Slovak University of Technology in Bratislava
Faculty of Civil Engineering
Department of Mathematics and Descriptive Geometry

Modern multivariate approach to time series modelling

Written part of dissertation examination

PhD. student: **Ing. Tomáš Bacigál**
Supervisor: **Prof. RNDr. Magda Komorníková, CSc.**
Consultant supervisor: **Prof. Ing. Ján Hefty, PhD.**

Bratislava 2005

Contents

1 Introduction

Many technical disciplines involved in civil engineering, such as geology, geodesy, hydrology, statics of structures and others deal with geometric and physical quantities to figure out processes that influence our environment (both original and man-made). Supported by advancements and automation on the field of measuring instruments, monitoring becomes robust and effective, yet demanding more appropriate methods of processing. The most obvious geometric concern in geodesy is to determine a position of particular points in time-space. For the purpose, variety of techniques has been developed to precisely measure all the related mediating parameters, yet there is a special one popularity of which has rocketed up in recent times.

In our work we focus on modelling time-series arisen from observations by NAVSTAR Global positioning system (GPS), which is satellite based navigational system developed and provided by the American Department of Defence and now widely used in civil sector. Observations had been performed daily in years 2001-2002 on GPS permanent stations, which takes part in EUREF Permanent Network representing a regional densification of global IGS (International GPS Service) net in Europe which is used, among other purposes, for regular monitoring of recent kinematics of the Earth's crust (see [10]). The standard outcome, being in the form of three coordinates (`X,Y,Z`) in geocentric coord.system, was transformed into local topocentric horizontal coordinate system (`n,e,v` - north, east, vertical component) - with the origin set into the mean position of the two year period - to be further processed.

To understand the term "time series" and to follow the up-coming theory smoothly, it may be necessary to recall some fundamental facts.

A discrete time series is a set of time-ordered data $\{x_{t1}, x_{t2}, \dots x_{tn}\}$ taken from observations of some phenomenon, usually at equally spaced time intervals. The subscript t is referred to as time, n denotes the length of the time series and x_t is assumed to be real. The main purpose of time series analysis is to understand the underlying mechanism that generates the observed data and, in turn, to forecast future values. We assume that the generating mechanism is probabilistic and that the observed series $\{x_1, x_2, \dots x_t, \dots\}$ is a realization of a stochastic process $\{X_1, X_2, \dots X_t, \dots\}$, i.e., a sequence of random variables. In the following the term time series refers both to observed data and stochastic process, in formulas contextually distinguished by lower and upper case, respectively. Generally a time series consists of components like:

- trend, the long-term component representing growth or decline over an extended period of time
- seasonal component, annually repeating pattern of changes constrained within the most natural periodicity
- cyclical component, a wavelike fluctuation around the trend
- residuals, usually stochastic remains after deterministic components removal

In practice, when modelling typical time series the first three (classes of) components are picked up by regression (mostly by so-called OLS - ordinary least squares - regression procedure) and the residuals are subjected to analysis known as Box–Jenkins methodology that covers a large family of linear models such as autoregressive (AR), moving average (MA), integrated ARMA (ARIMA) and the like. There is also a far larger family of nonlinear models, that become popular in recent years. The stages of analysis suggested for linear models are generally applicable with nonlinear ones too, and can be briefly summarized:

- Model specification. Using various summary statistics, decide on a class of models to be used for a particular data set and also approximately the number of lags required.
- Model estimation. Estimate the parameters of the selected model(s).
- Model evaluation. Use a variety of inference statistics and specification tests to judge the quality of model and, perhaps, compare it with other models. For example, the out-of-sample forecasting abilities of the models can be compared. If the model is unsatisfactory, consider a new specification.

Also as mentioning later, the modelling methods tend from univariate time series analysis to multivariate structures and gradually become an effective tool when looking for relations among inspected processes.

The thesis is organized as follows. Section 2 introduce the multivariate analysis considering linear relations within and among variables. Individually, in section 2.1 the data are pictured and some general ways of their processing are suggested, results of which are compared and some comments are given in section 2.3. In between, some theory and tests results are given in section 2.2, namely, first we test the single data for the presence of stochastic trend, which may contain some deterministic components, then our interest is turned to the presence of common trending behaviour. For this Engle-Granger and Johansen's tests are used and common trend variable is estimated by means of Gonzalo-Granger method. Subsequent two-dimensional transformation helps to better understand the theory of commonly integrated time series giving a nice geometrical application. The last subsection gives a statistical answer to whether k time series contain the same deterministic trend.

On the other hand, section 3 comes to multivariate modelling from the side, where nonlinear structures in data are assumed. Following Tsay's work, it mainly focus on testing the null hypotheses of linearity against the alternative threshold nonlinearity, finding proper parameters for multivariate threshold autoregressive model and finally, building this model. Some results are given, but deeper analysis and comparison is still subject to study.

Finally, the fourth section summarizes our work.

2 Linear modelling

Despite the quite general title, it is not the aim of this section to give a complex overview of linear modelling methods, interested reader is referred e.g. to [8], [1] and [4]. Our intention is rather specific as we use standard models in Box-Jenkins methodology and underline the non-univariate approach, when looking for (linear) relationships between two or more variables.

2.1 Data and processing approach

As mentioned in the introduction, observations at permanent GPS stations were put to use, specifically those denoted BOR1, GOPE, POTS, HFLK and PENC were tested for common deterministic trend (section 2.2.4) and we chose BOR1 (Poland) for the majority of analyses in the parent section. Because of significantly lower precision and negligible linear trend in vertical direction, we only deal here with the two time series n_t and e_t, each containing 730 data points. Figure 1 shows two dimensional representation of point variation on Earth's surface and Figure 2 time plot for each coordinate.

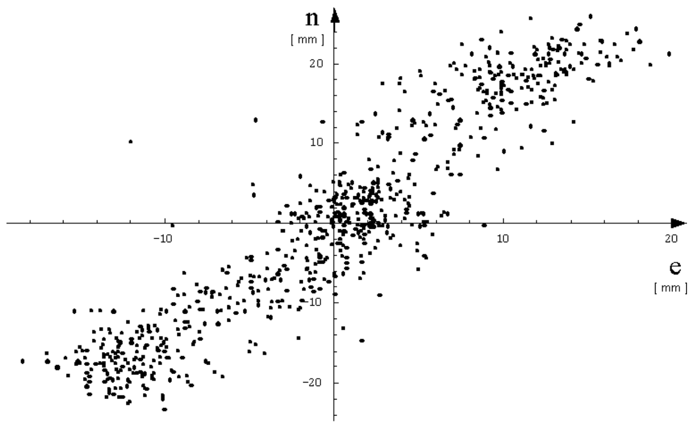


Figure 1: Daily record of point's position in a ground plane.

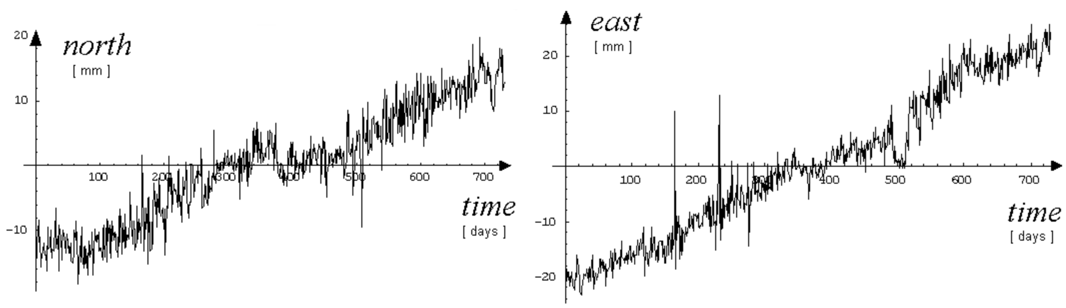


Figure 2: Time plot of point's position variation.

There's easily seen the data following linear trend with a high level of fit. It's a consequence of the long-term drift of the Eurasian tectonic plate, anyway, this overt drift is pretty suitable for applying several approaches of data processing mostly used in mathematical statistics and for showing a plus of the proposed key procedure.

Basically, we may treat our data as (a) two one-dimensional time series or (b) use the fact that both series just reflect the same systematic and random disturbing effects, in other words, they are significantly interconnected.

The first approach has been and still is the most preferred way of time series modelling in general, providing solid results in fitting. However, there is a slightly higher danger here of modelling spurious processes and, consequently, coming to misleading interpretations. The standard procedure includes modelling polynomial (linear) and periodic (seasonal) trends, and then applying Box-Jenkins methodology to cover some residual autocorrelations. This is well described in [12] and we'll re-enter it later in more details.

As for the (b)-group, it's an reasonable tendency of evolution in data processing to look for further relations and to develop more effective techniques (gratefully using computers), such as turning from single equations to vector representation of mathematical relations, etc. Vector regression analysis gives additional information about modelled processes and the way they are linked together (in the form of cross-correlation matrices, basically). We chose this modern approach as the second alternative to be compared in conclusion.

Still staying in the last group, we should introduce a theory largely elaborated by econometricians and given a name "cointegration". For brief explanation, two non-stationary I(1) time series (means integrated of the order 1 ,having the first differences stationary) are cointegrated, if one of their linear combinations is I(0) and hence stationary. There are several tests for cointegration, for details and references see [2],[5]. The most used ones was employed for proving our series to be cointegrated, the procedures are briefly described in section 2.2.2.

2.2 Cointegration

In this section, we perform some tests at first to find out what kind of trend is present in the data and to prove cointegration relation between our two time series. This is essential for applying common trend methodology in the later subsection. The latest one is an extra dose of theory, under which tests for common deterministic trends are applied over observations of several GPS stations.

If speaking about trend, it must be understood there are deterministic and stochastic trend being dealt with in time series theory and are often defined in the context of autoregressive models. Time series generated by deterministic trend (DT) model display mean or trend-reverting behaviour, while those generated by stochastic trend (ST) model lack the reverting forces. An illustrative example of DT model can be $X_t = \delta t + \varepsilon_t$ and ST model $X_t = \delta + X_{t-1} + \varepsilon_t = X_0 + \delta t + \sum_{i=1}^{t} \varepsilon_i$, where ε_t is $\mathrm{N}(0, \sigma^2)$ random process and the model of ST is called random walk (i.e., AR parameter by X_{t-1} equals unity) with drift δ. The term $\sum_{i=1}^{t} \varepsilon_i$ is now called the stochastic trend, but we may see it may be accompanied by (any) deterministic trend component. The key difference is that ST series can deviate from this trend for lengthy period of time.

In our geometrical application, we have good reason to believe the trend in n_t, e_t has deterministic nature, however many times it is not the case.

2.2.1 Testing for stochastic trend

For identifying the stochastic trend, there are two groups of methods

- tests for unit roots
- stationarity tests.

An example of the second set of methods may be KPSS test or LM test. However, here we focus on widely used *augmented Dickey-Fuller* (ADF) tests from the first group (for details see [5], p.80). Procedure starts with choosing the order p of AR(p) model standardly from AIC, BIC information criteria, and continues by performing auxiliary regression

$$\Delta X_t = \mu + \delta t + \rho X_{t-1} + \phi_1 \Delta X_{t-1} + \cdots + \phi_{p-1} \Delta X_{t-p+1} + \varepsilon_t \,, \qquad (1)$$

where X_t represents a variable, t is time, μ, δ, ρ, ϕ are parameters and ε residuals. ρ is the parameter of interest, for which we need to compute test statistic $t(\hat{\rho}) = \hat{\rho}/SE(\hat{\rho})$, SE denotes standard error. The relevant null hypothesis is that $\rho = 0$ against alternative $\rho < 0$ (one-sided test), that means if $t(\hat{\rho}) > t_{critical}$ then we do not reject H_0 of unit root and hence, series contain stochastic trend (ST). Otherwise there is no ST and we may solely think of eventual deterministic trend (DT). Note, that test statistic does not follow standard asymptotic distribution, some critical values are provided, for example, in [5], p.82. The procedure was executed three times, firstly omitting both deterministic components (constant μ and trend δt), then including only constant and finally both of them. Table 1 shows the results, that speak clearly for the primacy of deterministic trend in both time series.

Table 1: Augmented Dickey-Fuller test.

time series	deterministic component	$t(\hat{\rho})$	t_{crit} ($\alpha = 0.05$)	conclusion
n_t	none	-1.70	-1.95	$\rho = 0$ indicates ST
	constant	-1.69	-2.86	$\rho = 0$ indicates ST
	constant & trend	-7.59	-3.41	$\rho < 0$, DT accepted
e_t	none	-0.95	-1.95	$\rho = 0$ indicates ST
	constant	-0.96	-2.86	$\rho = 0$ indicates ST
	constant & trend	-7.19	-3.41	$\rho < 0$, DT accepted

2.2.2 Testing for cointegration

Having found trending behaviour of both our time series (within the framework of AR(p) model), it is natural to investigate whether these I(1) processes are "commonly integrated", i.e., there exist a common trend pattern. There has been devised several methods of testing for cointegration.

The first, *Engle-Granger two steps method* comes from single-equation model of two variables $X_{1,t}, X_{2,t}$ and works as follows. Residuals u_t from static regression

$$X_{2,t} = \beta_0 + \beta_1 X_{1,t} + u_t \tag{2}$$

are used in auxiliary regression

$$\Delta\hat{u}_t = \gamma_0 + \rho\hat{u}_{t-1} + \gamma_1\Delta\hat{u}_{t-1} + \cdots + \gamma_p\Delta\hat{u}_{t-p} + \varepsilon_t \ , \tag{3}$$

and the t-test for the significance of ρ is evaluated. When $\rho = 0$ (that is H_0, $t(\hat{\rho}) > t_{crit}$), u_t has a unit root and thus (2) does not reflect a stationary cointegration relationship. Otherwise, when $\rho < 0$, that is, when $t(\hat{\rho})$ is significantly negative, X_{1t}, X_{2t} are cointegrated. Some critical values are given in [5], page 217, test results in Table 2 shows indisputable presence of cointegration. By the way, R^2 (index of

Table 2: Engle-Granger testing for cointegration

regression of	det.component	$t(\hat{\rho})$	t_{crit} ($\alpha = 0.05$)	conclusion
e_t on n_t	constant	-8.69	-3.37	cointegration
	constant & trend	-9.56	-3.80	
n_t on e_t	constant	-9.71	-3.37	cointegration
	constant & trend	-9.90	-3.80	

determination) by the regression of n_t on e_t is slightly higher, therefore this regression is to be more preferred here.

Engle-Granger is useful when we analyze two time series, but it may become less useful for increasing number of time series. This occurs, e.g, if we decide to include the third coordinate observations. Hence, multivariate methods appear to be more helpful.

To better understand cointegration and all associate terms, let's describe two time series by following VAR(1) model

$$\begin{bmatrix} 1 & \delta \\ 1 & \eta \end{bmatrix} \begin{bmatrix} X_{1,t} \\ X_{2,t} \end{bmatrix} = \begin{bmatrix} \mu_1^* \\ \mu_2^* \end{bmatrix} + \begin{bmatrix} \rho_1 & \delta\rho_1 \\ \rho_2 & \eta\rho_2 \end{bmatrix} \begin{bmatrix} X_{1,t-1} \\ X_{2,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t}^* \\ \varepsilon_{2,t}^* \end{bmatrix} , \tag{4}$$

where $\delta \neq \eta$, μ_1^*, μ_2^* are intercept terms and $\varepsilon_{1,t}^*, \varepsilon_{2,t}^*$ are assumed to be mutually independent white noise error processes. Multiplying both sides with the inverse of the left-hand side matrix and subtracting the one period lagged $\boldsymbol{X}_{t-1}$ from both sides gives

$$\boldsymbol{\Delta X}_t = \boldsymbol{\mu} + \boldsymbol{\Pi X}_{t-1} + \boldsymbol{\varepsilon}_t \ , \tag{5}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\varepsilon}_t$ are functions of $\boldsymbol{\mu}^*$ and $\boldsymbol{\varepsilon}_t^*$, respectively. Interesting matrix is $\mathbf{\Pi}$, because when $0 \leq \rho_i < 1$ for $i = 1, 2$, $\mathbf{\Pi}$ has full rank 2, on the other hand, when $\rho_1 = \rho_2 = 1$, the rank of $\mathbf{\Pi}$ is equal to 0. Now interesting is cointegration case, when (for example) $\rho_1 = 1$ and $0 \leq \rho_2 < 1$, the matrix $\mathbf{\Pi}$ can be written as

$$\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}^\top , \tag{6}$$

where $\boldsymbol{\beta} = [1\ \eta]^\top$ is the cointegration parameters vector. $\boldsymbol{\beta}\boldsymbol{X}_t$ is an equilibrium (or long-run) relation between $X_{1,t}, X_{2,t}$, and the parameter matrix $\boldsymbol{\alpha}$ reflects the speed of adjustment toward equilibrium. Equation (5) incorporating (6) is called a vector error correction model.

Multivariate method of testing for cointegration, proposed above, comes by considering again the VAR(p) model, more convenient if written in error correction format

$$\boldsymbol{\Delta X}_t = \boldsymbol{\mu} + \mathbf{\Gamma}_1 \boldsymbol{\Delta X}_{t-1} + \cdots + \mathbf{\Gamma}_{p-1} \boldsymbol{\Delta X}_{t-p+1} + \mathbf{\Pi} \boldsymbol{X}_{t-p} + \boldsymbol{\varepsilon}_t , \tag{7}$$

where $\mathbf{\Pi}$ contains the information on possible cointegration relations between the m (in our case $m = 2$) elements of $\boldsymbol{X}_t$. If $\mathbf{\Pi}$ is close to rank deficiency, there may be cointegration. The *Johansen's method* is such a statistical method to investigate the rank of $\mathbf{\Pi}$ (which, essentially, amounts to a multivariate extension of the univariate ADF method). The procedure goes like this. First, we perform regressions

$$\begin{aligned} \boldsymbol{\Delta X}_t &= \boldsymbol{a}_0 + \boldsymbol{a}_1 \boldsymbol{\Delta X}_{t-1} + \cdots + \boldsymbol{a}_{p-1} \boldsymbol{\Delta X}_{t-p+1} + \boldsymbol{r}_{0t} , \\ \boldsymbol{X}_{t-p} &= \boldsymbol{b}_0 + \boldsymbol{b}_1 \boldsymbol{\Delta X}_{t-1} + \cdots + \boldsymbol{b}_{p-1} \boldsymbol{\Delta X}_{t-p+1} + \boldsymbol{r}_{1t} , \end{aligned} \tag{8}$$

then construct the matrices $\boldsymbol{S}_{00},\ \boldsymbol{S}_{10},\ \boldsymbol{S}_{11},\ \boldsymbol{S}_{01}$ of $(m \times m)$ from

$$\boldsymbol{S}_{ij} = \frac{1}{n} \sum_{t=1}^{n} \boldsymbol{r}_{it} \boldsymbol{r}_{jt}^\top \qquad \text{for } i, j = 0, 1 . \tag{9}$$

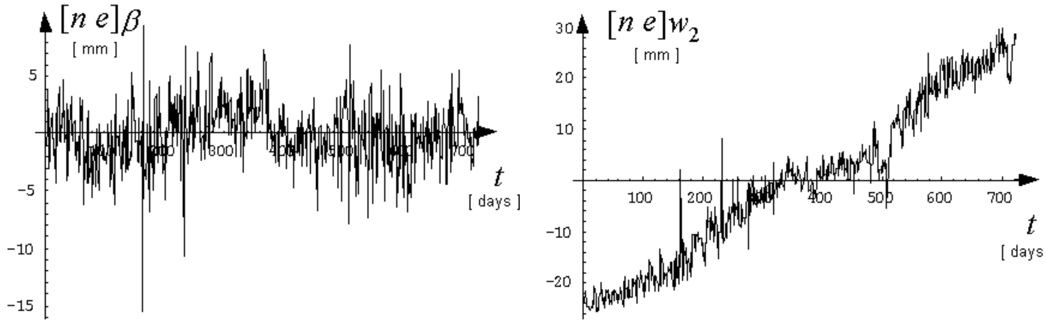
The next step is to solve the eigenvalue problem $|\lambda S_{11} - S_{10} S_{00}^{-1} S_{01}| = 0$ which gives the eigenvalues $\hat{\lambda}_1 \geq \cdots \geq \hat{\lambda}_m$ and the corresponding eigenvectors $\hat{\boldsymbol{\beta}}_1$ through $\hat{\boldsymbol{\beta}}_m$. Now, a test for the rank of $\mathbf{\Pi}$ can be performed by testing how many eigenvalues λ_i equal zero. The first test statistic (which is a likelihood ratio test) $\lambda_{trace} = -n \sum_{i=r+1}^{m} \ln{(1 - \hat{\lambda}_i)}$ tests the null hypothesis of at most r cointegration relations against the alternative there are more of them, while the second test statistic $\lambda_{max} = -n \ln{(1 - \hat{\lambda}_r)}$ can be used to test the null of $r - 1$ against r cointegration relations (vectors). H_0 shall not be rejected if test statistic is smaller than critical value. Table of critical values can be found, e.g., in [5] on page 224, our case evaluation is summarized in Table 3.

Having found 1 cointegration relation $[n_t\ e_t]\hat{\boldsymbol{\beta}}_1$, it's not a bad idea to plot it (see Figure 3a) for later comparison. Anyway, when there are r cointegration relations, among m variables, there has to be $(m - r)$ independent common stochastic

Table 3: Johansen's tests for cointegration ($m = 2$)

r	test	statistic	$\lambda_{crit}(m-r)$ ($\alpha = 0.05$)	conclusion
0	λ_{trace}	127.36	17.95	H_0 rejected
1	λ_{trace}	1.08	8.18	H_1 rejected, r=1 cointegr.vector
1	λ_{max}	126.28	14.90	H_0 rejected
2	λ_{max}	1.08	8.18	H_1 rejected, r-1=1 cointegr.vector

trends in the system. Gonzalo and Granger proposed a method to estimate the stochastic trends, procedure that use the Johansen's but differs in eigenvalue problem $|\lambda \boldsymbol{S}_{00} - \boldsymbol{S}_{01}\boldsymbol{S}_{11}^{-1}\boldsymbol{S}_{10}| = 0$, solution of which has the same eigenvalues $\hat{\lambda}_i$ but different eigenvectors $\hat{\boldsymbol{w}}_1, \ldots, \hat{\boldsymbol{w}}_m$. Because in our case $r = 1$, only one common stochastic trend variable can be constructed (using eigenvector $\hat{\boldsymbol{w}}_{r+1}$), that is $[n_t\ e_t]\hat{\boldsymbol{w}}_2$, plotted in Figure 3b.



a) b)

Figure 3: a) Cointegration relation between n_t, e_t, and b) common stochastic trend.

2.2.3 Geometrical aspect of common trend

Once having found cointegration, it's naturally leading us to investigate a common trend ([11]). We look for linear combination

$$\begin{aligned} y_t &= \gamma_1 n_t + \delta_1 e_t \\ x_t &= \gamma_2 n_t + \delta_2 e_t \end{aligned} \tag{10}$$

such that y represents a common trend direction and x_t is a stationary trend-free variable, orthogonal to y_t. In the light of our geometrical application, it's easy to rewrite a general common trend problem into familiar transformation (in 2D cartesian system)

$$\begin{aligned} y_t &= n_t \cos\alpha + e_t \sin\alpha \\ x_t &= -n_t \sin\alpha + e_t \cos\alpha \end{aligned} \tag{11}$$

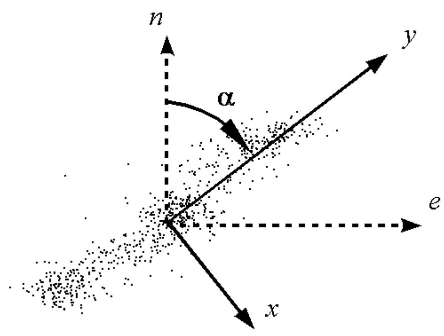


Figure 4: Transformation into common trend direction

as shown in Figure 4. The angle α can be determined either from analysis of stochastic trend

$$e_t = a_0 + b_0 n_t \ , \qquad \tan\alpha = b_0 \tag{12}$$

or analysis of deterministic trend starting at linear regression

$$n_t = a_1 + b_1 t + u_{1,t} \ , \qquad e_t = a_2 + b_2 t + u_{2,t} \ , \tag{13}$$

where t denotes time and a, b regression parameters. If we place (13) into (11) and focus on series x_t, which is supposed to be trend-free, then

$$\begin{aligned} x_t &= -(a_1 + b_1 t + u_{1,t})\sin\alpha + (a_2 + b_2 t + u_{2,t})\cos\alpha \ , \\ x_t &= (a_2\cos\alpha - a_1\sin\alpha) + \underbrace{(b_2\cos\alpha - b_1\sin\alpha)}_{0}\ t + (u_{2,t}\cos\alpha - u_{1,t}\sin\alpha) \end{aligned} \tag{14}$$

(linear trend term in x_t is eliminated), so

$$\tan\alpha = \frac{b_2}{b_1} \ . \tag{15}$$

All right, we have got a new couple of time series y_t, x_t. At this point it is more than interesting to realize that Figures 3 and 5 show the same variables. Figure 5 is situated in section 2.3 where the modelling procedure continues and come to results. Next subsection is a little cutaway showing a useful test for common deterministic pattern on multivariate time series.

2.2.4 Testing for common deterministic trend slopes

In section 2.2.1 we tested our time series separately for the presence of stochastic trend, which might contain constant and linear deterministic component. After modelling both of the components, tests showed no random-walk behaviour in each data set. Now it is of interest to examine if two or more of such a trend-stationary time series have the same slope. More concretely, do all the five concerned points (realized by permanent stations) move to the north with the same upward-moving trend? The question includes also the west-east direction and the velocity in the resultant too.

Such a hypothesis can be written as linear restrictions on the slope parameters across the series and we can apply the multivariate linear trend tests [6]. Consider the multivariate trend model

$$\begin{aligned} z_{1,t} &= \mu_1 + \beta_1 t + u_{1,t} \\ z_{2,t} &= \mu_2 + \beta_2 t + u_{2,t} \\ &\vdots \\ z_{k,t} &= \mu_k + \beta_k t + u_{k,t} \end{aligned} \tag{16}$$

that can be compactly written as $\boldsymbol{z}_t = \boldsymbol{\mu} + \boldsymbol{\beta} t + \boldsymbol{u}_t$, where μ and β are classical constant and linear trend parameters, u denotes residuals and k is the number of time series, in our case $k = 5$. We are interested in testing hypotheses of the form

$$\mathrm{H}_0 : \boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{r}\,, \qquad \mathrm{H}_1 : \boldsymbol{R}\boldsymbol{\beta} \neq \boldsymbol{r},\,, \tag{17}$$

where $\boldsymbol{R}$ is $q \times k$ matrix and $\boldsymbol{r}$ is a $q \times 1$ vector of known constants. The linear hypotheses of (17) are quite general, they include linear hypotheses on slopes within given trend equations ($q = k - 1$) as well as joint trend hypotheses across equations ($q = k$). According to [6] we apply two F-tests, both test statistics are functions of the following HAC (heteroskedasticity autocorrelation) variance covariance matrix estimator. Let $\hat{\boldsymbol{\mu}}$ and $\hat{\boldsymbol{\beta}}$ denote the stacked single equation OLS estimates and $\hat{\boldsymbol{u}}_t = \boldsymbol{z}_t - \hat{\boldsymbol{\mu}} - \hat{\boldsymbol{\beta}} t$ be the residuals. Define

$$\hat{\boldsymbol{\Omega}}_{HAC} = \hat{\boldsymbol{\Gamma}}_0 + \sum_{j=1}^{n-1} (1 - \frac{j}{L})(\hat{\boldsymbol{\Gamma}}_j + \hat{\boldsymbol{\Gamma}}_j^\top)\,, \tag{18}$$

which is the Bartlett kernel estimator, where $\hat{\boldsymbol{\Gamma}}_j = \frac{1}{n}\sum_{t=j+1}^{n} \hat{\boldsymbol{u}}_t \hat{\boldsymbol{u}}_{t-j}^\top$ and L is the truncation lag or bandwidth. Usually a consistent $\hat{\boldsymbol{\Omega}}_{HAC}$ is needed, yet [6] offers an alternative, where $L = n$. Although it does not result in consistent estimator, valid testing is still possible because of asymptotic proportionality and moreover it has certain advantage coming from the choice of bandwidth. It holds that

$$\hat{\boldsymbol{\Omega}}_{L=n} = \frac{2}{n^2} \sum_{t=1}^{n} \hat{\boldsymbol{S}}_t \hat{\boldsymbol{S}}_t^\top\,, \tag{19}$$

where $\hat{\boldsymbol{S}}_t = \sum_{j=1}^{t} \hat{\boldsymbol{u}}_j$. It is also convenient to express an element of $\hat{\boldsymbol{\beta}}$ as

$$\hat{\beta}_i = \left(\sum_{t=1}^{n} \tilde{t}^2 \right)^{-1} \left(\sum_{t=1}^{n} \tilde{t} z_{i,t} \right) \qquad \text{for } i = 1, 2, \ldots, k\,, \tag{20}$$

where $\bar{t} = \frac{1}{n}\sum_{t=1}^{n} t$ and $\tilde{t} = t - \bar{t}$. Now the first of test statistics can be defined

$$F_1 = (\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})^\top \left[\boldsymbol{R} (\sum_{t=1}^{n} \tilde{t}^2)^{-1} \hat{\boldsymbol{\Omega}}_{L=n} \boldsymbol{R}^\top \right]^{-1} (\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})/q\,. \tag{21}$$

Following [6] we also consider an alternative to $\hat{\boldsymbol{\Omega}}_{L=n}$ which is constructed using $\tilde{t}\hat{\boldsymbol{u}}_t$ instead of $\hat{\boldsymbol{u}}_t$. Because $\tilde{t}\hat{\boldsymbol{u}}_t$ is not a vector of stationary time series, establishing consistency of HAC estimator would be difficult if even feasible, yet again if we use $L = n$, the asymptotic behaviour of the HAC estimator can be derived. We can write

$$\tilde{\boldsymbol{\Omega}}_{L=n} = \frac{2}{n^2} \sum_{t=1}^{n} \tilde{\boldsymbol{S}}_t \tilde{\boldsymbol{S}}_t^\top , \tag{22}$$

where $\tilde{\boldsymbol{S}}_t = \sum_{j=1}^{t} (j - \bar{t})\tilde{\boldsymbol{u}}_j$, and then the second test statistic is

$$F_2 = n(\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})^\top \left[\boldsymbol{R}(\frac{1}{n}\sum_{t=1}^{n} \tilde{t}^2)^{-1} \tilde{\boldsymbol{\Omega}}_{L=n} (\frac{1}{n}\sum_{t=1}^{n} \tilde{t}^2)^{-1} \boldsymbol{R}^\top \right]^{-1} (\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})/q \, . \tag{23}$$

The null hypothesis in (17) is rejected if test statistic F_1 (F_2) exceeds critical value given for q restrictions in [6], Table 3 (Table 2, alternatively). It is worth noting that due to practical reasons indices of the F-statistics has been swapped in our work.

The asymptotic distribution theory for these F statistics is nonstandard and was developed for the case where the errors are covariance stationary. Simulation evidence reported by [6] suggests that the F-tests suffers much less from over-rejection problem caused by strong positive serial correlation than the compared standard alternative, whereas the power of F-s is slightly lower. Finite sample simulation evidence in [6] also suggested that the performance of the tests are improved when $\hat{\boldsymbol{\Omega}}$ estimator is computed using VAR(1) prewhitening. However, this we do not do here.

The standard alternative to F_1 and F_2 is a Wald test based on consistent $\hat{\Omega}_{HAC}$ estimator, which uses the same Bartlett kernel. For $\hat{\Omega}_{HAC}$ to be consistent, the bandwidth L must increase as the sample increases but at the slower rate. As referred in [6], the rate $\sqrt[3]{n}$ minimizes the approximate mean square error for $\hat{\Omega}$ and considering this in (18), the Wald test is defined as

$$W = (\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r})^\top \left[\boldsymbol{R}(\sum_{t=1}^{n} \tilde{t}^2)^{-1} \hat{\boldsymbol{\Omega}}_{HAC} \boldsymbol{R}^\top \right]^{-1} (\boldsymbol{R}\hat{\boldsymbol{\beta}} - \boldsymbol{r}) \, . \tag{24}$$

Asymptotic distribution of the Wald test is χ^2 with q degrees of freedom.

For reference we recommend to see an interesting application of this theory in [7] too.

Back to our application, we have got 5 time series from permanent stations denoted in international framework as BOR1(Poland), GOPE(Bohemia), POTS(Germany), HFLK(Austria), PENC(Hungary) and they make vector $\boldsymbol{z}_t$ in respective order. Table 4 contains OLS estimates of parameter β in units mm/year for the north and east coordinate time series n_t, e_t and even for common trend direction time series y_t obtained from (11) with (15).

Table 4: Trend parameter estimates [mm/year]

point:	BOR1	GOPE	POTS	HFLK	PENC
$\hat{\beta}_{n_t}$	14.2	14.7	14.4	13.4	12.2
$\hat{\beta}_{e_t}$	22.3	23.1	21.1	21.5	23.9
$\hat{\beta}_{y_t}$	26.4	27.4	25.5	25.3	26.9

By defining the $q \times k$ matrix $\boldsymbol{R}$ and $q \times 1$ vector $\boldsymbol{r}$

$$\boldsymbol{R} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \qquad \boldsymbol{r} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_1 \\ \hat{\beta}_1 \\ \hat{\beta}_1 \end{pmatrix},$$

the null hypothesis (17) says that the first (reference) time series has the same slope as the rest of time series. Moving the zero column in $\boldsymbol{R}$ and filling $\boldsymbol{r}$ with corresponding parameter, slopes of other series are tested.

We also consider a joint test, when $q = k$, $\boldsymbol{R}$ is k-dimensional identity matrix and $\boldsymbol{r}$ contains trend parameter of an average time series, $\bar{n}_t = \frac{1}{k}\sum_{j=1}^{k} (\boldsymbol{n}_t)_j$ for instance. Test results are collected in Table 5 and 6, critical values for F-tests are given in a separate line bellow. Every Wald test statistic is provided with p-value, which (given null hypothesis is valid) represents a probability that we get the particular value or even the more contradictory to tested hypothesis, in other words it expresses a probability the trend slopes are statistically the same.

First to conclude, all the tests confirmed common deterministic trend in north direction coordinate for all the series except for PENC. Although the joint test for $q = 5$ did not reject the null hypothesis due to quite a large set of variables, from the five tests within equations it is obvious that this last slope does not correspond to the others. As for other two variables, e_t and y_t, no significant deterministic relation was found. Therefore next we took foursomes into account (Table 6) removing seemingly most troublesome series, and observed how it changed the test statistics. Removing PENC accentuated the common trend in n_t, and pointed slightly at promises in y_t. It is useful now to repeat that slope of y_t represents a velocity in the direction a particular point moves, and does not contain information about direction.

At last, we tested triads, of which two appear quite interesting and are summarized in Table 6. First (BOR1, POTS, HFLK) was found to have statistically the same linear trend significant for every direction, though the p-value of Wald test leaves some space for questions. In short and simple, these three points move in the same direction and at the same speed. Second (BOR1, GOPE, PENC) shows no significant harmony in direction, yet it indicates a common velocity.

Honestly, we don't know the cause of this minor effects in tested parameters, there may be some speculations uttered about local instability of given points in particular direction or some residual systematic components in time series that weren't taken

Table 5: F_1, F_2 and Wald test for all 5 time series and selected triads.

station of reference trend	F_1			F_2			W					
	n_t	e_t	y_t	n_t	e_t	y_t	n_t	p-value	e_t	p-value	y_t	p-value
BOR1	17.2	96.6	501.4	16.7	57.7	281.0	0.82	0.935	2.85	0.583	13.88	0.008
GOPE	34.5	68.6	75.4	29.0	95.7	64.6	1.43	0.839	4.73	0.216	3.19	0.526
POTS	22.6	215.1	314.1	19.0	209.3	171.4	0.94	0.919	10.34	0.035	8.46	0.075
HFLK	32.3	159.6	415.7	31.3	188.6	215.8	1.55	0.818	9.31	0.054	10.66	0.031
PENC	143.2	338.8	526.3	198.0	170.0	290.4	9.78	0.044	8.39	0.078	14.34	0.006
Joint	21.5	218.7	423.0	24.5	127.2	218.9	1.50	0.913	7.85	0.164	13.51	0.019
Critical value for q=4 and α=5%: 46.8 (F_1), 43.8 (F_2) and for joint test(q=5): 70.1 (F_1), 78.3 (F_2)												
BOR1	0.6	1.4	1.5	1.0	3.2	3.8	0.03	0.988	0.08	0.962	0.09	0.954
POTS	2.4	4.8	3.0	3.8	6.5	2.9	0.10	0.953	0.16	0.923	0.07	0.964
HFLK	7.9	14.9	11.2	7.9	9.0	7.0	0.19	0.907	0.22	0.895	0.17	0.917
Joint	6.0	33.1	16.5	4.6	33.2	12.8	0.17	0.982	1.23	0.746	0.48	0.924
BOR1	16.4	27.8	4.8	27.9	22.6	5.0	0.69	0.709	0.56	0.756	0.12	0.940
GOPE	9.5	82.6	4.4	24.9	78.3	3.1	0.61	0.735	1.93	0.381	0.08	0.963
PENC	124.2	18.7	4.6	137.7	14.7	4.4	3.40	0.183	0.36	0.834	0.11	0.948
Joint	30.8	117.9	46.4	33.8	129.3	26.4	1.25	0.740	4.79	0.188	0.98	0.807
Critical value for q=2 and α=5%: 40.7 (F_1), 43.8 (F_2) and for joint test(q=3): 68.7 (F_1), 73.4 (F_2)												

Table 6: Joint test for common trend slope of 4-time-series set.

Station removed from set	F_1			F_2			W					
	n	e	y	n	e	y	n	p-value	e	p-value	y	p-value
POTS	25.1	295.0	149.4	36.7	131.4	94.6	1.81	0.771	6.49	0.166	4.67	0.323
HFLK	24.5	120.6	252.7	23.4	119.6	157.3	1.15	0.886	5.91	0.206	7.77	0.100
PENC	11.5	119.3	504.4	12.0	48.4	266.9	0.59	0.964	2.39	0.665	13.18	0.010
Critical value for joint test (q=4 and α=5%): 69.3 (F_1), 76.7 (F_2)												

into account in model specification and which might cause spurious estimates. Nevertheless, although the data visibly show significant linear trend behaviour caused by tectonic plate drift, the tests rejected common deterministic trend for two of the observed points. It is subject to study, why this happened.

2.3 Last comment on computation and results

Here we start at the point of getting common trend direction time series and naturally the trend-free as well. The next step is to model it the same way as two univariate series (the a) approach), at first by subtracting linear trend and seasonal component, then by testing it for residual auto-correlations and applying Box-Jenkins methodology. For comparing purposes we decided to include only annual seasonality and exclude any cyclical component. Figure 5 shows both series fitted by corresponding deterministic model. Correlogram of residuals confirmed the presence

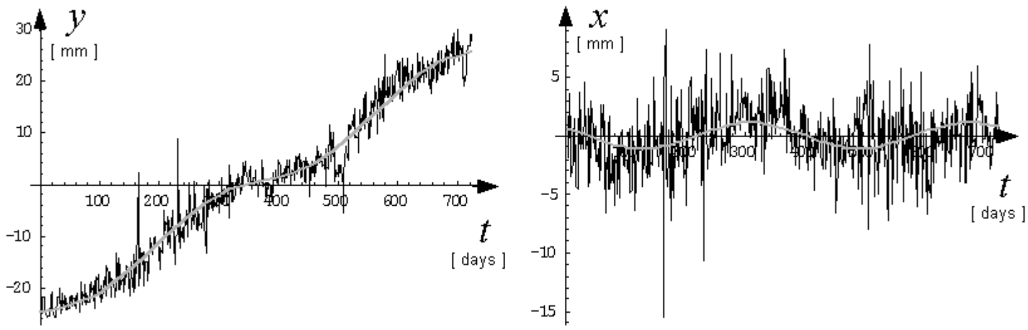


Figure 5: New series y_t and x_t fitted by linear and/or cyclical trend.

of significant correlations. This small residual dependencies may further be modelled by ARMA,ARCH, GARCH or some kind of TAR models, however here we simply employ the more standard AR(p) (autoregressive model of order p) defined

$$y_t = \Phi_0 + \Phi_1\, y_{t-1} + \Phi_2\, y_{t-2} + \cdots + \Phi_p\, y_{t-p} + \varepsilon_t\ , \tag{25}$$

where $\Phi_i\ /i = 1, ...p\ /$ are parameters, ε_t white noise. The order p is chosen either from plot of residuals' variances (Fig. 6a, watch the relative steepness) or information criteria (Fig. 6b, find a minimum), where AIC is Akaike and BIC Schwarz inf.criterion. Taking both results into account, there's no doubt y_t, x_t should be modelled by AR(2) and AR(4), respectively.

Model of y_t, x_t is ready, schematically $y_{m_t}, x_{m_t} = trend + seasonality + AR(p)$, however, this is not a final point we are supposed to come to. The new, model series must be transformed back to (`n, e`) system. If (11) is written in matrix notation, transformation matrix $\mathbf{M}_{\mathtt{n,e}\to\mathtt{y,x}}$ is clearly orthogonal and therefore a backward transformation can easily be performed

$$\begin{pmatrix} n_{m_t} \\ e_{m_t} \end{pmatrix} = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} y_{m_t} \\ x_{m_t} \end{pmatrix}$$

(because $\mathbf{M}_{\mathtt{y,x}\to\mathtt{n,e}} = \mathbf{M}^{-1}_{\mathtt{n,e}\to\mathtt{y,x}} = \mathbf{M}^{\mathsf{T}}_{\mathtt{n,e}\to\mathtt{y,x}}$). For visual review Figure 7 joins original data with the model.

One of the two cardinal purposes of data processing (that's: to understand and be able to forecast) is the next values prediction (Figure 8). It can be utilized well for

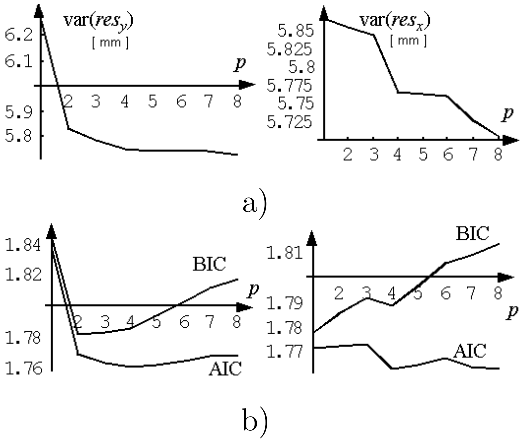


Figure 6: Order p determination: a) residuals' variation, b) information criteria.

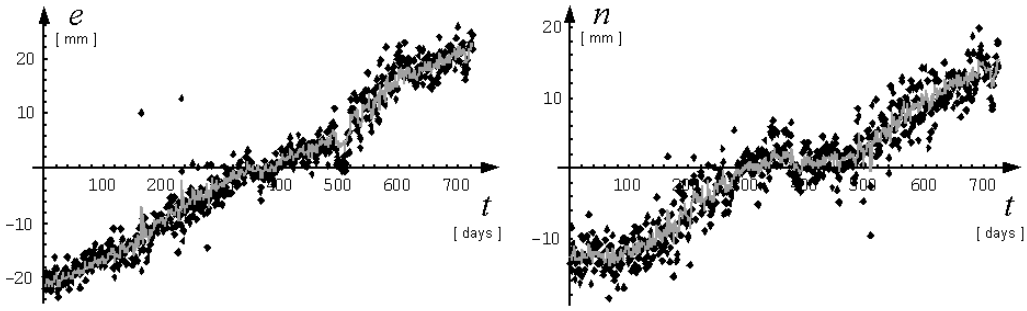


Figure 7: Original time series (black) and model (grey).

comparing the methods. We did it. Having computed model values for next 5 days and got the corresponding GPS measurements, we decided to quantify prediction efficiency by these measures:

$$\text{mean square error} \qquad mse = \frac{1}{k}\sum_{t=1}^{k}(real_t - model_t)^2\,, \tag{26}$$

$$\text{mean percentage error} \qquad mpe = \frac{1}{k}\sum_{t=1}^{k}\frac{real_t - model_t}{real_t}\,100\%\,, \tag{27}$$

where k is a number of predicted time points.

Now we finally come to results. First to mention are the parameters of deterministic model, i.e trend and seasonality, shown in Table 7.

These results are approximately the same for all three methods (excepting those relating to y_t, x_t, of course), and serve for data description. There's pretty seen the quantity of Eurasian tectonic plate long-term drift (25.2mm per year) and the effect of seasonal forces in particular direction, too.

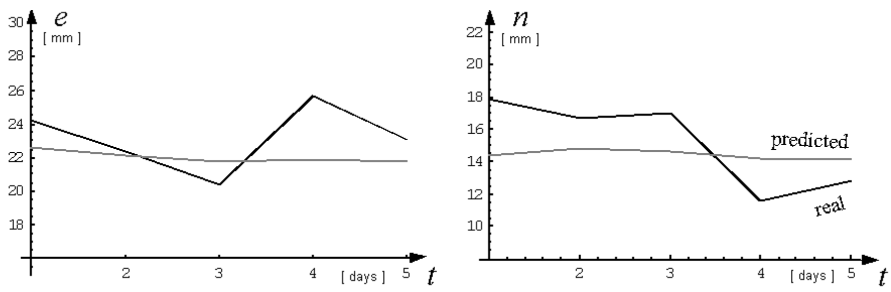


Figure 8: Prediction

Table 7: Deterministic model parameters

trend	n	13.2 mm / year	common trend y 25.2 mm / year	
	e	21.5 mm / year		
seasonality	n	amplitude = 2.2 mm		period = 365days
	e	1.6 mm		
	y	2.5 mm		
	x	1.1 mm		

What is more interesting is certainly in Table 8, which contains results from each method in separate line, namely mean square and mean percentage error of predicted values per variable. This is accompanied by the order of autoregressive model, properly chosen according to information criteria. *Mse* and *mpe* speak

Table 8: Mean square and mean percentage error of predicted values.

method	variable	order p	mse [mm^2]	mpe [%]
1.) independent univariate time series	n	1	7.40	5.08
	e	4	3.70	2.88
2.) multivariate time series	n	2	8.13	5.44
	e	2	5.06	5.13
3.) respecting common trend	n	2 (y)	**5.90**	**4.04**
	e	4 (x)	**4.10**	**2.49**

positively for the method that respects the presence of common trend.

On top of that, if outliers are removed using criterion of triple standard deviation (1% confidence level), better accuracy is attained (Table 9).

Table 9: *mse* and *mpe* of predicted values after removing outliers.

method	variable	order p	mse [mm^2]	mpe [%]
1.) independent univariate time series	*n*	1	7.37	4.94
	e	4	3.34	1.85
2.) multivariate time series	*n*	4	7.52	5.01
	e	4	4.23	4.20
3.) respecting common trend	*n*	4 (y)	**5.59**	**0.50**
	e	4 (x)	**3.33**	**1.97**

3 Nonlinear modelling

In the past, evolution of time series analysis made chiefly for linear models construction - theoretically as well as practically. There were reasons for it: nice theoretical results, possibility to apply variety of achievements grown up on the cultivated fields of linear regression models, relatively simple numerical algorithms and satisfactory practical results. However, despite all this arguments we cannot deny the fact, that many disciplines deals with mechanisms and links of obviously nonlinear character. That's why nonlinear modelling attracts more and more attention. Again, as with linear models it is not the aim to categorize the models here, a very brief overview can be found in [8], pages 5-7, or [4], pages 191-203. Instead, we want to introduce a new trend in modelling that the linear and nonlinear one have in common. The talk is about a multivariate approach. It is being applied to a class of regime-switching models - threshold autoregressive model (TAR) - which is fairly simple and widely used nowadays. For illustration, regime-switching idea may effectively be utilized with daily river flow rate time series which reflects two states of snow in the mountains according to outside temperatures (melting/solid). Of course, the examples are many more, and we presume, that processes influencing our observations may have such nonlinear character.

In the next subsections two components (north and east) of point's position (from permanent GPS observations) in horizontal coordinate system are taken to obtain bivariate time series, which consequently are tested for nonlinearity and modelled using bivariate threshold autoregressive model. Whole procedure, of course, can easily be generalized to more than two-variate series.

3.1 Model introduction

May we have a time series of n time-points, there are several ways to model it. One large family of models that are strongly suitable for modelling stochastic processes, are those arising from Box-Jenkins methodology such as ARMA etc. We will be interested in autoregressive (AR) models, defined

$$y_t = \Phi_0 + \Phi_1\, y_{t-1} + \cdots + \Phi_p\, y_{t-p} + \varepsilon_t\,, \tag{28}$$

where y_t will denote a variable in general, Φ-s are parameters of an AR model and ε_t represents residuals with white noise properties. This is linear model and as such, it may fit only linear dependencies. But what if we know our time series are nonlinear (excluding common trend and seasonality) but piecewise linear, changing it's behaviour by activation of some factor.

We get threshold autoregressive model (TAR), e.g.

$$y_t = \begin{cases} \Phi_1^{(1)}\, y_{t-1} + \cdots + \Phi_p^{(1)}\, y_{t-p} + \varepsilon_t^{(1)} & \text{if } z_{t-d} \le r, \\ \Phi_1^{(2)}\, y_{t-1} + \cdots + \Phi_p^{(2)}\, y_{t-p} + \varepsilon_t^{(2)} & \text{if } z_{t-d} > r, \end{cases} \tag{29}$$

where z_t is a threshold variable, r is a threshold and their relation delimits constituent regimes of the model. Letter d denotes time lag (delay). Because there is often a need to process more than a single vector of measurements at once (sometimes given with some explanatory time series), we will speak about multivariate TAR model

$$\boldsymbol{y}_t = \boldsymbol{\Phi}_0^{(j)} + \sum_{i=1}^{p} \boldsymbol{\Phi}_i^{(j)} \boldsymbol{y}_{t-i} + \boldsymbol{\varepsilon}_t^{(j)} \quad \text{if } r_{j-1} < z_{t-d} \leq r_j, \tag{30}$$

where $\boldsymbol{y}_t = (y_{1t} \ \ldots \ y_{kt})$, $\boldsymbol{\Phi}_0^{(j)}$ is constant term for regime j, and y_{kt} denotes k^{th} univariate time series nested in $\boldsymbol{y}_t$.

As for $\boldsymbol{y}_t$ I put to use GPS observations at permanent station GOPE which are given as point coordinates in horizontal coordinate system (`n, e, v` - north, east and vertical component), in our work just only n_t and v_t are played with (see Figure 9). Usually the components have been processed separately, yet this means a risk

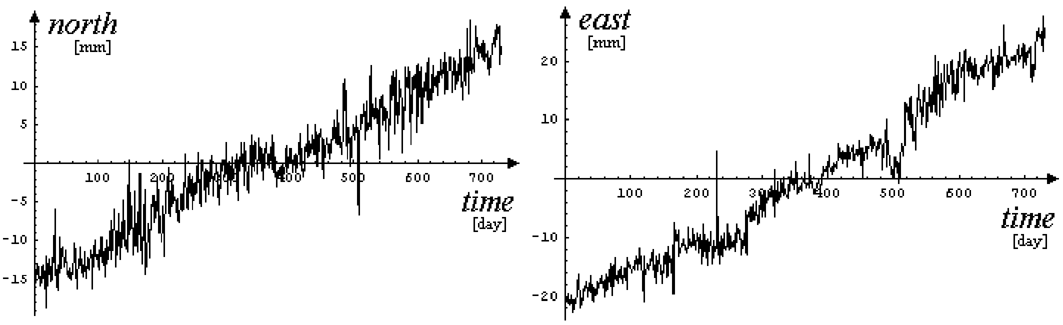


Figure 9: Two vectors of GPS observations of length $n = 730$ days

of some information loss, as they are obviously somehow correlated. That's why we have focused on multivariate modelling.

Now we have data, type of model and we assume that the threshold variable z is known, but the delay d, the order p of AR model and threshold r are not (for simplicity we restrict the case to 2 regimes).

The goal is threefold:

1. To find proper order p of AR model.

2. To make sure, that time series are not linear using test developed by prof. Tsay.

3. To choose the best delay and threshold values, and consequently to build up the final shape of multivariate model.

The order p selection is performed in a classical way on every data set either by

- using a Levinson-Durbin estimation procedure, where for each $p = 1, \ldots p_{max}$ an covariance matrix is computed and their determinants are plotted to find the significant stop of decreasing tendency, or

- employing three information criteria AIC, BIC, HQIC which are to be minimized by the most appropriate order value.

For illustration see Figure 6 or please, refer to [12].

3.2 Testing for nonlinearity

Null hypothesis H_0: $\boldsymbol{y}_t$ is linear.
Alternative hypothesis H_1: $\boldsymbol{y}_t$ follows a threshold model.

Following [13], we utilize standard least square regression framework:

$$\boldsymbol{y}_t = \boldsymbol{A}_t\boldsymbol{\Phi} + \boldsymbol{\varepsilon}_t, \qquad t = h+1, \dots n \tag{31}$$

where $h = \max(p, d)$, $\boldsymbol{A}_t = \begin{pmatrix} 1 & \boldsymbol{y}_{t-1} & \boldsymbol{y}_{t-2} & \dots & \boldsymbol{y}_{t-p} \end{pmatrix}$ is regressor and $\boldsymbol{\Phi}$ denotes parameter matrix. If H_0 holds, then least square estimates are useful, otherwise the estimates are biased under H_1.

Now, let the ordering of the threshold variable z be rearranged increasingly so that $z_{(i)}$ is the smallest element of $S = \{z_{h+1-d}, \dots z_{n-d}\}$ and $t(i)$ is the time index of $z_{(i)}$. Therefore $z_{(i)} = z_{t(i)}$ and the autoregression is

$$\boldsymbol{y}_{t(i)+d} = \boldsymbol{A}_{t(i)+d}\boldsymbol{\Phi} + \boldsymbol{\varepsilon}_{t(i)+d}, \qquad i = 1, \dots n-h\,. \tag{32}$$

It is important to see that the dynamics of the $\boldsymbol{y}_t$ series has not changed (i.e. the independent variable of $\boldsymbol{y}_t$ is $\boldsymbol{A}_t$ for all t). What has changed is the ordering by which the data enter the regression setup. This means an effective transformation of threshold model into a *changepoint* problem.

To detect model change consider the idea:
If $\boldsymbol{y}_t$ is linear, then recursive least squares estimates of the arranged regression is consistent so that the predictive residuals approach white noise (consequently, predictive residuals are uncorrelated with the regressor $\boldsymbol{A}_{t(i)+d}$). Let

$$\hat{\boldsymbol{\eta}}_{t(m+1)+d} = \frac{\boldsymbol{y}_{t(m+1)+d} - \boldsymbol{A}_{t(m+1)+d}\hat{\boldsymbol{\Phi}}_m}{\left[1 + \boldsymbol{A}_{t(m+1)+d}\boldsymbol{V}_m\boldsymbol{A}_{t(m+1)+d}^\top\right]^{1/2}} \tag{33}$$

be the standardized predictive residual of regression (32), where

$$\boldsymbol{V}_m = \left[\sum_{i=1}^{m} \boldsymbol{A}_{t(i)+d}^\top \boldsymbol{A}_{t(i)+d}\right]^{-1}$$

and $\hat{\boldsymbol{\Phi}}_m$ is the estimate of arranged regression (32) using data points associated with the m smallest values of z_{t-d}.

Next, there comes a regression

$$\hat{\boldsymbol{\eta}}_{t(l)+d} = \boldsymbol{A}_{t(l)+d}\boldsymbol{\Psi} + \boldsymbol{w}_{t(l)+d}, \qquad l = m_0+1, \dots n-h\,. \tag{34}$$

where m_0 denotes the starting point of recursive least squares estimation ($m_0 \approx 3\sqrt{n}$). The problem of interest is to test the hypothesis H$_0$: $\boldsymbol{\Psi} = 0$ versus H$_1$: $\boldsymbol{\Psi} \neq 0$ in (34). Tsay [13] designed a test statistic

$$C(d) = [n - h - m_0 - (kp + 1)] \times [\ln(\det S_0) - \ln(\det S_1)], \tag{35}$$

where

$$S_0 = \frac{1}{n - h - m_0} \sum_{l=m_0+1}^{n-h} \hat{\boldsymbol{\eta}}_{t(l)+d}^{\top} \hat{\boldsymbol{\eta}}_{t(l)+d},$$

$$S_1 = \frac{1}{n - h - m_0} \sum_{l=m_0+1}^{n-h} \hat{\boldsymbol{w}}_{t(l)+d}^{\top} \hat{\boldsymbol{w}}_{t(l)+d},$$

and $\hat{\boldsymbol{w}}_t$ is the least square residual of regression (34). Under the null that $\boldsymbol{y}_t$ is linear (and some regularity conditions), $C(d)$ is asymptotically a χ^2 random variable with $k(pk + 1)$ degrees of freedom. If $C(d) < \chi^2_{df}$, we do not reject the null hypothesis.

Table 10: Results of testing for nonlinearity

p (df)	d	$C(d)$	χ^2		p-value
			$\alpha = 0.05$	$\alpha = 0.01$	
2 (10)	1	29.4	18.3	23.2	0.0010
	2	15.1			0.128
	3	23.2			0.010
	4	8.4			0.406
	5	11.9			0.290
	6	15.8			0.104
	7	25.3			0.005
	8	21.9			0.034
	9	13.2			0.213
	10	18.9			0.041
4 (18)	1	41.4	28.9	34.8	0.0014
	2	21.1			0.278
	3	30.2			0.035
	4	14.2			0.281
	5	15.6			0.383

Note. The test is most powerful, if d is correctly specified.

3.3 Model building and results

First we aim at choosing the best values of delay and threshold.

a) One way is to apply conditional least squares estimation.
Assume that p and s (number of regimes) are known, then parameters of model (for now a bit simplified)

$$\boldsymbol{y}_t = \begin{cases} \boldsymbol{A}_t \boldsymbol{\Phi}_1 + \boldsymbol{\Sigma}_1^{1/2} \boldsymbol{a}_t & \text{if } z_{t-d} \leq r, \\ \boldsymbol{A}_t \boldsymbol{\Phi}_2 + \boldsymbol{\Sigma}_2^{1/2} \boldsymbol{a}_t & \text{if } z_{t-d} > r, \end{cases} \tag{36}$$

where $\boldsymbol{a}_t = (a_{1t} \ \ldots \ \boldsymbol{a}_{kt}) \sim \mathrm{N}(\mathbf{0}, \mathbf{I})$,
are $(\boldsymbol{\Phi}_i, \boldsymbol{\Sigma}_i, r, d)$. Putting the possible values of r and d into grid $\{1, 2, \ldots d_0\} \times \{r_{min}, r_{min} + step, \ldots r_{max}\}$ model (36) reduces to two separated multivariate linear regressions from which the least squares estimates of $\boldsymbol{\Phi}_i$ and $\boldsymbol{\Sigma}_i$ $(i = 1, 2)$ are readily available:

$$\hat{\boldsymbol{\Phi}}_i(r,d) = \left(\sum_t^{(i)} \boldsymbol{A}_t^\top \boldsymbol{A}_t\right)^{-1} \left(\sum_t^{(i)} \boldsymbol{A}_t^\top \boldsymbol{y}_t\right) \tag{37}$$

$$\hat{\boldsymbol{\Sigma}}_i(r,d) = \frac{\sum_t^{(i)} \left(\boldsymbol{y}_t - \boldsymbol{A}_t \hat{\boldsymbol{\Phi}}_i^*\right)^\top \left(\boldsymbol{y}_t - \boldsymbol{A}_t \hat{\boldsymbol{\Phi}}_i^*\right)}{n_i - k}, \tag{38}$$

where $\sum_t^{(i)}$ denotes summing over observations on regime i, $\hat{\boldsymbol{\Phi}}_i^* = \hat{\boldsymbol{\Phi}}_i(r,d)$, n_i is number of data points in regime i and k $(k < n_i)$ the dimension of $\boldsymbol{A}_t$. It becomes clear that conditional least squares estimates of r and d should minimize the sum of squares of residuals

$$(\hat{r}, \hat{d}) = \arg\min_{r,d} S(r,d) \tag{39}$$

where $S(r,d) = (n_1 - k)\mathrm{Tr}[\hat{\boldsymbol{\Sigma}}_1(r,d)] + (n_2 - k)\mathrm{Tr}[\hat{\boldsymbol{\Sigma}}_2(r,d)]$.

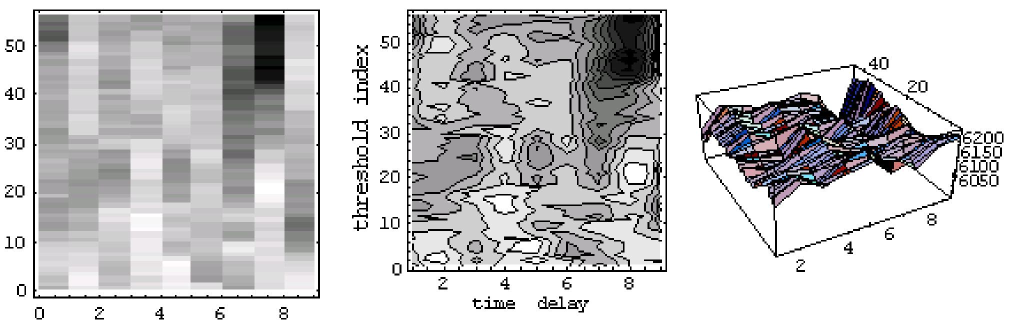


Figure 10: Density, contour and 3D plot of $S(r,d)$; lower axis represents delay d in days, side axis $r \in \langle -2.6, \ 3.0\rangle$ [mm]

Table 11: Results of conditional estimation

p	r [mm]	d [day]	S [mm²]
2	1.89	8	6013.9
	- 0.36	1	6136.5
	- 1.06	1	6137.9
	- 0.35	3	6138.4

b) Besides this, we may apply Akaike information criterion AIC to the same grid $r \times d$.

In fact, it comes along with and supplement the least squares estimation procedure and, of course, there are other parameters defining the multivariate threshold model that could be selected by the criterion

$$AIC(p, s, d, r) = \sum_{(j=1)}^{s} [n_j \ln(\det \hat{\mathbf{\Sigma}}_j) + 2k(kp + 1)] \tag{40}$$

with

$$\hat{\mathbf{\Sigma}}_j = \frac{1}{n_j} \sum_{t}^{(j)} \hat{\boldsymbol{\varepsilon}}_t^{(j)\top} \hat{\boldsymbol{\varepsilon}}_t^{(j)},$$

where n_j is the number of data points in regime j, $\sum_t^{(j)}$ denotes summing over observations in regime j and $\hat{\boldsymbol{\varepsilon}}_t^{(j)}$ are residuals.

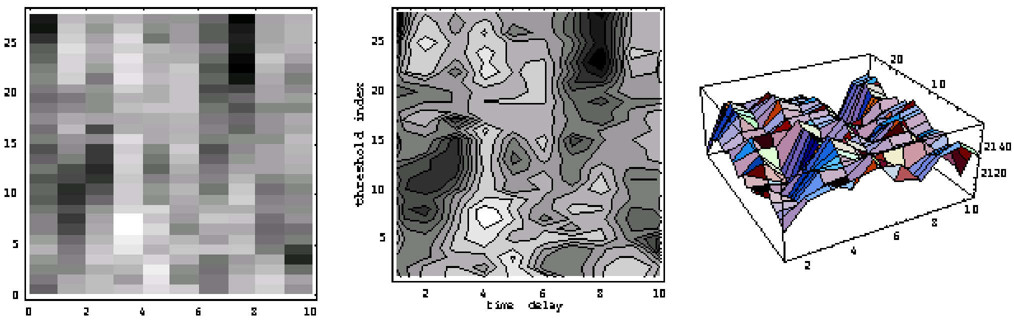


Figure 11: AIC mapped over grid $r \times d$, $r \in \langle -2.6,\ 3.0 \rangle$ [mm], $d \in \{1, 2 \ldots 10\}$ [day]

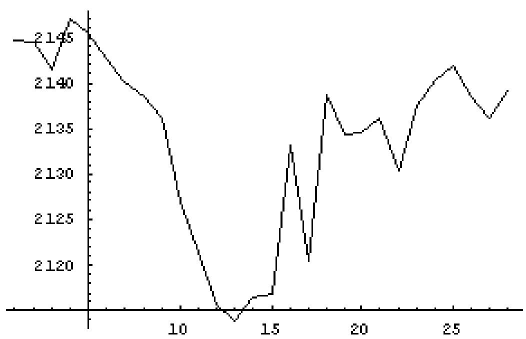


Figure 12: AIC vs. threshold grid index for $d = 3$

There's easily seen a pretty good agreement among the methods, however still partial and shall be the subject to further study. Basically, we prefer those values confirmed by the majority of demonstrated procedures, rather smaller than higher values... but of course, it should depend on practical expectations most.

Table 12: Results of AIC model selection

p	r [mm]	d [day]	AIC
2	1.91	8	2100
	- 0.30	3	2110
	0.25	1	2120
	- 0.35	1	2121

Respecting all previous results, the final shape of model has been selected, built up and is shown in Table 13 and Table 14 and visually compared with original data in Figure 13. However, decision is not so easy and some comparisons to other methods and confrontation with practical purposes are needed.

Table 13: Model variables and characteristics

$p = 2$	$d = 1$ day	$r = -0.35$ mm	$s = 2$ regimes	$z_t = y_{1t}$

Table 14: Parameters and covariance matrices

Φ_1	
- 0.010237	- 0.274496
0.412028	0.0171475
0.005351	0.359622
- 0.017014	- 0.027737
0.053311	0.417337

Φ_2	
0.152080	0.166899
0.226559	0.033515
- 0.108756	0.492913
0.185507	0.041789
0.001387	0.236399

Σ_1 [mm²]	
4.736	- 0.287
- 0.287	3.194

Σ_2 [mm²]	
4.399	- 0.898
- 0.898	4.692

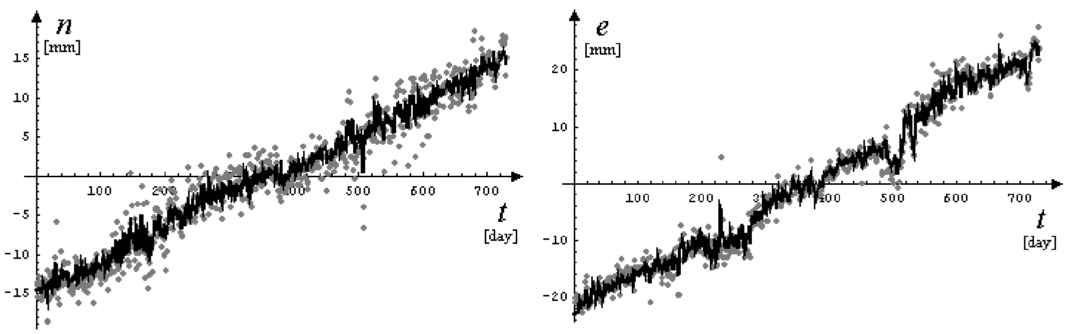


Figure 13: Visualized fit of the built model. Original data are represented by dotted, model by joined plot of a) north and b) east component.

Here we have shown one possible way of processing of geodetic data, that may be extended to three-or-more-regimes models and models including some exogenous variables, but what should be considered about the proposed procedure as the major contribution to time series analysis applied in geodesy is treating the data as set of mutually depending variables effectively describable by multivariate modelling approach rather than by the univariate one.

4 Conclusion

Multivariate modelling methods count doubtlessly among the modern approaches in time series processing, allowing us to capture wider area of causalities, which are of our interest. Supported by powerful data collecting technologies such like GPS, we are able to model relations of time space entities, namely the varying geometrical positions of points. They reflects processes that interact locally, regionally and globally. We concentrated here on permanent GPS observations, that serves for regular monitoring of the Earth's crust kinematics (and for other research purposes). The time series of north and east horizontal coordinates clearly show the long-term drift of Eurasian tectonic plate as linear deterministic trend, which is common to every point in the area. Besides this, the time series visibly reflect other effects, both common and unique. They can be mutually distinguished by the methods of multivariate modelling as proposed in this thesis.

We have introduced some contemporary linear modelling techniques as well as nonlinear ones, most of which originates from econometrics. Firstly the statistical tests to prove the presence of stochastic or even deterministic trend, then the presence of a common trend, in other words, we have shown that two time series of one point's coordinates are commonly integrated, so that they can be decomposed to stationary time series of cointegration relation and common trend series. This decomposition were also geometrically demonstrated by two-dimensional linear transformation and utilized for further modelling in order to compare a forecast performance with alternatives, classic multivariate and univariate method. There was also performed a test for common deterministic trend applied to particular coordinate over five points, i.e., north, east and common trend direction.

As for nonlinear models, we utilized multivariate threshold autoregressive model, restricted to two regimes and applied to trendy ground coordinates time series as in previous. It is easily extendable to more than two variables and regimes as well.

5 Theme and propositions of the thesis project

Theme of the thesis:

Contemporary methods of time series modelling and their application in geodesy

Propositions of the thesis:

- Cointegration
  - Testing for common trends
  - Transformation into common trend direction
- Testing for common periodic components
- Multivariate threshold autoregressive model
  - with exogenous threshold variable
  - using aggregation operators
- Threshold cointegration
- Multivariate autoregressive modelling using copulas
- Application in geodesy

6 Objective and prospective contribution

Thesis objective:

To examine common behaviour of multivariate set of data by statistical tests and to utilize new progressive procedures in their processing, focusing on permanent GPS observations

Prospective contribution:

- Incorporating the theory of cointegrated processes into geodetic data models should reveal more of the character of surveyed phenomenons.
- Tests for common features should mathematically support the specialist's opinion about data when choosing the most appropriate model.
- Multivariate approach has an obvious ambition to be the preferred way of automatized processing on computers.
- Statistical tests helps to reveal nonlinearity in time series behaviour so that threshold models can manage some large "awkward" shocks, for instance.
- Joining the idea of cointegration and threshold should in proper case improve performance of final model.
- The new theoretical approach that make use of copulas is expected to widen a specialist's horizon of promising modelling methods.
- All the proposed theories are new to implement useful procedures into geodetic data processing.

7 References

1. Arlt, J., Arltová, M.: Finanční časové řady, Grada publishing, Praha (2003).

2. Bognár, T.: Cointegration and coordinate transformation applied for geodetical time series, Journal of Electrical Engineering 12/S, vol. 55 (2004), 25–28.

3. Bruyninx, C., Kenyeres, A., Takacs, B.: EPN data and product Analysis for improved velocity estimation: First results, International Association of Geodesy Symposia, vol. 125 (2001)

4. Cipra, T.: Analýza časových řad s aplikacemi v ekonomii, Alfa, Praha (1986).

5. Franses, P.H.: Time Series Models for Business and Economic Forecasting, Cambridge University Press, Cambridge (1998).

6. Franses, P.H., Vogelsang, T.J.: Testing for Common Deterministic Trend Slopes, Center for Analytic Economics Working Paper 01-15, Cornell University (2002).

7. Fomby, T.B., Vogelsang, T.J.: Tests of Common Deterministic Trend Slopes Applied to Quarterly Global Temperature Data, Advances in Econometrics Volume 17, Chapter 2, Elsevier Science (2003), 29–43.

8. Granger, C.W.J., Teräsvirta: Modelling Nonlinear Economic Relationships, Oxford University Press, Oxford (1993).

9. Hefty, J.: The permanent Modra-Piesok GPS station and its long-term and short-term stability, Slovak Journal of Civil Engineering (2001), 31–37.

10. Hefty, J., Husár, L.: Družicová geodézia/Globálny polohový systém, Slovenská technická univerzita v Bratislave (2003).

11. Komorníková, M., Komorník, J.: Time series models for Earth's crust kinematics, Kybernetika, vol.38 (2002), n.3, 383-387.

12. Reference and user's guide to Time series pack for Mathematica, Wolfram Research (1995).

13. Tsay, R.S.: Testing and modeling multivariate threshold models, Journal of the American Statistical Association, vol. 93 (1998), 1188–1202.

Publications

Bacigál, T.: Multivariate Threshold Autoregressive Models in Geodesy, Journal of Electrical Engineering 12/S, Volume 55 (2004), 91–94.

Bacigál, T.: Modeling points position time series with respect to common trend, Zborník príspevkov konferencie PRASTAN 2004, Kočovce 17–21. 5. 2004, 5-10.

Bacigál, T., Komorníková, M.: Modelling points position time series in the light of cointegration, Proceedings of INGEO 2004 (CD-ROM edition), INGEO 2004 Bratislava 11–13. 11. 2004.

Bacigál, T., et al.: Variance components estimate in 2D geodetic network, Proceedings of INGEO 2004 (CD-ROM edition), INGEO 2004 Bratislava 11-13. 11. 2004.

The thesis project follows the assignment of grants:

- VEGA 1/1033/04
- VEGA 1/0273/03
- APVT - 20 - 003204