Nat. Volatiles & Essent. Oils, 2021; 8(4): 3252-3259

Physical Correlation and Some of its Dangerous Properties

Mikhail D.Gerasimenko[1], Nikolay V.Shestakov[2], Valerii M.Kamornyi[3]

[1,2]*Institute of Applied Mathematics FEBRAS, Radio st., 7, Vladivostok 690041, Russia*
[1,2,3]*Far Eastern Federal University, Off E502a, 10 Ajax Bay, Russky Island, Vladivostok 690922, Russia*

Abstract

Apart from algebraic correlations, we need to take into account physical correlations in observations due to inestimable systematic errors. This is especially important in the case of GPS measurements. Different investigators have obtained values of correlation coefficients of up to 0.85 for identical satellite configuration due to physical correlations. The attempt has been to introduce these correlations or some correlation functions into post-processing of GPS measurements in order to improve the results. As illustrated in our report however, it may be very problematic in practice and could lead to extraordinarily absurd results. One such unexpected effect may be artificial increase of the weight P of the results. When weight p_i even of one observation approaches zero ($p_i \to 0$), we could obtain P→∞. The result is clearly absurd and has no physical meaning. That is why the formal estimates of accuracy do not reflect the real situation in the case of physically correlated observations. We encountered this phenomenon when developing a code for optimal GPS geodetic network design for fault-mechanics studies. Considering this undesirable phenomenon, we do not exclude the necessity for revision of some methods used in improving GPS processing results. It would be particularly important in methods that use different non-diagonal covariance matrices of observations. We have proposed a method for correction of physical correlation which solved a numerical problem in the optimal design of geodetic networks.

Keywords: Correlation Matrix, Extraordinary Influence of Correlation Coefficients, Optimal Design, Physical Correlation, Geodetic Network, GPS.

1. Introduction

Today the GPS is widely used in geodesy and geodynamics. In connection to this, there has been a significantly increased interest in adjustment of correlated observations. GPS data for closely distributed stations are physically correlated because they are made in practically identical atmosphere conditions, using the same satellite constellation and are thus affected by almost the same systematic errors. Therefore, we not only need to take into account algebraic correlations but physical ones as well in data processing. For example, [1] analyzed the multipath effect at the same site for two consecutive days and obtained values of correlation coefficients due this error source of up to 0.85 for identical satellite configurations. The influence of physical correlation on accuracy estimates of GPS measurements is shown by [2] as well. These correlations have a considerable influence on the evaluation of GPS measurements [3]. Sometimes this problem is solved by introducing into the post-processing stage the correlation matrix of GPS observations [4, 5]. Very often the correlation coefficients of this matrix are calculated using some experimental or intuitive assumptions which might not reflect the real physical situation.

Sometimes these correlation coefficients have an extraordinary influence on processing results and in the most complicated cases could eventually lead to absurd results. To our knowledge, this problem has not been adequately addressed in any current literature available to us. There are only a few examples of such effects for strongly correlated data [6]. Probably it could be due to the large correlation matrix involved, significant processing volume and its complication. As a result, many details in the computation process are hidden from researchers, especially if all calculations are performed by a computer. Exaggerated improvement in the accuracy of the final results may be misinterpreted as the correct choice of the modeling function used for calculation of correlation coefficients. In reality, however such accuracy improvement may be as a result of mathematical manipulations and not reflect the real situation.

From our experience, no significant results have been obtained during the last decades in correlation errors processing theory. Usually, the problem discussed here is solved by formally applying the principles of correlation theory [7].But this “formal” usage of mathematical methods for modeling physically correlated observations could lead to extraordinary results as shown in the next sections using simple numerical examples.

2. Methodology. Weighted Mean of Correlated Observations of the Same Variable

Let *n* correlated observations of the same variable *y* be given as a vector of observations $\boldsymbol{Y} = (\boldsymbol{y_1}, \boldsymbol{y_2}, \dots, \boldsymbol{y_n})^{\boldsymbol{T}}$with non-identical weight$\boldsymbol{p_i}$. The cofactor matrix of the vector of observations is given by equation

$$\mathbf{Q} = \mathbf{P}^{-1/2}\mathbf{R}\mathbf{P}^{-1/2} \qquad (1)$$

Where $\mathbf{P} = \boldsymbol{diag}(\boldsymbol{p_1}, \boldsymbol{p_2}, \dots, \boldsymbol{p_n})$ is the diagonal weight matrix and$\mathbf{R}$ is the correlation matrix. The matrix $\mathbf{R}$ has units on the main diagonal. Its non-diagonal elements r_{ij} are correlation coefficients of observations y_i and y_j.

The least square solution of a system of equations

$$\mathbf{A}y - Y = V \ , \qquad (2)$$

gives us the estimate of unknown quantity

$$\hat{y} = (\mathbf{A}^T\mathbf{Q}^{-1}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{Q}^{-1}Y \ , \qquad (3)$$

where matrix $\mathbf{A}^{\boldsymbol{T}} = (\mathbf{1}, \mathbf{1}, \dots, \mathbf{1})$ consists of units only,

$$\mathbf{Q}^{-\mathbf{1}} = \mathbf{P}^{\mathbf{1/2}}\mathbf{R}^{-\mathbf{1}}\mathbf{P}^{\mathbf{1/2}} \qquad (4)$$

is the weight matrix,$\boldsymbol{V}$is the vector of residuals. The value

$$P_{\hat{y}} = \mathbf{A}^T\mathbf{Q}^{-1}\mathbf{A} = \sum_{i=1}^{n}\sum_{j=1}^{n} q_{ij}^{(-1)} \ , \qquad (5)$$

is the weight of weighted mean of correlated observations $\boldsymbol{y_i}$, where $\boldsymbol{q_{ij}^{(-1)}}$are the elements of the weight matrix $\mathbf{Q^{-1}}$. Therefore the estimate of the varianceŷare given as

$$\sigma_{\hat{y}}^2 = \frac{1}{n-1}\frac{V^T\mathbf{Q}^{-1}V}{P_{\hat{y}}} \ , \qquad (6)$$

Let us consider in more details the case when $\boldsymbol{n} = \mathbf{2}$, i.e. only two correlated observations $\boldsymbol{y_1}$ and $\boldsymbol{y_2}$ with the correlation coefficient $\boldsymbol{r_{12}} = \boldsymbol{r}$ are made. In this case we have

$$\mathbf{Q}^{-1} = \frac{1}{1-r^2}\begin{pmatrix} p_1 & -r\sqrt{p_1p_2} \\ -r\sqrt{p_1p_2} & p_2 \end{pmatrix}, \qquad (7)$$

$$P_{\hat{y}} = \frac{1}{1-r^2}\left(p_1 - 2r\sqrt{p_1p_2} + p_2\right), \qquad (8)$$

$$\hat{y} = \frac{\left(p_1 - r\sqrt{p_1p_2}\right)y_1 + \left(p_2 - r\sqrt{p_1p_2}\right)y_2}{p_1 - 2r\sqrt{p_1p_2} + p_2}, \qquad (9)$$

In the next section we make a preliminary inspection of a few particular examples using equations (7)-(9).

3. Results. A Few Particular Examples of Extraordinary Absurd Results

3.1 Example 1

Let's assume that both observations have the identical weights $p_1 = p_2 = p$. The equation (9) gives us the least square estimate

$$\hat{y} = (y_1 + y_2)/2.$$

This means that the arithmetic average value may not depend on the magnitude of the correlation coefficients. For two measurements it is obvious from the formula given above.

But for the weight of average value

$$P_{\hat{y}} = \frac{2p}{1+r} \ , \qquad (10)$$

a situation is not so simple. A correlation coefficient changes within the limits $-1 \leq r \leq 1$. Thus if r increases from 0 to +1 then weight $P_{\hat{y}}$decreases from $2p$to p. There are no serious objections to this result but the weight of one observation is ignored. Another situation if r changes from 0 to -1, the quantity $P_{\hat{y}} \geq 2p$. When $r \to -1$we obtain the weight $P_{\hat{y}} \to \infty$! This is a clearly absurd results and it is impossible in case of uncorrelated observations.

3.2 Example 2

Let's consider a case of non-identical weights. Let's assume $p_1 = p = 1,\ p_2 = cp = c$, where c is some constant. In that case we obtain

$$P_{\hat{y}} = \frac{1}{1-r^2}\left(1 - 2r\sqrt{c} + c\right)\ , \tag{11}$$

$$\hat{y} = \frac{\left(1 - r\sqrt{c}\right)y_1 + \left(c - r\sqrt{c}\right)y_2}{1 - 2r\sqrt{c} + c}, \tag{12}$$

From equation (11) follows that if $r > 0$ then weight $P_{\hat{y}} = min$ when $c = r^2$and then from equation (12) follows $\hat{y} = y_1$!

Moreover when $p_2 = c = 0$ (second observation is absent), we have again $\hat{y} = y_1$, butthe weight $P_{\hat{y}} = \frac{1}{1-r^2} \geq p_1$and when $|r| \to 1$we obtain the absurd result $P_{\hat{y}} \to \infty$. We obtain this result under the assumption that actually we carry out only one observation with weight $p_1 = 1$!

3.3 Example 3

From equation (12) as a particular case we could obtain that if $c = 1/r^2$(or $r = 1/\sqrt{c}$) then the estimate $\hat{y} = y_2$ with the weight $P_{\hat{y}} = c$. It means that the final results are fully independent from the first measurement, which is impossible in case of uncorrelated observations.

3.4 Example 4

Let's assume that we measured two height differences $h_1 = 101mm$ and $h_2 = 102mm$ with the physically correlation coefficient $r = 0.9$ and weights$p_1 = 1$ and $p_2 = c = 4$ from two original rappers with heights $H_1 = H_2 = 0\ mm$ to determine point. The height *H* of this point accordingly on the basis of common sense must be in the limits

$$h_1 \leq H \leq h_2\ . \tag{13}$$

Using the equation (12) we obtain $H \approx 102.57\ mm$! This result seems to be obviously absurd because it is more than numerical value of any of observations. For independent observations this phenomenon is not observed and $H \approx 101.8\ mm$. An almost identical example is given on p. 326 by Strang and Borre [6].

In order to remove the such absurd effect, it should be taken into account that the correlation coefficient must satisfies condition for two measurements of the same variable

$$r \leq min\left(\sqrt{\frac{p_1}{p_2}};\ \sqrt{\frac{p_2}{p_1}}\right) = r_0\ , \tag{14}$$

which is derived from natural inequalities $|v_1 + v_2| \leq |h_2 - h_1|$ for residuals to the measurements and evaluation of guaranteeing hit in the limits (13).

Then in our elementary example should be $r \leq 0.5$. When $r = 0.5$, the value *H*=102.0, which is more reasonable. However, here it is easy to see that provided the first measurement with a weight $p_1 = 1$not involved in the calculation of *H* that is not entirely correct. To remedy this situation, you should use the corrected correlation coefficient $\bar{r} = r \cdot r_0$. In our case$\bar{r} = r \cdot r_0 = 0.45$ and then $\mathrm{P} = 101.97$.

We now turn to equation (8) that is formally in our case it is suitable to determine the weight adjusted mark *H*. From (8) if $p_1 \to 0$ or $p_2 \to 0$, the weight adjusted determined mark $p_H \to \frac{max(p_1, p_2)}{1-r^2}$that is more than any weight of measurement. This is an absurd result!

4. Discussion. Application of Physical Correlation in Optimal Geodetic Network Design

The physical coefficients of correlation are functions of factors that contribute to the general features of the measurement results, i.e. make them dependent. Such factors can be practically the same atmospheric conditions at the points of measurements; the measurements are carried out by the same instrument and performer; an identical arrangement of obstacles and reflective surfaces at the points of satellite observations, etc. However, these functions explicitly raises major difficulties associated with the diversity and complexity of factors that contribute to the generality of the measurement results.

Therefore, the physical coefficients of correlation are assigned a priori and often without sufficient physical and/or mathematical justification. Frequently used covariance functions depend on some parameters whose numerical values are obtained experimentally or appointed from particular views of the authors about the nature of the factors that lead to physical dependence measurements. The main criterion for the selection of covariance functions and their parameters is usually "improving" the accuracy of the final result. However, as shown above, this "increase" accuracy can be a purely formal and meaningless, having nothing to do with reality. In other words, before taking into account possible physical correlation in a particular task, you need to make sure that you enter in processing of correlation coefficients do reflect the physical reality. Integrating physical correlation during mathematical processing of the measurements and used correlation functions is dedicated toa rather extensive literature [2, 3, 4, etc.].

Not touching these questions, let us discuss more details on its account when optimizing the weights of the designed networks. In this case we assume that the physical coefficients of correlation (in numerical or analytical form) adequately reflect the real physical dependence measurements. Our task is to overcome the numerical problem of optimal design, described above, the essence of which boils down to the fact that when reducing the weight of any planned measure, the weight function of the results of measurements can be increased, i.e. general the solution is absurd.

Let us see more detail on the account of physical correlation when we optimizing the weights of measurements in the designed networks. Our task is to ensure a reasonable solution, in order to overcome the numerical problem of optimal design similar to Example 2 and 4 where the whole solution is obtained absurd.

Note that in Example 2 we followed the formal mathematics that is used when performing calculations on a computer. Nevertheless, we should not forget and physics! The fact that when the weight of the second measurement $p_2 = \mathrm{c} = 0$, i.e. it is simply not available, to talk about any correlation of measurements do not make sense and should be put value $r = 0$. Then $P_{\hat{y}} = p_1 = 1$, that should be within the meaning. This is even more true that when $p_2 \to 0$, root mean error $m_2 \to \infty$ and the possible systematic influence, which are the reason for the introduction of the physical correlation, "drowned" in the random measurement errors.

Moreover, the above numerical problems occurs only for dependent measurements with different weights. In this case in the computer calculations necessary to adjust the coefficients of physical correlation so that when a weight $p_i \to 0$ or $p_j \to 0$ the correlation coefficient $r_{ij} \to 0$. Suitable corrector can use the ratio of the minimum and maximum weights corresponding measurements; instead of physical correlation coefficient r_{ij} must be used in the optimization coefficient

$$\bar{r}_{ij} = r_{ij} \frac{\min(p_i, p_j)}{\max(p_i, p_j)} \ , \qquad (15)$$

which allows the problem posed, what is easy to see in the considered example above.

Additionally, we note that the revised by formula (15) the correlation coefficient also satisfies condition (14) for two measurements of the same variable that is removes theabsurd effects demonstratedinExample4.

Using, for example, formula (15) for adjusting the correlation coefficients yielded positive results in the optimal GPS geodetic network design for fault-mechanics studies by the method proposed in [8].The coefficients of the physical correlations were calculated according to the formula by Johnson and Wyatt [5]

$$r_{ij} = \frac{L^2}{L^2 + (d_i - d_j)^2} \ , \qquad (16)$$

Where L is a scaling length, $(d_i - d_j)$ is a distance between geodetic points i and j. In some conditions correlation coefficients r_{ij} may have very big values. For example, if $L = 5\,km$ km as assumed by [5] and $d_i - d_j = 1\,km$ we have $r_{ij} = 0.96$accordingly. As already mentioned above, the abnormal influence of those large correlation coefficients was manifested as accuracy improvement in the estimated fault parameters when useless GPS stations were eliminated from observation scheme by decreasing the weights to insignificant level of GPS measurements which were made on these sites.

Results of optimization for $L = 1\ km$ and $\left|d_i - d_j\right| \geq 1\ km$, which corresponds $r_{ij} \leq 0.5$, and using correction (15) were close to the results obtained by [8] that was obtained without taking into account the physical correlation..

It should be noted that the values of the coefficients of physical correlation, calculated by the formula (16) under condition $\left|d_i - d_j\right| < L$, will be greater than 0.5, that are unlikely to corresponds to the physical reality, and the use of optimization without adjustment (15) do not provide any reasonable solution. Correction coefficients according to formula (15) removes the problem.

5. Conclusions

At a first glance, the above discussed effects may appear to relate to some extreme cases only, with no practical use. Also, we have theoretically considered a case of two physically correlated observations only. But even in this simplest case we demonstrated unexpected peculiarities and absurd results which cannot be left without attention. What other unexpected effects can appear when processing correlated observations from spacious geodetic networks? By other words, investigations in correlated observations and their usage in practice is not completely solved. Considering this undesirable phenomenon, we do not exclude the necessity for revision of some methods used in improving GPS processing results especially those that make use of different non-diagonal elements of covariance matrices of observations. In some emergencies the formal estimates of accuracy do not reflect the real situation in case of physically correlated observations where accuracy may be artificially improved. Nevertheless, even considering this undesirable phenomenon in the case of two correlated observations, we have solved the numerical problem of obtaining a stable solution in the case of developing a code for optimal geodetic network design taking into account the possible dependence of projected measurements.

Acknowledgments

The work is executed within the framework of the state task of the Institute of Applied Mathematics of the Far Eastern Branch of the Russian Academy of Sciences.

Conflicts of Interest

The authors declare no conflict of interest.

References

Georgiadou Y.; Kleusberg A. On Carrier Signal Multipath Effects in Relative GPS Positioning. ManuscrGeod 1988, 13, 172-179, https://www.researchgate.net/publication/281570462_On_Carrier_Signal_Multipath_Effects_in_Relative_GPS_Positioning

El-Rabbany A.; The effect of physical correlations on the ambiguity resolution and accuracy estimation in GPS differential positioning. PhD Thesis 1994, Department of Geodesy and Geomatics Engineering, University of New Brunswick, Tech rep 170.

Schwieger V.; An Approach to Determine Correlations between GPS Monitored Deformation Epochs. Proc of the 8th FIG International Symposium on Deformation Measurements, Hong Kong Polytechnic University: 17-26, China, 25-28 June 1996.

Ananga N.; Coleman R.; Rizos C. Variance-covariance estimation of GPS networks. Bull Geodesique 1994, 68, 77-87.

Johnson H.O.; Wyatt F.K. Geodetic network design for fault-mechanics studies. ManuscrGeod 1994, 19, 309-323, https://www.scec.org/publication/160

Strang G.; Borre K.; Linear algebra, geodesy, and GPS.; Wellesley-Cambridge Press, Massachusetts, 1997; 638 p.

Kemnits U.V.; Mathematical processing of the results of correlated observations. Nedra 1970, Moskow.

Gerasimenko M.D.; Shestakov N.V.; Kato T. On optimal geodetic network design for fault-mechanics studies. Earth Planet Space 2000, 52, 985-987, https://doi.org/10.1186/BF03352317

Goyal, Megha, and Urmil Verma. "Wheat yield prediction using weather based statistical model in northern zone of Haryana." Int. J. Humanit. Soc. Sci 7.4 (2018): 47-50.

Elsayed, Ashraf M., Yasser A. Shehata, and Ossama M. Badawy. "Comparing The Effect Of Simulation Based Learning Versus On-Scene Learning In Self-Efficacy On Marine Cadets." International Journal of Library & Educational Science (IASET: IJLES) 1 (2015): 1-12.

Kavitha, A. K. "The Importance of Internet and Factors Influencing Internet Usage of Teaching Faculties-A Study with Special Reference to Selected Arts Colleges Affiliated to Bharathiar University." International Journal of Libr (2016).

Banu, A. Shamim, and G. Sundharavadivel. "A Relationship between Work-Life Balance and Job Performance of Employees." International Journal of Human Resource Management and Research (IJHRMR) 9.2, Apr 2019, 53-58

Sampath, O., et al. "Correlation and Regression Studies of Rice Varieties Influenced by Combination of Plant Density and Fertilizer Levels." International Journal of Agricultural Science and Research (IJASR) 7.3, Jun 2017, 233-240