, in particular by showing that Bayesian statistics is a natural tool to be applied to an updating problem. Yet, the prior distribution reflecting the true information available on the geodatabase is still a problem. Specifically, the proposal of connecting all the points on the geodatabase by a plane pseudo-traverse is questionable in that the accuracy of each individual point will then depend on the quite arbitrary design of the traverse itself.

In this paper, we give precisely a solution to this problem when the prior information is assumed to consist barely of a relative accuracy prescription.

With “relative accuracy”, here we mean that given *any* two points P , Q on the geodatabase, the errors in their coordinates are such that

$$\frac{\sigma(|\underline{\mathbf{r}}_{PQ}|)}{|\underline{\mathbf{r}}_{PQ}|} = K \quad (1.1)$$

with K a given constant, $\underline{\mathbf{r}}_{PQ}$ the planar base vector between P and Q , and σ the r.m.s. of its modulus.

It turns out that (1.1) defines an essential feature of the error field $\underline{\mathbf{u}}(P) = (\delta x(P), \delta y(P))$ of the geodatabase, namely the variogram. In this way, the well-known kriging theory (Wackernagel 2003) becomes available for the

solution of the present problem. So we find a clear method to update the coordinates of all points, without considering them as possibly belonging to rigid bodies, i.e., without any further constraint.

This is the reason why we have entitled the paper “Bayesian intermezzo”; because the full solution under the Bayesian concept still has to be studied.

The relative accuracy of the geodatabase defines the variogram of the error random field

In this paragraph, we set up the prior stochastic model of the geodatabase on the basis of our hypotheses, namely the knowledge of the relative accuracy translated into formula (1.1).

So we assume that any identifiable point P on the geodatabase, with coordinates

$$\underline{\mathbf{R}}_P = \begin{pmatrix} X_P \\ Y_P \end{pmatrix}, \quad (2.1)$$

has a random position, due to previous noise propagation, with average

$$E\{\underline{\mathbf{R}}_P\} = \underline{\mathbf{r}}_P = \begin{pmatrix} x_P \\ y_P \end{pmatrix} \quad (2.2)$$

and the difference between the two, the position error vector $\underline{\mathbf{u}}(P)$, is a random field with zero average and finite variance.

Since the theory we are going to develop is fully translation invariant, the hypothesis $E\{\underline{\mathbf{u}}(P)\} = 0$, can be substituted by the much weaker

$$E\{\underline{\mathbf{u}}(P)\} = \underline{\mathbf{t}}, \quad (2.3)$$

a constant translation vector valid for the whole geodatabase.

This hypothesis is often useful if we have to compare geodatabase coordinates with GPS coordinates which might be given in a different geodetic datum.

Let us agree on the hypothesis that the vector random field $\underline{\mathbf{u}}(P)$ is intrinsically homogeneous and isotropic (cf. Matheron 1970).

Notice that this implies an extension of the usual kriging concept in the sense that we assume

$$\frac{1}{2} E\{|\underline{\mathbf{u}}(P) - \underline{\mathbf{u}}(Q)|^2\} = \gamma(d_{PQ}) \quad (2.4)$$

where d_{PQ} is the Euclidean distance between P and Q and *in addition* that the component of $\underline{\mathbf{u}}(P)$ along any fixed direction is again an intrinsic process such that

$$\frac{1}{2} E\{[\underline{\mathbf{e}} \cdot \underline{\mathbf{u}}(P) - \underline{\mathbf{e}} \cdot \underline{\mathbf{u}}(Q)]^2\} = f(d_{PQ}) \quad (2.5)$$

with f independent of the direction of the arbitrary unit vector $\underline{\mathbf{e}}$.

This implies that

$$f(d_{PQ}) = \frac{1}{2}E\{[\delta x_P - \delta x_Q]^2\} = \frac{1}{2}E\{[\delta y_P - \delta y_Q]^2\}, \tag{2.6}$$

and also

$$f(d_{PQ}) = \frac{1}{2}\gamma(d_{PQ}), \tag{2.7}$$

because the sum of the second and third term in (2.6) has to be $\gamma(d_{PQ})$.

$\forall \alpha$

$$\begin{aligned} f(d_{PQ}) &= E\{[\delta u(P, \alpha) - \delta u(Q, \alpha)]^2\} = \\ &= E\{[\delta x_P - \delta x_Q]^2 \cos^2 \alpha + [\delta y_P - \delta y_Q]^2 \sin^2 \alpha + 2(\delta x_P - \delta x_Q)(\delta y_P - \delta y_Q) \sin \alpha \cos \alpha\} \\ &= f(d_{PQ}) + 2 \sin \alpha \cos \alpha E\{(\delta x_P - \delta x_Q)(\delta y_P - \delta y_Q)\} \end{aligned}$$

and this implies (2.8).

We call $\gamma(d)$ the variogram of the field \underline{u} and we note that each component of \underline{u} has a variogram which is just a half of γ . Under these hypotheses, let us consider the random vector

$$\underline{R}_P = \underline{r}_P + \underline{u}(P). \tag{2.9}$$

Indeed we have

$$E\{|\underline{R}_P - \underline{R}_Q|^2\} = |\underline{r}_P - \underline{r}_Q|^2 + 2\gamma(d_{PQ}). \tag{2.10}$$

Now we need to compute $E\{|\underline{R}_P - \underline{R}_Q|\}$. We do that with an approximation up to the second order in $|\underline{u}|$.

Observe that for any finite vector \underline{r} and a small enough vector $\underline{\varepsilon}$ we have (denoting $|\underline{r}| = r$)

$$\begin{aligned} |\underline{r} + \underline{\varepsilon}| &= \sqrt{r^2 + 2\underline{r} \cdot \underline{\varepsilon} + \varepsilon^2} = r + \frac{\underline{r}}{r} \cdot \underline{\varepsilon} \\ &+ \frac{1}{2} \left[\frac{1}{r} \varepsilon^2 - \frac{1}{r} \left(\frac{\underline{r}}{r} \cdot \underline{\varepsilon} \right)^2 \right] + O(\varepsilon^3), \end{aligned} \tag{2.11}$$

We recall that $E\{\underline{u}_P - \underline{u}_Q\} = 0$ by hypothesis and use (2.7), to find, neglecting third order terms in $\underline{\varepsilon} = \underline{u}_P - \underline{u}_Q$,

$$\begin{aligned} E\{|\underline{R}_P - \underline{R}_Q|\} &= E\{|\underline{r}_P - \underline{r}_Q + \underline{u}_P - \underline{u}_Q|\} \\ &\cong |\underline{r}_P - \underline{r}_Q| + \frac{1}{2} \frac{1}{|\underline{r}_P - \underline{r}_Q|} [2\gamma(d_{PQ}) - \gamma(d_{PQ})] \\ &= d_{PQ} + \frac{1}{2} \frac{1}{d_{PQ}} \gamma(d_{PQ}) \end{aligned} \tag{2.12}$$

In addition it turns out that

$$E\{[\delta x_P - \delta x_Q][\delta y_P - \delta y_Q]\} = 0. \tag{2.8}$$

This last relation is easily derived by imposing that the field

$$\delta x_P \cos \alpha + \delta y_P \sin \alpha = \delta u(P, \alpha)$$

has $f(d_{PQ})$ as variogram irrespectively of the value of α .

In fact, using (2.6)

Squaring (2.12) and subtracting from (2.10) yields

$$\sigma^2\{|\underline{R}_P - \underline{R}_Q|\} = E\{|\underline{R}_P - \underline{R}_Q|^2\} - \{E[|\underline{R}_P - \underline{R}_Q|]\}^2 = \gamma(d_{PQ}), \tag{2.13}$$

which is the sought relation.

According to (1.1), (2.13) implies

$$\gamma(d_{PQ}) = k^2 d_{PQ}^2. \tag{2.14}$$

One interesting remark is that the model (2.14), which is indeed an authorized variogram (Wackernagel 2003), is not however compatible with a stationary covariance.

This is the reason why we have from the beginning pointed to kriging, rather than the ordinary Wiener–Kolmogorov theory, also known as collocation in geodesy.

The updating algorithm

Now that the relations (2.13), (2.14) are established, the full machinery of kriging theory is available to solve the updating problem formalized in the following.

We assume to have a geodatabase with points P identified by the vector

$$\underline{R}_P = \begin{vmatrix} X_P \\ Y_P \end{vmatrix} = \begin{vmatrix} x_P \\ y_P \end{vmatrix} + \begin{vmatrix} \delta x_P \\ \delta y_P \end{vmatrix} = \underline{r}_P + \underline{u}(P), \tag{3.1}$$

where $\underline{u}(P)$ is an intrinsic random field, homogeneous and isotropic, endowed with the known variogram $\gamma(d_{PQ})$, given, in the present context, by (2.14).

Moreover, we assume that new measurements have been performed involving part of the points $\{P_i\}$ so that we have from them, namely for a certain subset J of the indexes $\{i = 1, 2, \dots, N\}$, a new sequence of coordinates

$$\underline{N}_i = \underline{r}_i + \underline{v}_i, \quad i \in J, \tag{3.2}$$

where we have just denoted with an index i quantities related to the point P_i .

In (3.2), we suppose that

$$E\{\underline{v}_i\} = 0, \quad E\{\underline{v}_i \underline{v}_j^t\} = C_{v_i v_j}, \tag{3.3}$$

so that the full covariance of the vector $\underline{v} = [\dots \underline{v}_i^t \dots]^t$ is known. For the sake of simplicity, we shall assume that the index set J is just that of the first J positions. Since we are thinking of an updating of the geodatabase, we obviously think of the variances of the coordinates in \underline{N}_i as significantly smaller than the variances of the geodatabase errors \underline{u}_i . Moreover, we assume \underline{v}_i to be uncorrelated with the field $\underline{u}(P)$.

Now we observe that

$$\underline{R}_i - \underline{N}_i = \underline{u}_i - \underline{v}_i, \quad i \in J; \tag{3.4}$$

this relation means that we can consider $\underline{R}_i - \underline{N}_i$ as an observation of the random field $\underline{u}(P)$ at the updating points P_i , $i \in J$, with noise $-\underline{v}_i$. So, we shall put

$$\underline{u}_{oi} = \underline{u}_i - \underline{v}_i = \underline{R}_i - \underline{N}_i \quad i \in J, \tag{3.5}$$

\underline{u}_{oi} being the observation vector.

The problem is precisely how to predict $\underline{u}(P)$ at any other point of the geodatabase.

We look for a linear prediction, i.e., for a linear combination of the observed \underline{u}_{oi} ,

$$\hat{\underline{u}}(P) = \sum_{i=1}^J \lambda_i \underline{u}_{oi}, \tag{3.6}$$

unconditionally unbiased or, recalling (2.3), such that

$$E\{\hat{\underline{u}}(P)\} = \underline{t}. \tag{3.7}$$

Moreover we will require that the mean square prediction error E^2

$$E^2 = E\left\{\left|\underline{u}(P) - \hat{\underline{u}}(P)\right|^2\right\}, \tag{3.8}$$

be minimized. Note that the class of our predictors (3.6) is not the most general one, which would be of the form

$$\hat{\underline{u}}(P) = \sum_{i=1}^J \Lambda_i \underline{u}_{oi} \tag{3.9}$$

with Λ_i a set of 2×2 matrices. On the other hand, it is easy to see that a predictor like (3.9) could be determined only if we knew the full covariance of \underline{u} , which is not the case.

The form (3.6) in fact translates our prior hypothesis of isotropy, which is not a model for how the “true” error $\underline{u}(P)$ is but rather the formalization of our lack of prior information.

In other words, this is where the Bayesian concept plays its role. As usual the condition (3.7) implies that the coefficients λ_i have to satisfy the relation

$$\sum_{i=1}^J \lambda_i = 1. \tag{3.10}$$

Now the proof runs as that of ordinary kriging (Wackernagel 2003) and we report it here only briefly.

First one can prove that, under (3.10),

$$E^2 = 2 \sum_{i=1}^J \lambda_i \gamma_i - \sum_{i,k=1}^J \lambda_i \lambda_k \gamma_{ik} + \sum_{i,k=1}^J \lambda_i \lambda_k \text{Tr} C_{v_i v_k}, \tag{3.11}$$

where

$$\gamma_i = \gamma(d_{PP_i}), \quad \gamma_{ik} = \gamma(d_{P_i P_k}) \tag{3.12}$$

and we have used

$$E\{\underline{v}_i^t \underline{v}_k\} = \text{Tr} C_{v_i v_k}. \tag{3.13}$$

So by using the synthetic notation

$$\underline{\lambda} \equiv \{\lambda_i\}, \quad \underline{\gamma} = \{\gamma_i\}, \quad \Gamma = \{\gamma_{ik}\}, \quad \Sigma = \{\text{Tr} C_{v_i v_k}\}$$

we have

$$E^2 = 2 \underline{\lambda}^t \underline{\gamma} + \underline{\lambda}^t (-\Gamma + \Sigma) \underline{\lambda}. \tag{3.14}$$

In fact from

$$\begin{aligned} \underline{u}(P) - \hat{\underline{u}}(P) &= \underline{u}(P) - \sum_{i=1}^J \lambda_i \underline{u}_{oi} = \sum_{i=1}^J \lambda_i [\underline{u}(P) - \underline{u}_{oi}] \\ &= \sum_{i=1}^J \lambda_i [\underline{u}(P) - \underline{u}_i - \underline{v}_i], \end{aligned}$$

which holds because of (3.10), one derives

$$\begin{aligned} E^2 &= E\left\{\left|\underline{u}(P) - \hat{\underline{u}}(P)\right|^2\right\} \\ &= E\left\{\sum_{ik} \lambda_i \lambda_k [\underline{u}(P) - \underline{u}_i]^t [\underline{u}(P) - \underline{u}_k]\right\} + \sum_{ik} \lambda_i \lambda_k C_{v_i v_k}. \end{aligned}$$

Now using the identity

$$\begin{aligned} [\underline{u}(P) - \underline{u}_i]^t [\underline{u}(P) - \underline{u}_k] &= \frac{1}{2} |\underline{u}(P) - \underline{u}_i|^2 \\ &\quad + \frac{1}{2} |\underline{u}(P) - \underline{u}_k|^2 - \frac{1}{2} |\underline{u}_i - \underline{u}_k|^2 \end{aligned}$$

and again recalling (3.10), one easily derives (3.14).

The target function (3.14) can be minimized with respect to $\underline{\lambda}$ with the constraint (3.10) that is written, introducing the vector $\underline{e}' = [11\dots 1]$, as

$$\underline{e}'\underline{\lambda} = 1. \tag{3.15}$$

By using a Lagrange multiplier -2α , one finds the usual kriging equations

$$\begin{cases} (-\Gamma + \Sigma)\underline{\lambda} + \alpha\underline{e} = \underline{\gamma} \\ \underline{e}'\underline{\lambda} = 1 \end{cases} \tag{3.16}$$

Once $\underline{\lambda}$ is derived, $\hat{\underline{u}}(P)$ is computed through (3.6) and (3.14) provides its prediction error.

Then the updated geodatabase at the original points P_i , of coordinates $\underline{R}_{up}(P_i)$ is

$$\underline{R}_{up}(P_i) = \underline{R}_i - \hat{\underline{u}}(P_i) \tag{3.17}$$

and, since

$$\underline{R}_i - \hat{\underline{u}}(P_i) = \underline{r}_i + \underline{u}(P_i) - \hat{\underline{u}}(P_i), \tag{3.18}$$

the prediction error at P_i becomes the error of the updated geodatabase.

We conclude the paragraph with a remark. Namely, in addition to the above results, one would like as well to get hold of the translation between geodatabase and updated coordinates. It is known that the optimal estimate can be obtained only if we know the full covariance structure of $\underline{u}(P)$.

But since this is not available, we can still have a non-optimal estimate, for instance by taking

$$\hat{\underline{t}} = \frac{1}{j} \sum_{i=1}^j \underline{u}_{oi}. \tag{3.19}$$

Though we are not able to give the variance of this estimator, (3.19) is sufficient for many practical purposes.

An extension and a few examples

Since most of the updating work is nowadays done by GPS and since, when a positioning service is not available, a

$$\begin{aligned} E^2 = E\{[\delta\underline{u}_P - \lambda\delta\underline{u}_{o2}]^2\} &= E\{[(\underline{u}_P - \underline{u}_1) - \lambda\{\underline{u}_2 - \underline{u}_1\}]^2\} + \lambda^2 E\{|\underline{v}|^2\} = \\ &= 2\gamma(d_{PP_1}) + 2\lambda^2\gamma(d_{P_2P_1}) - 2\lambda E\{(\underline{u}_P - \underline{u}_1)'(\underline{u}_2 - \underline{u}_1)\} + \lambda^2 Tr\{C_{vv}\} \end{aligned} \tag{4.6}$$

On the other hand

$$E\{(\underline{u}_P - \underline{u}_1)'(\underline{u}_2 - \underline{u}_1)\} = \frac{1}{2} E\{|\underline{u}_P - \underline{u}_1|^2 + |\underline{u}_2 - \underline{u}_1|^2 - |\underline{u}_P - \underline{u}_2|^2\} = \gamma(d_{PP_1}) + \gamma(d_{P_2P_1}) - \gamma(d_{PP_2}). \tag{4.7}$$

natural outcome of GPS observations is base vectors between couples of points or even local networks of base vectors, it is interesting to see whether and how the above developed theory can account for this kind of measurements.

We will do that taking the case of single base, between points P_1 and P_2 ,

$$\underline{b} = \underline{N}_2 - \underline{N}_1 = \underline{r}_2 - \underline{r}_1 + \underline{v}_2 - \underline{v}_1 \tag{4.1}$$

leaving to the reader the easy generalization to more bases. We assume that $E\{\underline{v}\} = E\{\underline{v}_2 - \underline{v}_1\} = 0$ and that the covariance of \underline{v} is known. This case is not covered by ordinary kriging theory and then it requires some adjustment.

The point is that now the observations are themselves translation invariant and therefore they cannot convey any information on the absolute value of $\underline{u}(P)$.

Nevertheless, we can expect to be able to say something about variations of $\underline{u}(P)$ for instance about $\delta\underline{u}_P = \underline{u}(P) - \underline{u}(P_1)$. So we have as observation:

$$\begin{aligned} \delta\underline{u}_{o2} = \underline{u}_o(P_2) - \underline{u}_o(P_1) &= [\underline{R}(P_2) - \underline{R}(P_1)] - [\underline{N}(P_2) - \underline{N}(P_1)] = \\ &= \underline{u}(P_2) - \underline{u}(P_1) - (\underline{v}_2 - \underline{v}_1) = \underline{u}_2 - \underline{u}_1 - \underline{v} = \delta\underline{u}_2 - \underline{v} \end{aligned} \tag{4.2}$$

and we try to estimate $\delta\underline{u}_P$ as a linear function of $\delta\underline{u}_{o2}$, namely

$$\delta\hat{\underline{u}}_P = \lambda\delta\underline{u}_{o2}. \tag{4.3}$$

It is extremely important to notice however that in this context

$$E\{\delta\underline{u}_P\} = E\{\underline{u}(P) - \underline{u}(P_1)\} = \underline{t} - \underline{t} = 0 \tag{4.4}$$

as well as

$$\begin{aligned} E\{\delta\hat{\underline{u}}_P\} &= \lambda E\{\delta\underline{u}_{o2}\} = \lambda E\{\delta\underline{u}_2 - \underline{v}\} \\ &= \lambda E\{\underline{u}_2 - \underline{u}_1\} = 0, \end{aligned} \tag{4.5}$$

so that there is no constraint on (4.3) to get an unbiased predictor, namely λ is a free variable. Now we proceed to compute the r.m.s. prediction error, i.e.

We call, for the sake of simplicity

$$\gamma(d_{PP_i}) = \gamma_i \quad i = 1, 2 \quad \gamma(d_{P_1P_2}) = \gamma_{12} \quad 2\Sigma_v = Tr\{C_{vv}\}$$

and then, combining (4.7) and (4.6), we find

$$E^2 = 2(1 - \lambda)\gamma_1 + 2\lambda(\lambda - 1)\gamma_{12} + 2\lambda\gamma_2 + 2\lambda^2\Sigma_v. \quad (4.8)$$

The minimum of E^2 is obtained at

$$\lambda = \frac{1}{2} \frac{\gamma_1 - \gamma_2 + \gamma_{12}}{\gamma_{12} + \Sigma_v}. \quad (4.9)$$

From (4.9) and (4.8), the prediction error can be computed. So we can put now

$$(\underline{R}_P - \underline{R}_1)_{up} = (\underline{R}_P - \underline{R}_1) - \delta \hat{\underline{u}}_P \quad (4.10)$$

and therefore we have

$$(\underline{R}_P - \underline{R}_1)_{up} = \underline{r}_P - \underline{r}_1 + \delta \underline{u}_P - \delta \hat{\underline{u}}_P. \quad (4.11)$$

Formula (4.10) tells us how to update the geodatabase, relative to the position of P_1 , while (4.11) tells us that the prediction error of $\delta \underline{u}_P$ is the new error of the updated geodatabase.

To get acquainted with the above concepts, we present two small artificial examples.

Example 1: assume you have a geodatabase with a prior relative error of $\sigma(d) = 2 \cdot 10^{-4}d$. We have two points P_1 and P_2 observed by GPS with errors $\underline{v}_1, \underline{v}_2$ such that

$$C_{v_1v_1} = C_{v_2v_2} = \sigma_v^2 I \quad \sigma_v = 0.1 \text{ m} \quad C_{v_1v_2} = 0.$$

The position of the two points is respectively -1 km and $+1$ km on the x -axis at the scale of the terrain.

The errors observed at P_1, P_2 are $\underline{u}_{o1}, \underline{u}_{o2}$, and are given in meters.

Note that, according with (2.14), $\gamma(d) = 4 \cdot 10^{-8}d^2$ (d given in meters) $= 4 \cdot 10^{-2}d^2$ (d given in kilometers), meaning that we shall express d in km but the result for γ will be in m^2 .

So we have

$$\Gamma = \begin{vmatrix} 0 & 4 \cdot 10^{-2}d_{12}^2 \\ 4 \cdot 10^{-2}d_{12}^2 & 0 \end{vmatrix} = \begin{vmatrix} 0 & 16 \cdot 10^{-2} \\ 16 \cdot 10^{-2} & 0 \end{vmatrix}$$

$$\Sigma_v = \begin{vmatrix} 2\sigma_v^2 & 0 \\ 0 & 2\sigma_v^2 \end{vmatrix} = \begin{vmatrix} 2 \cdot 10^{-2} & 0 \\ 0 & 2 \cdot 10^{-2} \end{vmatrix}$$

$$\underline{\gamma} = \begin{vmatrix} 4 \cdot 10^{-2}d_1^2 \\ 4 \cdot 10^{-2}d_2^2 \end{vmatrix}$$

and the conditioned normal system (3.16) becomes

$$\begin{cases} -2 \cdot 10^{-2}\lambda_1 + 16 \cdot 10^{-2}\lambda_2 + \alpha = 4 \cdot 10^{-2}d_1^2 \\ 16 \cdot 10^{-2}\lambda_1 - 2 \cdot 10^{-2}\lambda_2 + \alpha = 4 \cdot 10^{-2}d_2^2 \\ \lambda_1 + \lambda_2 = 1 \end{cases}$$

We notice that, in this system, λ is non-dimensional, α is in m^2 as the known terms, but d_1 and d_2 are expressed in km.

The solution of the system is

$$\begin{cases} \lambda_1 = \frac{1}{2} - \frac{1}{9}(d_1^2 - d_2^2) \\ \lambda_2 = \frac{1}{2} + \frac{1}{9}(d_1^2 - d_2^2) \\ \alpha = 10^{-2}(2d_1^2 + 2d_2^2 - 7) \end{cases}$$

So that, using the relation $d_1^2 - d_2^2 = 4x$ (x in km), the geodatabase correction is given by

$$\hat{\underline{u}}(P) = \frac{1}{2}(\underline{u}_{o1} + \underline{u}_{o2}) + \frac{4x}{9}(\underline{u}_{o2} - \underline{u}_{o1}).$$

The result nicely illustrates that, when $\underline{u}_{o1} = \underline{u}_{o2}$, $\hat{\underline{u}}(P)$ is just a common translation. The unit of $\hat{\underline{u}}(P)$ are meters as those of $\underline{u}_{o1}, \underline{u}_{o2}$.

Finally, we compute the new geodatabase error (in m^2) and we find

$$E^2(P) = 10^{-2} \left(\frac{8}{9}x^2 + 8y^2 + 1 \right).$$

Here, we read that at the origin the error is 0.1 m, which is a reasonable number, and that the error increases faster in the Y direction, which again is understandable, given the design of the updating points.

Example 2: we take the same situation as in Example 1, with the only difference that now we assume that instead of giving separately, the updated coordinates of P_1 and P_2 , we give the base vector $\underline{N}_2 - \underline{N}_1$.

Correspondingly (cf. 4.2), we will have an observation $\delta \underline{u}_2$ with a vector noise \underline{v} that now we assume to have covariance

$$C_{vv} = \sigma_v^2 I \quad \sigma_v = 0.02 \text{ m}.$$

Our purpose is to predict $\delta \hat{\underline{u}}(P)$ according to (4.3). The solution is almost given by (4.9) so we have only to use

$$\gamma_i = 4 \cdot 10^{-2}d_i^2 \text{ (km)}, \quad d_i^2 = |\underline{r}_P - \underline{r}_i|^2 \quad (i = 1, 2)$$

$$\gamma_{12} = 16 \cdot 10^{-2}, \quad \Sigma_v = \sigma_v^2 = 4 \cdot 10^{-4}$$

in that formula.

The result is

$$\lambda = \frac{2(x + 1)}{4.01};$$

we observe that indeed $\lambda = 0$ when $x = -1, y = 0$ because in that case $\delta \underline{u}_P = 0$.

On the other hand, the same is true when $x=-1$ and y whatever; this is explained by the fact that we have only one base vector observed, directed along x and this gives no information on $\delta\mathbf{u}$ when $\mathbf{r}_p - \mathbf{r}_1$ is along the y axis.

Finally, a computation of E^2 up to 10^{-4}m^2 , and neglecting 10^{-6}m^2 , gives

$$E^2(P) = 4 \cdot 10^{-2} \left[(x+1)^2 + y^2 - (x+1)^2 \left(1 - \frac{10^{-2}}{4} \right) \right].$$

As we see E^2 is zero for $y=0$, $x=-1$ as it should be. Also interesting is the case $y=0$, $x=1$, namely $P=P_2$, where $E^2(P) = 4 \cdot 10^{-2}\text{m}^2$, which is complying with the noise of $\delta\mathbf{u}_{02}$.

Conclusions

If one assumes that the vector field $\mathbf{u}(P)$ of the position error on a geodatabase is a homogeneous, isotropic intrinsic random field with constant mean and variogram (2.14), known a priori from the relative accuracy of the geodatabase, one can attack the problem of updating the position \mathbf{R}_p , of any point P on the geodatabase, by kriging theory.

With a typical GPS updating survey in mind, in the paper two cases are covered: that the coordinates of individual points P_i , $i \in J$, are re-determined or that base vectors $\mathbf{b}_{ik} = \mathbf{N}_{P_i} - \mathbf{N}_{P_k}$ are determined.

On the basis of the new observations, the new geodatabase coordinates are derived and their mean square error is computed.

As explicitly stated, this approach does not take into account any other geometric relation between points than that of the relative accuracy of the bases $|\mathbf{R}_i - \mathbf{R}_k|$.

If other constraints have to be considered, this can always be done a posteriori.

For instance, consider the case of Example 1 and assume that four points P_3, P_4, P_5, P_6 have been accordingly

updated, with coordinates $\mathbf{R}_{\text{up}3}, \mathbf{R}_{\text{up}4}, \mathbf{R}_{\text{up}5}, \mathbf{R}_{\text{up}6}$ with errors of known variances. Assume further that we want to impose the condition that P_3, P_4, P_5, P_6 are the vertices of a rectangle; then one can easily create the family of rectangles in the plane, depending on five parameters (namely the sides a, b and the three parameters of a roto-translation) and find the one that best fits the updated coordinates. Similar artifacts can be found to impose different conditions. Of course, a full rigorous theory would adjust the updating observations, taking the constraint into account all together. But this is left for future work.

References

- Arnold LM, Wright GL (2005) Analysing product-specific behaviour to support process dependent updates in a dynamic spatial updating model. *Trans GIS* 9(3):397–419
- Belussi A, Brovelli MA, Negri M, Pelagatti G, Sansò F (2006) Dealing with multiple accuracy levels in spatial databases with continuous update, *Proceedings of Accuracy 2006—7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences*, Instituto Geografico Portugues, Lisboa, Portugal, pp 203–212
- Brovelli MA, Zamboni G (2004) Adaptive transformation of cartographic bases by means of multiresolution spline interpolation. *Int Arch Photogramm Remote Sens Spat Inf Sci* 35(part B):206–211
- Brovelli MA, Zamboni G (2006) The usability of vectorization and a new point matching procedure as first step in conflating raster and vector maps. *Int Arch Photogramm Remote Sens Spat Inf Sci* 36(part 2/W40):85–91
- Leung Y, Ma J-H, Goodchild MF (2004) A general framework for error analysis in measurement-based GIS part 1: the basic measurement-error model and related concepts. *J Geogr Syst* 6(4):325–354
- Matheron G (1970) *La théorie des variables régionalisées, et ses applications*. Fascicule 5 des cahiers du Centre de Morphologie Mathématique de Fontainebleau
- Rigaux P, Scholl M, Voisard A (2002) *Spatial databases: with application to GIS*. Morgan Kaufmann Series in Data Management Systems
- Wackernagel H (2003) *Multivariate geostatistics: an introduction with applications*, 3rd edn. Springer, Berlin