On Analysis and Forecasting of Surface Movement and Deformation: Some AR-Models and Their Application

Olga Waelder

1 Mathematical Modeling

We assume two independent variables *X, Y* (forecasting factors) describing pressure and temperature, which influence the dam deformation *Z*. This restriction concerning the use of only two variables is made related to the further model applications to the two-dimensional data set that is given in Yuanzhong and Litao (2005). Obviously, our models can be simply generalised for more independent variables. We consider here four following models:

$$Z_1(k) = C + \sum_{i=1}^{p} \alpha_i X(k-i) + \sum_{i=1}^{p} \beta_i Y(k-i) + \varepsilon \qquad (1)$$

$$Z_2(k) = C \cdot Z_2(k-1) + \sum_{i=1}^{p} \alpha_i X(k-i) + \sum_{i=1}^{p} \beta_i Y(k-i) + \varepsilon$$

$$Z_3(k) = C + \sum_{i=0}^{p} \alpha_i X(k-i) + \sum_{i=0}^{p} \beta_i Y(k-i) + \varepsilon$$

$$Z_4(k) = C \cdot Z_4(k-1) + \sum_{i=0}^{p} \alpha_i X(k-i) + \sum_{i=0}^{p} \beta_i Y(k-i) + \varepsilon$$

The random variable $\varepsilon \sim N(0, \sigma^2)$ is normally distributed with mean zero and (unknown) variance σ^2. Parameter p describes the so called "depth" of the recursion of AR-models (it corresponds to the number of months which should be taken into consideration in our application).
With help of methods of the adjustment theory we can estimate the unknown coefficients C and α_i, β_i, $i = 0(or\ 1)..p$ using a sufficient number M of equations corresponding to the given measurements $x(j)$, $y(j)$, $z(j)$, $|j| = M$. M is the number of months, when the measurements are sampled. This number M should satisfy $M \leq 2p+1$ for models 1 and 2, $M \leq 2p+3$ for models 3 and 4. For example, we have for model 1:

$$z_1(p+1) = C + \sum_{i=1}^{p} \alpha_i x(p+1-i) + \sum_{i=1}^{p} \beta_i y(p+1-i) + \varepsilon$$

$$z_1(p+2) = C + \sum_{i=1}^{p} \alpha_i x(p+2-i) + \sum_{i=1}^{p} \beta_i y(p+2-i) + \varepsilon \qquad (2)$$

$$\ldots$$

$$z_1(p+M) = C + \sum_{i=1}^{p} \alpha_i x(p+M-i) + \sum_{i=1}^{p} \beta_i y(p+M-i) + \varepsilon$$

After setting these equations to matrix form like (2')

$$\bar{z} = A \cdot \bar{u} + \bar{\varepsilon},$$

$$\bar{z}^T = [z(p+1), \ldots, z(p+M)], \quad \bar{u}^T = [C, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_p],$$

$$A(k,1) = 1, k = 1..M \qquad (2')$$

$$A(k, l+1) = x(p+k-l), \quad k = 1..M, \quad l = 1...p$$

$$A(k, l+p+1) = y(p+k-l), \quad k = 1..M, \quad l = 1..p$$

we get the well-known solution of (2') corresponding to:

$$\bar{u} = (A^T A)^{-1} A^T \bar{z} \qquad (3)$$

The accuracy of the model fitting is obtained by:

$$\hat{\sigma} = \sqrt{\frac{(A\bar{u} - \bar{z})^T (A\bar{u} - \bar{z})}{M - 2p - 1}} \qquad (4)$$

The statistical goodness-of-fit can be proved by the empirical mean and correlation coefficient between real measurements and their forecasted values, which can be calculated using the proposed mathematical models. It should be expected, that:

$$E(\varepsilon) \sim \hat{\varepsilon} = \bar{z} - A\bar{u} \approx 0 \qquad (5)$$

$$\hat{\rho} = \rho(A\bar{u}, \bar{z}) \approx 1$$

Models 2, 3, 4 can be analogously handled. The value $\hat{\sigma}^2$ from (4) can be used as an estimator for the unknown variance σ^2 of $\varepsilon \sim N(0, \sigma^2)$ and applied for forecasting discussed below.

2 Case study: Monitoring movement and deformation of a gangue dam (a gold mine in Shandong Province, Chine)

Here, we use the data set from Table 1 in Yuanzhong and Litao (2005), p. 91. The monthly dam deformations are presented in Figure 1.
At first, we discuss modeling and forecasting the deformation for the last four months (K = 30, 31, 32 and 33 from Table 1 in Yuanzhong and Litao (2005), p. 91) based on measurements obtained during months 1–29, see Figure 1 (left).

Table 1: Real measurements and their forecasting for the last four months (cited from Yuanzhong and Litao (2005), p.91, 92)

Month number, K	30	31	31	33
Deformation, *real*, mm	38	54	50	79
Deformation, *forecasted*, mm	45.3	48.0	52.6	49.6

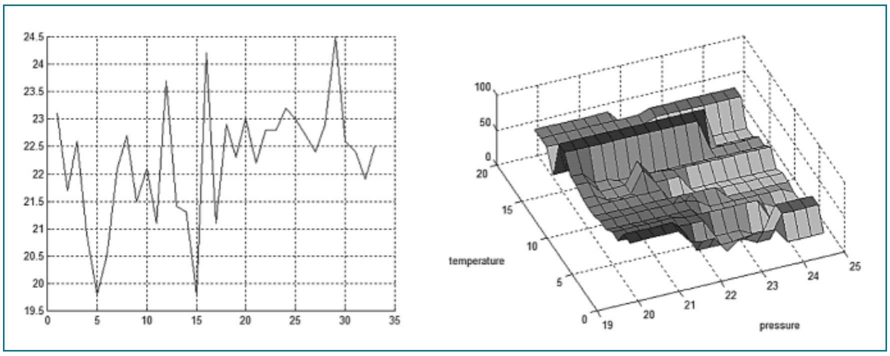

Fig. 1: The monthly deformations (left) and deformations related to pressure and temperature (right)

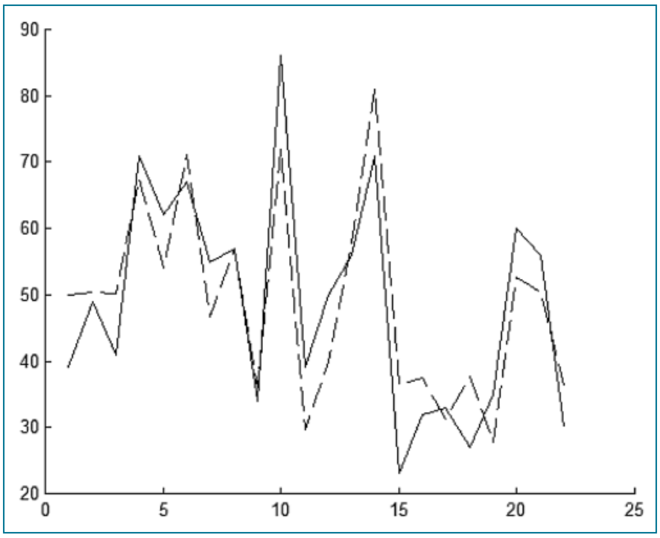

Fig. 2: The real deformations (solid line) and their forecasting values (dotted line) by model 1 from (1) with $p = 7$

Real measurements and their forecasted values corresponding to the approach described in Yuanzhong and Litao (2005) are presented in Table 1.

Remark 1: In this case the sum of absolute differences between the real and forecasted deformations is $d = 45.3$. We apply models 1–4 from (1) for different values of parameter p and obtain the coefficients of the corresponding multiple linear regression given in Tables 2–5. Figure 2 shows a comparison between real measurements and their estimated values.

The forecasting is modelled as follows. At first, we use a multiple regression model for the trend prediction. At second, we simulate 100 times a normally distributed parameter ε with mean zero and variance $\hat{\sigma}^2$ and use the obtained maximal *absolute* values for calculating the limits of the forecasting interval. Of course, one can also use the $1.5 \cdot \sigma$-rule for these limits.

In Table 6 the forecasted values and the sum of their absolute differences between real measurements (see Table 1) and their forecasted values are presented.

3 Outlook and discussion

It can be seen from Tables 1 and 6, that a long-time forecasting is mostly useless. This fact is well-known in approximation theory. It is more meaningful to make forecasting only for the single month 30 based on last 29 months, than for four future months 30–33.

Further, if additional information about pressure, temperature and deformation in the month 30 is still given, one should fit a new multiple regression model based on the months 1 –30, and then forecast the value of deformation in the month 31 and so on. The choice of a multiple regression model 1–4 from (1) should be strongly depending on the expert knowledge about the "true nature" of a dam deformation process.

It is clear, that other regression models are possible. For example, one can use a generalized model like (6):

$$Z_5(k) = C + \sum_{i=0}^{p} \alpha_i \frac{X(k-i)}{Y(k-i) + eps} + \sum_{i=0}^{p} \beta_i Y(k-i) + \varepsilon \qquad (6)$$

The parameter *eps* helps to correct zeros of temperatures. We use $eps = 9$ here because this value leads to the minimum of d, c.f. Figure 4. In this case the following results are obtained, c.f. Table 6:

$\hat{p} = 0.72$, $\hat{\varepsilon} = -0.028$ mm $\times 10^{-9}$ and $d = 25.94$, c.f. Remark 1. The parameter $p = 4$ leads to $\hat{\sigma} = 14.47$.

The forecasted values for the months $K = 30 - 33$ are: 39.70 ± 21.71, 44.66 ± 21.71, 50.49 ± 21.71, 64.59 ± 21.71.

Figure 4 shows the values of d depending on the choice of *eps* in (6). We can see, that the optimal value corresponds to approx. $eps = 9$. Figure 5 presents the result of forecasting for the month 30 using this model, c.f. Figure 3.

Finally, the "point-exactly" forecasting can never exist, because of unknown, future process oscilations caused by the random parameter ε.

Graphically, interval-related, short-time forecasting can be presented as in Figure 3. The mathematical software MATLAB is used for numeric implementation and visualisation.

Table 2: The coefficients of the multiple linear regression and some goodness-of-fit characteristics c.f. (4), (5) for the model 1 from (1) based on measurements from the first 29 months

$p=2$	$p=3$	$p=4$	$p=5$	$p=6$	$p=7$
$C=108.31$	$C=48.25$	$C=81.65$	$C=17.81$	$C=143.85$	$C=132.98$
$\alpha_1=1.75$	$\alpha_1=0.62$	$\alpha_1=0.19$	$\alpha_1=-0.64$	$\alpha_1=-1.17$	$\alpha_1=-0.18$
$\beta_1=-0.92$	$\beta_1=-0.95$	$\beta_1=-0.90$	$\beta_1=-0.97$	$\beta_1=-0.68$	$\beta_1=-0.69$
$\alpha_2=-3.77$	$\alpha_2=-3.61$	$\alpha_2=-3.31$	$\alpha_2=-2.0$	$\alpha_2=0.5$	$\alpha_2=2.35$
$\beta_2=-1.18$	$\beta_2=-1.49$	$\beta_2=-1.46$	$\beta_2=-1.6$	$\beta_2=-1.45$	$\beta_2=-1.4$
	$\alpha_3=3.77$	$\alpha_3=4.01$	$\alpha_3=3.9$	$\alpha_3=0.72$	$\alpha_3=0.36$
	$\beta_3=0.09$	$\beta_3=0.21$	$\beta_3=0.21$	$\beta_3=0.48$	$\beta_3=0.31$
		$\alpha_4=-1.68$	$\alpha_4=-2.47$	$\alpha_4=-2.21$	$\alpha_4=-2.43$
		$\beta_4=0.01$	$\beta_4=-0.26$	$\beta_4=-0.5$	$\beta_4=-0.63$
			$\alpha_5=3.51$	$\alpha_5=5.17$	$\alpha_5=5.25$
			$\beta_5=-0.19$	$\beta_5=0.52$	$\beta_5=0.62$
				$\alpha_6=-6.86$	$\alpha_6=-8.00$
				$\beta_6=0.23$	$\beta_6=0.34$
					$\alpha_7=-0.9$
					$\beta_7=0.49$
$\hat{\sigma}=12.03$	12.23	13.11	13.47	12.88	13.94
$\hat{\varepsilon}=-0.29$ mm $\times 10^{-9}$	0.14	0.25	0.02	– 0.29	– 0.32
$\hat{\rho}=0.70$	0.74	0.74	0.78	0.85	0.88

Table 3: The first coefficient of the multiple linear regression and some goodness-of-fit characteristics c.f. (4), (5) for the model 2 from (1) based on measurements from the first 29 months

$p=2$	$p=3$	$p=4$	$p=5$	$p=6$	$p=7$
$C=-0.03$	$C=-0.07$	$C=-0.06$	$C=-0.1$	$C=0.14$	$C=0.01$
$\hat{\sigma}=12.69$	12.32	13.34	13.42	13.40	14.37
$\hat{\varepsilon}=-3.44$ mm	– 1.32	– 1.35	– 0.31	– 1.16	– 0.63
$\hat{\rho}=0.66$	0.73	0.73	0.79	0.83	0.87

Table 4: The first coefficient of the multiple linear regression and some goodness-of-fit characteristics c.f. (4),(5) for the model 3 from (1) based only on measurements from the first 29 months

$p=1$	$p=2$	$p=3$	$p=4$	$p=5$	$p=6$
$C=139.72$	$C=169.64$	$C=141.43$	$C=113.46$	$C=120.38$	$C=232.08$
$\hat{\sigma}=10.59$	11.11	11.69	12.16	12.56	12.47
$\hat{\varepsilon}=-0.09$ mm $\times 10^{-9}$	0.23	0.03	0.32	0.01	0.54
$\hat{\rho}=0.77$	0.78	0.79	0.81	0.85	0.89

Table 5: The first coefficient of the multiple linear regression and some goodness-of-fit characteristics c.f. (4), (5) for the model 4 from (1) based on measurements from the first 29 months

$p=1$	$p=2$	$p=3$	$p=4$	$p=5$	$p=6$
$C=0.18$	$C=0.19$	$C=0.14$	$C=0.23$	$C=0.24$	$C=0.26$
$\hat{\sigma}=11.75$	12.55	12.39	12.40	12.79	13.93
$\hat{\varepsilon}=-4.36$ mm	– 4.11	– 2.16	– 1.37	– 1.28	– 1.72
$\hat{\rho}=0.71$	0.70	0.76	0.81	0.84	0.86

Table 6: Forecasted intervals and the sum of their absolute differences* d *between real measurements and centres of forecasting intervals produced by four models from (1)

Modell 1

	K = 30	K = 31	K = 32	K = 33	d
$p = 2$	39.73 ± 18.05	49.76 ± 18.05	61.07 ± 18.05	45.55 ± 18.05	50.48
$p = 3$	35.02 ± 18.35	54.89 ± 18.35	71.79 ± 18.35	49.24 ± 18.35	55.42
$p = 4$	33.35 ± 19.67	56.23 ± 19.67	70.61 ± 19.67	45.32 ± 19.67	61.16
$p = 5$	25.64 ± 20.21	61.08 ± 20.21	69.62 ± 20.21	45.13 ± 20.21	67.04
$p = 6$	31.52 ± 19.32	64.34 ± 19.32	52.07 ± 19.32	52.55 ± 19.32	51.21
$p = 7$	33.45 ± 29.1	65.95 ± 29.1	50.28 ± 29.1	54.82 ± 29.1	40.97

Modell 2

	K = 30	K = 31	K = 32	K = 33	d
$p = 2$	48.75 ± 19.04	60.13 ± 19.04	65.87 ± 19.04	45.40 ± 19.04	66.35
$p = 3$	39.29 ± 18.48	62.12 ± 18.48	79.33 ± 18.48	53.39 ± 18.48	64.35
$p = 4$	39.32 ± 20.02	61.55 ± 20.02	78.95 ± 20.02	53.08 ± 20.02	63.74
$p = 5$	35.17 ± 20.13	68.48 ± 20.13	75.39 ± 20.13	52.95 ± 20.13	68.76
$p = 6$	27.91 ± 20.1	58.57 ± 20.1	54.05 ± 20.1	45.25 ± 20.1	52.45
$p = 7$	37.67 ± 21.56	71.96 ± 21.56	52.95 ± 21.56	60.04 ± 21.56	40.20

Modell 3

	K = 30	K = 31	K = 32	K = 33	d
$p = 1$	35.47 ± 15.89	31.67 ± 15.89	60.87 ± 15.89	62.02 ± 15.89	52.71
$p = 2$	36.61 ± 16.67	31.49 ± 16.67	59.75 ± 16.67	56.81 ± 16.67	55.83
$p = 3$	39.64 ± 17.54	32.73 ± 17.54	58.71 ± 17.54	60.73 ± 17.54	49.89
$p = 4$	39.32 ± 18.24	34.27 ± 18.24	55.69 ± 18.24	61.45 ± 18.24	44.29
$p = 5$	47.14 ± 18.84	33.25 ± 18.84	59.94 ± 18.84	67.18 ± 18.84	51.66
$p = 6$	47.41 ± 22.05	35.12 ± 22.05	57.29 ± 22.05	54.78 ± 22.05	59.80

Modell 4

	K = 30	K = 31	K = 32	K = 33	d
$p = 1$	28.89 ± 17.63	31.19 ± 17.63	61.90 ± 17.63	54.58 ± 17.63	68.24
$p = 2$	28.86 ± 18.83	29.95 ± 18.83	63.11 ± 18.83	54.73 ± 18.83	70.57
$p = 3$	36.28 ± 18.59	29.64 ± 18.59	59.79 ± 18.59	58.52 ± 18.59	56.35
$p = 4$	30.44 ± 18.6	27.38 ± 18.6	50.35 ± 18.6	51.86 ± 18.6	61.67
$p = 5$	38.51 ± 19.19	25.88 ± 19.19	54.90 ± 19.19	57.13 ± 19.19	55.40
$p = 6$	37.68 ± 20.9	27.20 ± 20.9	55.13 ± 20.9	49.34 ± 20.9	61.90

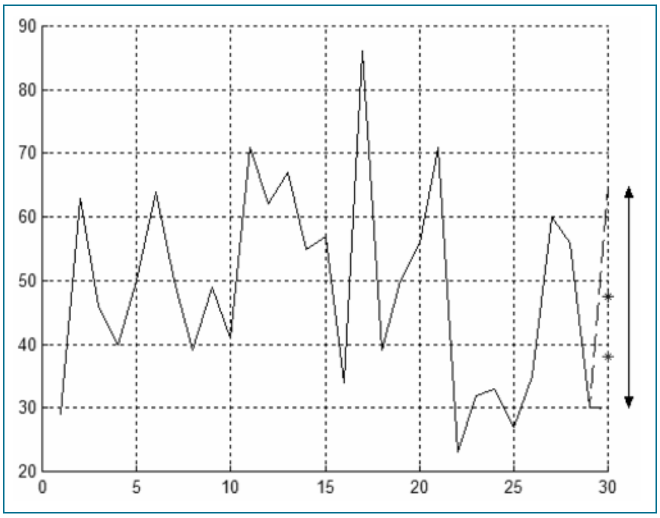

Fig. 3: Graphical presentation of the interval-related, short-time forecasting (dotted lines, arrows) by model 1 from (1) with $p = 7$ for the dam deformation in the month 30 based on measurements from the months 1–29. In this case $\hat{\rho} = 0.88$, $\hat{\varepsilon} = -0.32 \times 10^{-9}$ and $\hat{\sigma} = 13.94$ are obtained. The real deformation in the month 30 is drawn as the under star. The centre of its forecasting interval is drawn as the upper star

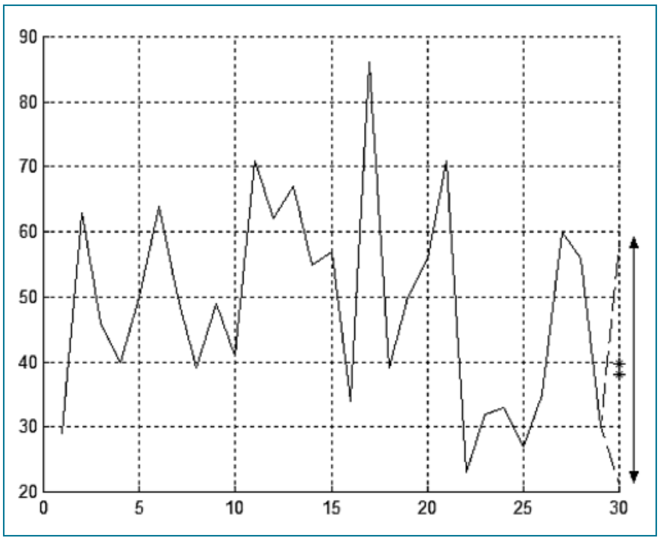

Fig. 5: Graphical presentation of the interval-related, short-time forecasting (dotted lines, arrows) by the model given in (6). The real deformation in the month 30 is drawn as the under star. The centre of its forecasting interval is drawn as the upper star

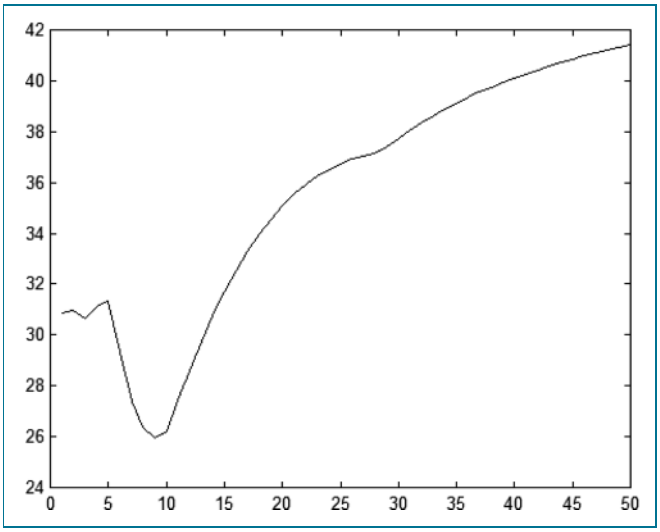

Fig. 4: Dependence of values d *on the choice of* eps *for the multiple regression model in (6)*

Abstract

The movement and deformation of surface can be considered as a random and dynamic time-varying process. Methods of multiple regression analysis can help to recognize the linear structure of such processes and to make the forecast. Qualified monitoring and forecasting the surface movement and deformation can sometimes prevent natural disasters.

Related to this problem, a two-step-modeling is proposed in the paper Yuanzhong and Litao (2005). Here, we discuss some alternative models, which can be applied as well to the trend analysis of the dam deformation as to its short-time forecasting. These models are mainly based on methods of the adjusment theory, see for example Wolf (1979). Some ideas come from applied regression analysis, see Draper and Smith (1998).

This paper presents a completion to the approach proposed in Yuanzhong and Litao (2005). For this reason, a case study of deformation analysis and short-time forecasting is also related to a gangue dam (a gold mine in Shandong Province, Chine).

References

[1] Draper, N.R.; Smith, H. (1998): Applied Regression Analysis, Willey-Interscience

[2] Yuanzhong, L.; Litao, H. (2005). Multiple Recursion Forecast of Surface Movement and Deformation. AVN (Allgemeine Vermessungs-Nachrichten), 3/2005, pp. 88-92

[3] Wolf, H. (1979): Ausgleichungsrechnung. Ferd. Dümmlers Verlag, Bonn

Anschrift der Verfasserin:
Dr. rer. nat. Olga Waelder,
Senior Scientist, Institute for Cartography,
Technical University of Dresden, Mommsenstrasse 13,
D-01062 Dresden,
E-Mail: olga.waelder@tu-dresden.de