Modeling point's position time series with respect to common trend

Tomáš Bacigál
KMaDG SvF STU, Radlinského 11, Bratislava 831 68
bacigal@math.sk

1. Introduction

Many technical disciplines involved in civil engineering, such as geology, geodesy, statics of structures and others deal with position of particular points in time-space 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. In this paper we'll 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. Observations had been performed daily in years 2001-2002 on GPS permanent station Borowiec (BOR1, Poland) which takes part in EUREF Permanent Network representing a regional densification of global IGS net in Europe which is used, among other purposes, for regular monitoring of recent kinematics of the Earth's crust (see [2]). 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 in the mean position of the two year period - to be further processed.

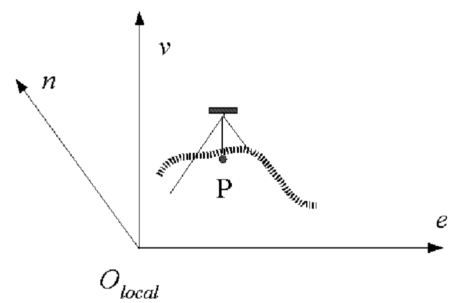

Figure 1: Scheme of GPS installation and topocentric horizontal coordinate system

Because of significantly lower precision and negligible linear trend in vertical direction, we only deal here with the two time series n and e each containing 730 data points. Figure 2 shows two dimensional representation of point variation on Earth's surface and Figure 3 time plot for each coordinate.

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.

2. Data processing

Basically, we may treat our data as (a) two independent 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.

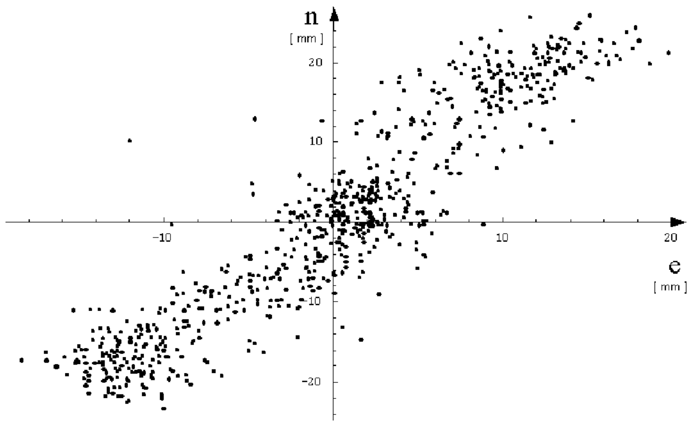

Figure 2: Daily record of point's position in a ground plane.

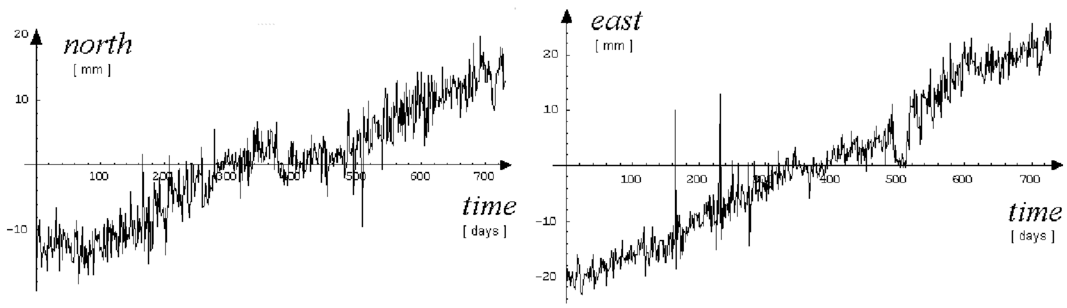

Figure 3: Time plot of point's position variation.

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 [4] 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 [1]. The most used ones was employed for proving our series to be cointegrated.

Once having found cointegration, it's naturally leading us to investigate a common trend ([3]). We look for linear combination

$$\begin{aligned} y &= \gamma_1 n + \delta_1 e \\ x &= \gamma_2 n + \delta_2 e \end{aligned} \qquad (1)$$

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

$$\begin{aligned} y &= n\cos\alpha + e\sin\alpha \\ x &= -n\sin\alpha + e\cos\alpha \end{aligned} \tag{2}$$

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

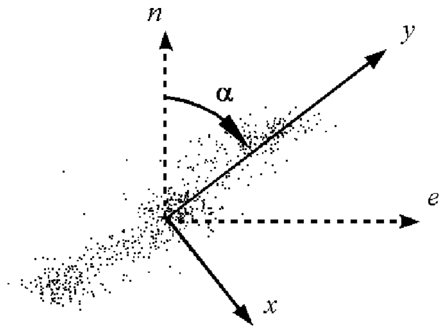

Figure 4: Transformation into common trend direction

tic trend

$$e_t = a_0 + b_0 n \ , \qquad \tan\alpha = b_0 \tag{3}$$

or analysis of deterministic trend starting at linear regression

$$n_t = a_1 + b_1 t + e_{1,t} \ , \qquad e_t = a_2 + b_2 t + e_{2,t} \ , \tag{4}$$

where t denotes time and a, b regression parameters. If we place (4) into (2) and focus on series x, which is supposed to be trend-free, then

$$\begin{aligned} x_t &= -(a_1 + b_1 t + e_{1,t})\sin\alpha + (a_2 + b_2 t + e_{2,t})\cos\alpha \ , \\ x_t &= (a_2\cos\alpha - a_1\sin\alpha) + \underbrace{(b_2\cos\alpha - b_1\sin\alpha)}_{0}\, t + (e_{2,t}\cos\alpha - e_{1,t}\sin\alpha) \end{aligned} \tag{5}$$

(linear trend term in x is eliminated), so

$$\tan\alpha = \frac{b_2}{b_1} \ . \tag{6}$$

All right, we have got a new couple of time series y, x. The next step is to model it the same way as two univariate series ((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 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

$$y_t = \Phi_0 + \Phi_1\, y_{t-1} + \Phi_2\, y_{t-2} + \ldots + \Phi_p\, y_{t-p} + \varepsilon_t \ , \tag{7}$$

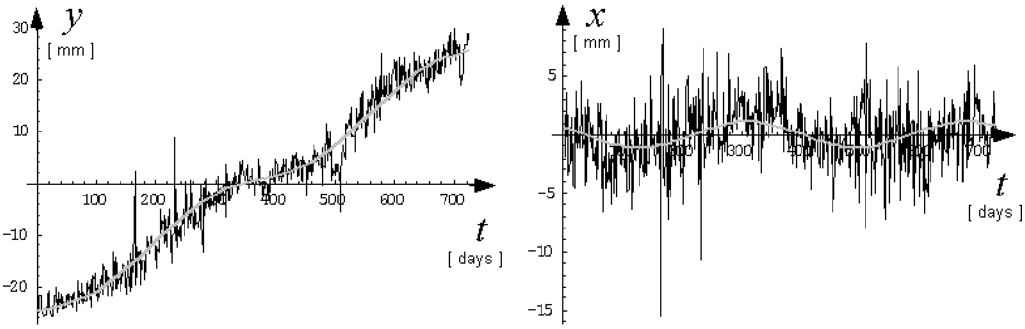

Figure 5: New series y and x fitted by linear and/or cyclical trend.

where Φ_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

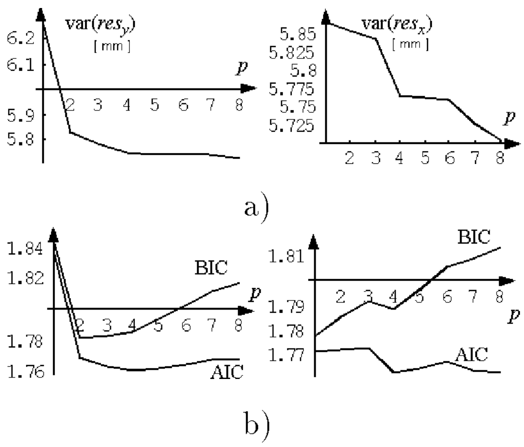

Figure 6: Order p determination: a) residuals' variation, b) information criteria.

inf.criterion. Taking both results into account, there's no doubt y, x should be modelled by AR(2) and AR(4), respectively.

Model of y, x is ready, schematically $y_m, x_m = trend + seasonality + AR(p)$, 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 (2) is written in matrix notation, transformation matrix $\mathbf{M}_{n,e->y,x}$ is clearly orthogonal and therefore a backward transformation can easily be performed

$$\begin{pmatrix} n_m \\ e_m \end{pmatrix} = \begin{pmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} y_m \\ x_m \end{pmatrix}$$

(because $\mathbf{M}_{y,x\to n,e} = \mathbf{M}^{-1}_{n,e\to y,x} = \mathbf{M}^{\mathrm{T}}_{n,e\to y,x}$). 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 (Fig. 8). It can be utilized well for 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:

$$\text{mean square error} \qquad mse \;=\; \frac{1}{k}\sum_{t=1}^{k}(real_t - model_t)^2\,, \qquad (8)$$

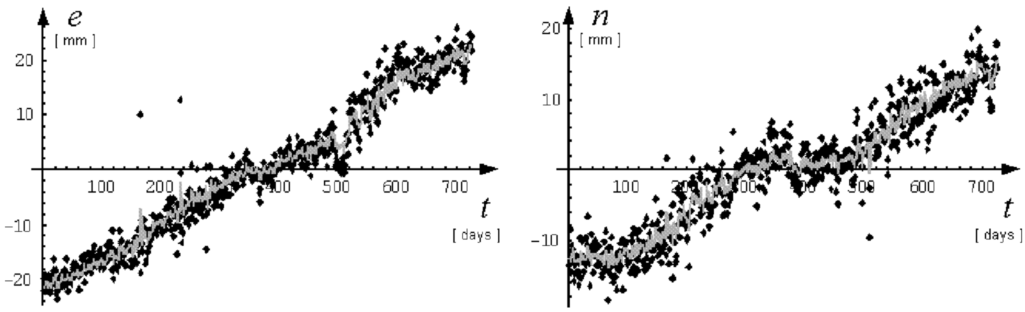

Figure 7: Original time series (black) and model (grey).

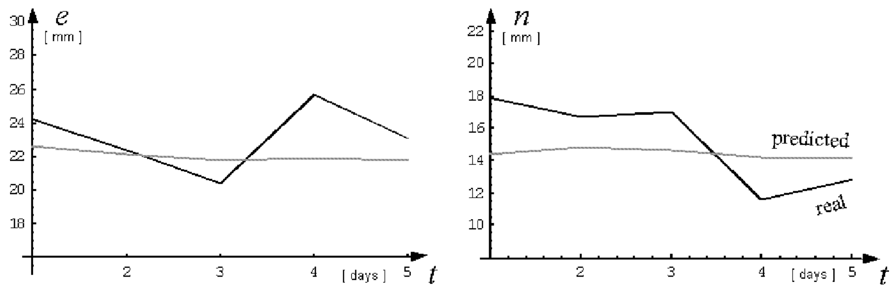

Figure 8: Prediction

$$\text{mean percentage error} \qquad mpe \;=\; \frac{1}{k}\sum_{t=1}^{k}\frac{real_t - model_t}{real_t}\,100\% \;, \tag{9}$$

where k is a number of predicted time points.

3. Results

First to mention are the parameters of deterministic model, i.e trend and seasonality, shown in Table 1.

➧ 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		

Table 1: Deterministic model parameters

These results are approximately the same for all three methods (excepting those relating to y, x, 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.

What is certainly more interesting is in Table 2, 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 positively for the method that respects the presence of common trend.

However, if outliers are removed using criterion of triple standard deviation (1% confidence level), better accuracy is attained (Figure 3).

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**

Table 2: Mean square and mean percentage error of predicted values.

method	variable	order p	mse [mm^2]	mpe [%]
1.) independent univariate time series	n	1	7.37	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**

Table 3: mse and mpe of predicted values after removing outliers.

Acknowledgement. Research supported by grants APVT-20-023402, VEGA 1/0273/03 and VEGA 1/0033/04

References

1. Bognár, T.: Cointegration and coordinate transformation applied for geodetical time series, forthcoming Journal of Electrical Engineering (2005), ISCAM 2004.

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

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

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