

## Error Analysis of Regular Observations

**Momchil Minchev**

University of Mining and Geology, Sofia, Bulgaria  
e-mail: minchev@netissat.bg

### Introduction

In classical error theory observations are regarded as realizations of random values which allow estimates of specified qualities to be derived. If the experiment is conducted under stable environmental parameters, without disturbances, the observations will be normally distributed and estimates of lowest possible variances could be obtained. Otherwise, efficient estimates cannot be computed, and besides the inherent random errors would appear also biases due to external systematic effects.

The biases are usually regarded as deterministic values, modeled on the basis of specific theories, hypotheses, assumptions, etc. It is commonly accepted to perform geodetic surveys on particular occasions and with strictly defined purposes, as well as to introduce into the observation results corrections that would compensate the influence of the known systematic errors.

When empirical data series of considerable length and spatial coverage are available, processes and events whose instant state is realized through the systematic errors come to the fore. That requires the use of more powerful analysis tools, and also enhancement of the traditional terms and methodologies.

### Sporadic and regular data

Sporadic data results from singular observations designed without sufficient a priori information on the impact of external factors. Unlikely, regular data is a product of comprehensively designed observations realized considering the available external factor models  $L = L(E_1, E_2, \dots)$ . Ultimate regularity is achieved in case of continuous observations with period  $T_{\min}$ . Permanent geodetic networks, e.g. EPN, are typical examples for regular data sources (fig. 1).

If the external factors are not properly considered at design time, and the observation repeatability is defined mechanically, regarding just the technical limitations  $(\Delta t_{\min}, \sigma)$ , resulting data would be of periodic but not regular nature.

Errors of three types are distinguished within the classical error theory: gross errors, systematic errors and random errors. From practical standpoint that is completely reasonable, but not for analysis of regular data, when such definitions become inadequate. It is therefore necessary to introduce definitions as *jump*, *trend*, *white noise*, *additive/multiplicative disturbances*, etc., as well as to make clear the representativeness of the raw data samples with regard to time, i.e. instant data or averaged data over a particular interval [Minchev, 2007].

Not going into interpretation details, it is obvious that the North and East component graphs at fig. 1 clearly display a trend and jumps – on the events of transition from one reference frame to another or equipment upgrade. Unlikely, the height graph consists of a periodic component with amplitude of about 20 mm and wave length of 1 year apparent on the common noise background.

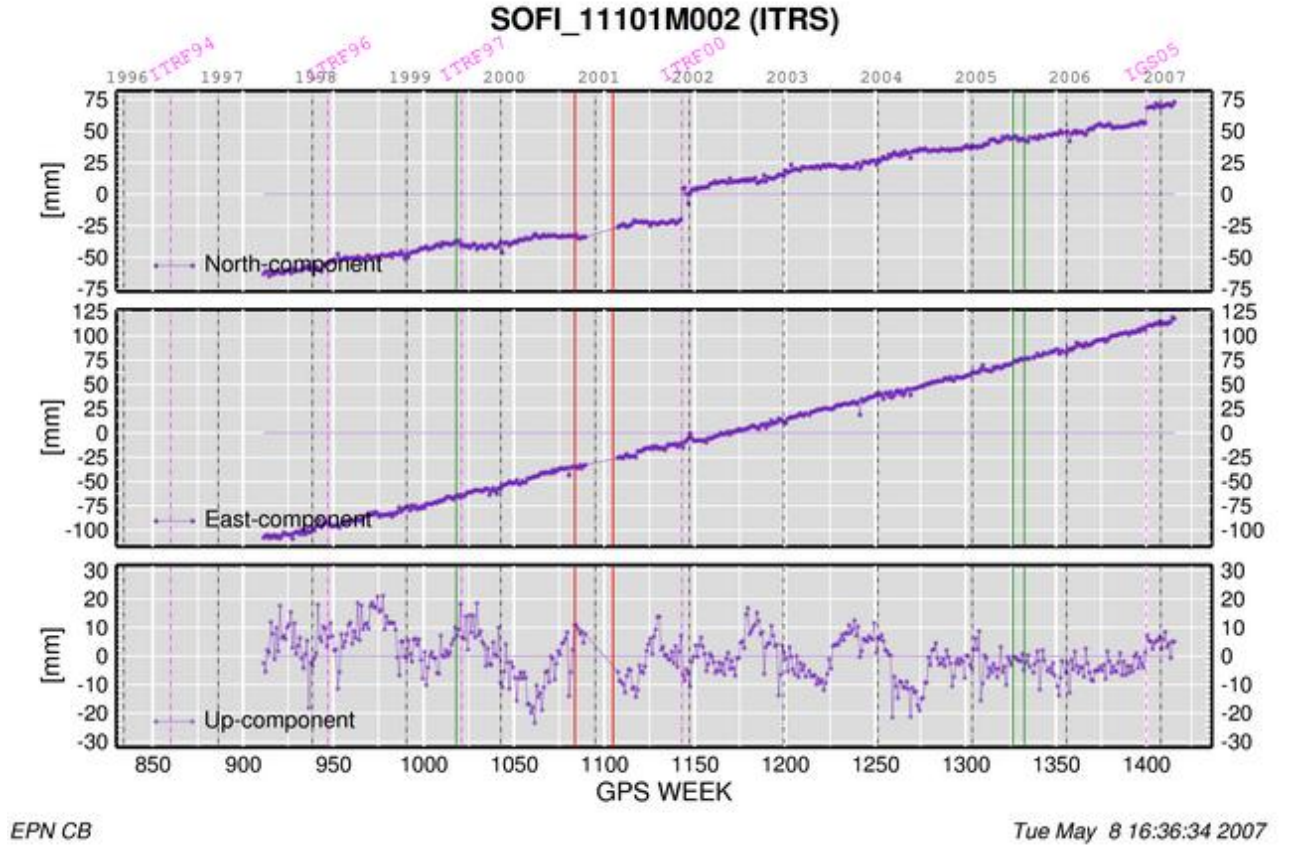


Figure 1. ITRS time series as typical regular data: SOFI Euref Permanent Network station  
Courtesy EPN CB – Royal Observatory of Belgium

### Modeling external factors impact

Observed data behavior is actually formed under the influence of multiple factors  $E_i \in \mathbf{E}$ , part of which ( $i=1-r$ ) identified both as phenomena and mechanism of impacting the measurements -

$$L = L(E_1, E_2, \dots, E_r). \quad (1)$$

If the values of those factors (e.g. atmosphere and ionosphere status, earth-crust dynamics, etc.) are defined during the observation session, their effect could be compensated by introducing relevant corrections. Otherwise, parameterization of each factor is necessary:

$$E_i = E_i(e_{i1}, e_{i2}, \dots, e_{ik}), \quad (2)$$

followed by linearization and including into the mathematical model along with the initial parameter set  $\mathbf{x}_0$ . For instance, the first case (1) is typical when regarding the Earth rotation parameters impact, whereas the second one (2) is used when modeling the atmosphere refraction. The later may consist of diurnal and season cycles, trends and other features expressed by the parameters  $e_{ij}$ . Thus, the deterministic part of the model is set up.

The rest of the external factors ( $e_{ij}$ ,  $i > r$ ) is not reliably identified, or their manner of impacting the measured data is not completely studied. Therefore, their total effect is presented as a

stochastic process  $x(t)$ , resulting from overlapping of finite number of sine waves of various amplitudes ( $A$ ), periods ( $T = 1/f$ ) and phases ( $\varphi$ ):

$$x(t) = \bar{x}(t) + \sum_{i=1}^n A_i \sin(2\pi f_i t + \varphi_i), \quad (3)$$

where  $\bar{x}(t)$  - mean section value at instant  $t$ . If the frequency ratio  $f_i/f_{i+1}$  is a rational number for all  $i$ -s, then the base period  $T_0$  of  $x(t)$  obtains finite values and the process is periodic, else it is undefined and the process is specified as non-periodic. In the first case  $x(t)$  could be presented also as a harmonic process -

$$x(t) = \bar{x}(t) + \sum_{k=1}^{\infty} (a_k \cos kt + b_k \sin kt), \quad (4)$$

with a base frequency  $f_0 = 1/T_0$ ;  $a_i, b_i$  - Fourier coefficients [Minchev, 2003].

Based on the inequality

$$T_i < \frac{1}{m} T', \quad (5)$$

where  $T_i$  - period of the  $i$ -th component in (3),  $m$  - sampling index, indicative for the sampling rate achievable at observation periods  $T'$ , long-period and short-period components can be distinguished in the spectrum of  $x(t)$ . According to the Nyquist criterion known from digital signal processing, a  $f_i$  harmonic of the  $x(t)$  frequency domain displays systematic behaviour if  $f_i \leq 2f'$ , where  $f' = 1/T'$  - observation frequency, and is to be modeled in the deterministic part of the design, along with the other parameters:

$$\mathbf{x}^T = [\mathbf{x}_0^T, \mathbf{x}_1^T, \mathbf{x}_2^T, \dots, \mathbf{x}_r^T], \quad (6)$$

where  $\mathbf{x}_0$  - initial design parameters,  $\mathbf{x}_i, i = 1, 2, \dots, r$  - external parameters (1, 2) -

$$\mathbf{x}_i^T = [e_{i1}, e_{i2}, \dots, e_{ik}]. \quad (7)$$

Hence the mathematical model equations are achieved

$$\mathbf{B}\mathbf{x} + \mathbf{G}\mathbf{s} + \mathbf{l} = \mathbf{v}', \quad (8)$$

where  $\mathbf{B} = \partial \mathbf{L} / \partial \mathbf{x}$  - design matrix corresponding to deterministic part

$$\mathbf{B}\mathbf{x} + \mathbf{l} = \mathbf{v}, \quad (9)$$

$\mathbf{G} = \partial \mathbf{L} / \partial \mathbf{s}$  - design matrix corresponding to the stochastic parameters  $\mathbf{s}$ , identified in (1) as short-period components of  $x(t)$ ,  $\mathbf{l} = \mathbf{L}^0 - \mathbf{L}'$ ,  $\mathbf{L}'$  - observation vector,  $\tilde{\mathbf{L}} = \mathbf{L}' + \mathbf{v}$  vector of estimates of the observed values  $\mathbf{L}$ ,  $\mathbf{v}' = \mathbf{v} - \mathbf{G}\mathbf{s}$  - unmodeled part (noise). Further, the solution follows the standard least squares procedure yielding new parameters whose meaning is to be further interpreted.

With the collection of new knowledge and enhancement of the mathematical models used, the relative part of the undefined factor impact decreases, which leads to reduction of the random errors and their gradual turn into white noise. Hence could be identified new jumps, trends and other changes due to unknown factors of smaller amplitude, which leads to determination of a next part of the noise, and so on, until reaching the limitation of the period and accuracy of the available observations  $\mathbf{L}'(T', \sigma)$ .

In case of sporadic observations, signal cannot be reliably distinguished from noise because the external factors  $E_i$  could be displayed as either systematic or random errors, depending on the observation interval  $\Delta t$  and the apparent phases  $\varphi_{E_i}$ . Therefore, systematic and random errors are attributed only to non-regular observations, and must always be related to the value of  $\Delta t$ .

### Regular observation design

For setting up an observation program are important the a priori information on the external factor  $E_i$  impact, and the technical limitations  $L'(\Delta t_{\min}, \sigma)$ . For the successful sampling of the  $E_i$  effect with period  $T_{E_i} = \frac{1}{f_{E_i}}$ , the condition  $f_{E_i} \leq 2f_L$  must be fulfilled, where  $f_L$  - observation frequency. Generally, the observation period is defined depending on the sampling rate:  $T_L = \frac{k}{m} \Delta t_{E_i}$ , where  $k \leq 2m$  - integer numbers.

If the  $E_i$  factor is stationary ( $f_{E_i} = \text{const}$ ) over a given time span  $[t_1, t_2]$ , the following equation is valid

$$T_L = nT_{E_i} + \Delta t_{E_i}, \quad (10)$$

where for each  $n$   $nT_{E_i} \in [t_1, t_2]$ ,  $\Delta t_{E_i} = \frac{T_{E_i}}{m}$ . That practically allows identifying in observation series of given period impacts of factors of shorter periods (fig. 2).

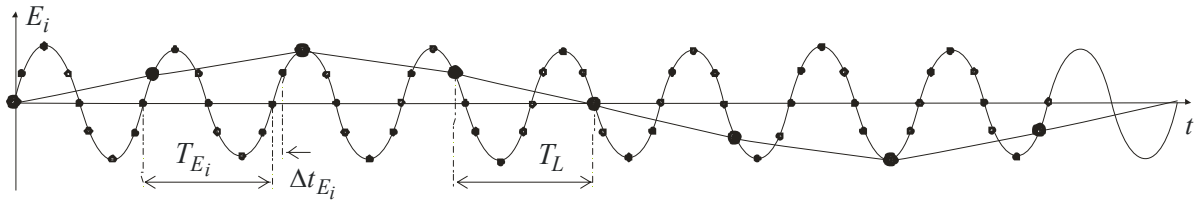


Figure 2. Observations under stationary external factor  $E_i$

In case the observation period is a multiple of an external factor,  $T_L = kT_{E_i}$ ,  $k$  - integer, the phase  $\varphi_{E_i}$  will be the same at all observation epochs, thus introducing the constant bias  $\Delta L(E_i) = A_{E_i} \sin(2\pi f + \varphi)_{E_i}$ . If the phase is close to  $90^\circ$  or  $270^\circ$  (at fig. 3  $\varphi_i = 90^\circ$ ), it is possible  $\Delta L$  to reach significant values  $\Delta L > t_\beta \sigma$ ,  $t_\beta$  - quantile of Student's distribution at  $\beta$  confidence probability. Thus the observations can be "cleared" from the  $E_i$  factor impact but the estimates will be biased. This effect could be of use for differential analysis and should be considered in the observation program design. Otherwise, results of unrealistic accuracy could be achieved.

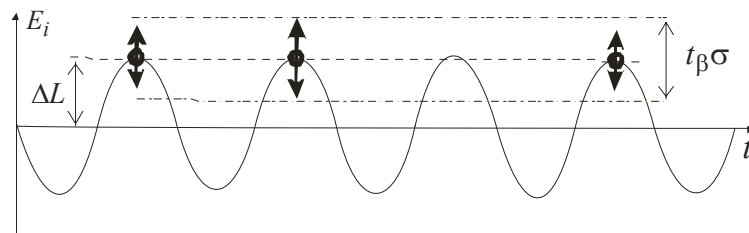


Figure 3. Multiple observation and external factor periods

Although two observation groups -  $(L'_1, L'_2, \dots)_A$ ,  $(L'_1, L'_2, \dots)_B$ , are normally distributed, respectively  $N(\bar{L}_A, s_A)$ ,  $N(\bar{L}_B, s_B)$ , and of equal accuracy -  $s_A \approx s_B$ , the results in the first case are apparently more optimistic only because they cover a relatively short time span where the unmodelled external effects remain practically unchanged, respectively  $\bar{s}_A \approx s_A$ , where  $\bar{s}_A$  - effective value of  $s_A$  impacted by  $E_i$  (fig. 4). As  $\Delta L_A > t_\beta \bar{s}_A$ , such estimate would be a misleading one. The broader interval in the second case,  $t_\beta \bar{s}_B$ , where  $\bar{s}_B > s_B$ , allows for achievement of more realistic results, as well as to look for a simplified way for compensating the unmodelled effects, e.g. a suitable trend. Similar cases are typically met when analysing static GPS observations, where shorter series yield apparently “more accurate” results than the longer ones.

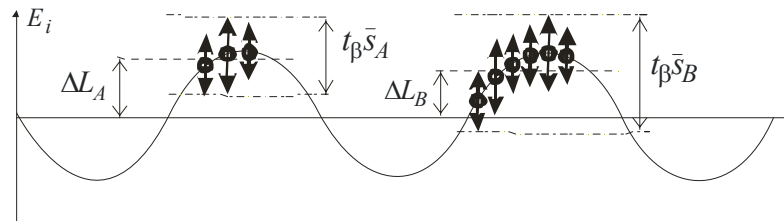


Figure 4. Observation session length effects

## Conclusion

Analyses of sporadic and regular observations require different approaches. Classical error theory is applicable for sporadic data processing whilst regular data, including GNSS observations collected at permanent stations, should be treated using more common and powerful methods. Signal processing tools allow for modeling the external factors impact and separating the inherent instrumental errors. Thus classical error theory could be extended renovated.

## References

- Minchev M. Methodological aspects of observation errors classification. UACEG Jubilee scientific conference, Sofia, May 2007
- Minchev M. An interpretation of the observation errors. Innovations in mine surveying and geoengineering, St. Konstantin and Elena, 16-20 June, 2003, 105-110