Akademie věd České republiky
Ústav teorie informace a automatizace, v.v.i.

Academy of Sciences of the Czech Republic
Institute of Information Theory and Automation, v.v.i.

RESEARCH REPORT

Petr Nedoma and Petr Pecha

Predictive Monitoring of Radiation Situation

No. 2173 June 2007

Project GAČR No. 102/07/1596

ÚTIA AV ČR, v.v.i., P.O.Box 18, 182 08 Prague, Czech Republic
Fax: (+420)286890378, http://www.utia.cz, E-mail: utia@utia.cz

Contents

1 Introduction

This report describes *predictive monitoring* of radiological situation on measuring units of the Early Warning Network (EWN) of the Czech Republic.

The monitoring is based on predictive capabilities of simple models estimated in Bayesian way. The report exploits ready models and algorithms and focuses on their achievable performance. The underlying theory can be found in many sources, e.g. [1, 2]. The report itself is written for non-mathematicians and the performance is inspected using extensive simulation studies. Typical results are summarized here.

The algorithms essentially check natural background dose-rate measurements with the aim to recognize a non-natural increase in the measurements. The recognition is to be made as quickly as possible – this information could be vital for early warning as well as in early stages of a nuclear release.

Autoregressive models and mixture models are used. Estimates of their parameters are recursively updated according to the adopted Bayesian methodology. These models suit well to the considered technical problem. With them, the values to be measured can be predicted and the following key question answered:
"What is the probability that the measurement in a measuring station of the EWN will reach a threshold value in couple of measuring instants?". In this way, a nuclear release can be detected safely before actually reaching a dose-rate threshold value. Consequently, the corresponding preventive actions can be speeded up while the level of false alarms is kept low.

In this study, non-natural trends in measurements are simulated as no real data reflecting releases are available from the EWN. The tests are as extensive as possible and processing algorithms are summarized in appendix to ensure reproducibility of algorithms. MATLAB (MathWorks) [3] has been selected as the experimental environment. Resulting software modules for predictive monitoring are available in the form of C-coded modules.

2 Operations of EWN

The EWN of the Czech Republic has been established after 1990 to provide overview information about radiation situation across the territory and to raise an alarm in case of unreported accident outside the ČR.

Each EWN measuring site provides values of integrals of dose-equivalent over 5 minutes. Consequently, measurement noise is suppressed substantially. The EWN operates in one of three modes – standard mode, alert mode and emergency mode. The modes of operation differ in frequency of data transmission to Center of Radiation Monitoring Network. The modes are started when the dose rate measurement reaches a threshold value.

This report is dedicated to the problem of prediction that the measurements will exceed an "emergency level" that is set uniformly to the value of $\mu Sv/hour$. The level cannot be reached by the natural background radiation.

3 Measurements available

The dose-rate measurements analyzed have been supplied by National Radiation Protection Institute in the form of 5-minutes measurements of the Czech EWN, sites Churáňov, Temelín and Dukovany.

The Churáňov is a mountain measuring site with relatively higher natural background, the Temelín and Dukovany are places of the nuclear power stations. One hour averages are used for the purpose of the predictive monitoring. Measurements from 1. 1. 2006 to 24. 4. 2007 are available. The original data and the one hours averages are described in Fig. 3.

The data characteristics are:

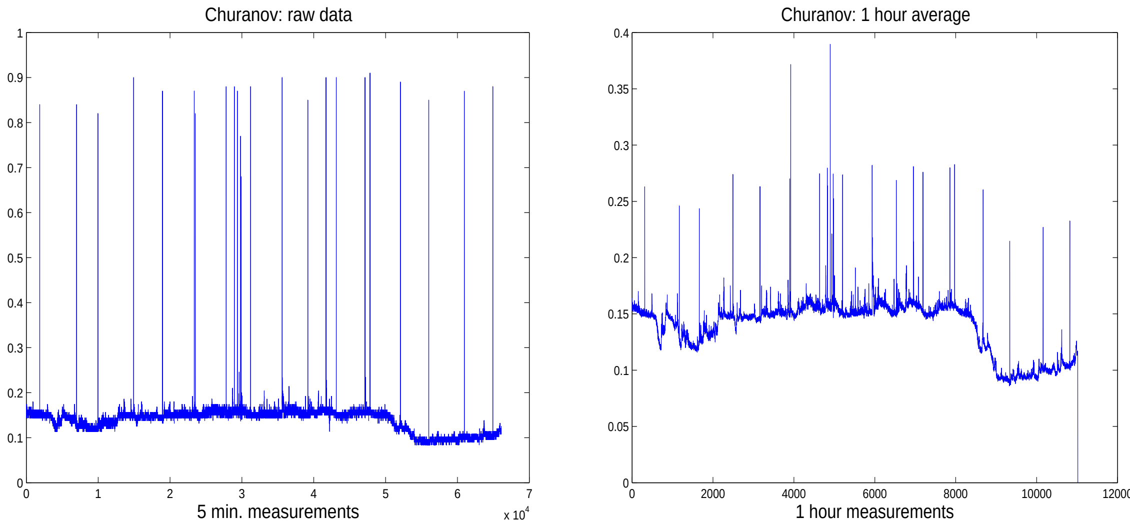

Figure 1: Raw data and 1 hours averages, site Churáň

site	mean	median	minimum	maximum	st. deviation
Churáňov	0.139	0.149	0.086	0.390	0.023
Temelín	0.130	0.134	0.093	0.271	0.015
Dukovany	0.09	0.093	0.058	0.371	0.008

The data histograms are displayed in Fig. 3, Fig. 3 and Fig. 3, respectively.

The peaks in measurements are produced by radon daughter products in aerosol form washed out by rain, the slow trends in measurements are caused by year seasons and the "stationary fluctuations" reflect changes of terrestrial and cosmic radiation.

4 Mixture model

A sequence $d(\mathring{t}) \equiv (d_1, \ldots, d_{\mathring{t}})$ of data records d_t, labelled by the discrete time $t \in t^* \equiv \{1, \ldots, \mathring{t}\}$ is observed and mutual relationships are searched for. They are modelled by the joint probability density function (pdf)

$$f(d(\mathring{t})|\Theta) = \prod_{t \in t^*} f(d_t|d(t-1), \Theta).$$

Both the joint pdf and respective factors are conditioned on unknown parameters Θ. The considered *parametric mixture* model has the form

$$f(d_t|d(t-1), \Theta) = \sum_{c \in c^*} \alpha_c f(d_t|d(t-1), \Theta_c, c). \tag{1}$$

The individual pdfs $f(d_t|d(t-1), \Theta_c, c)$ are called parametric components. The unknown parameter Θ consists of probabilistic weights of components $\alpha \equiv (\alpha_1, \ldots, \alpha_{\mathring{c}}) \in \alpha^* \equiv \left\{\alpha_c \geq 0, \sum_{c \in c^*} \alpha_c = 1\right\}$ and by individual parameters $\Theta_c,\ c \in c^*$, of components. The components are decomposed by the chain rule

$$f(d_t|d(t-1), \Theta_c, c) = \prod_{i \in i^*} f(d_{i;t}|\psi_{ic;t}, \Theta_{ic}, i, c), \tag{2}$$

where $f(d_{i;t}|\psi_{ic;t}, \Theta_{ic}, i, c)$ are called parametric factors. They predict scalar entries $d_{i;t}$ of d_t called factor outputs. They are assumed to depend on regression vectors $\psi_{ic;t}$ that consist of current values of other record entries $d_{j;t},\ j > i$ and several delayed record values $d_{t-k},\ k \geq 1$. The factorization (2) allows us to combine entries of logical and continuous nature, to consider factors of different types.

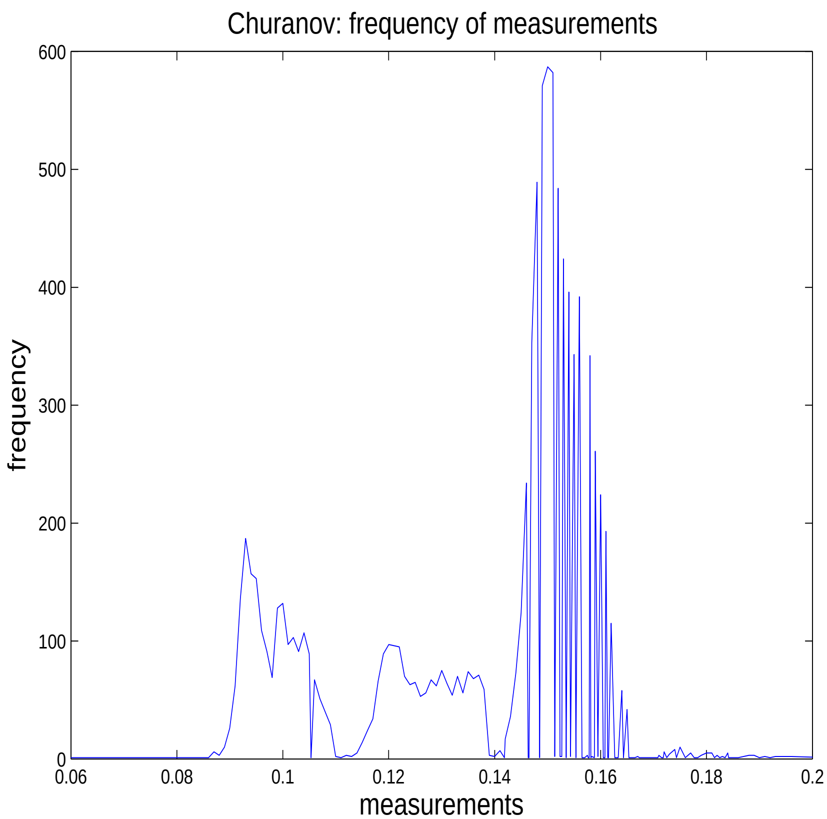

Figure 2: Histogram of data, site Churáňov

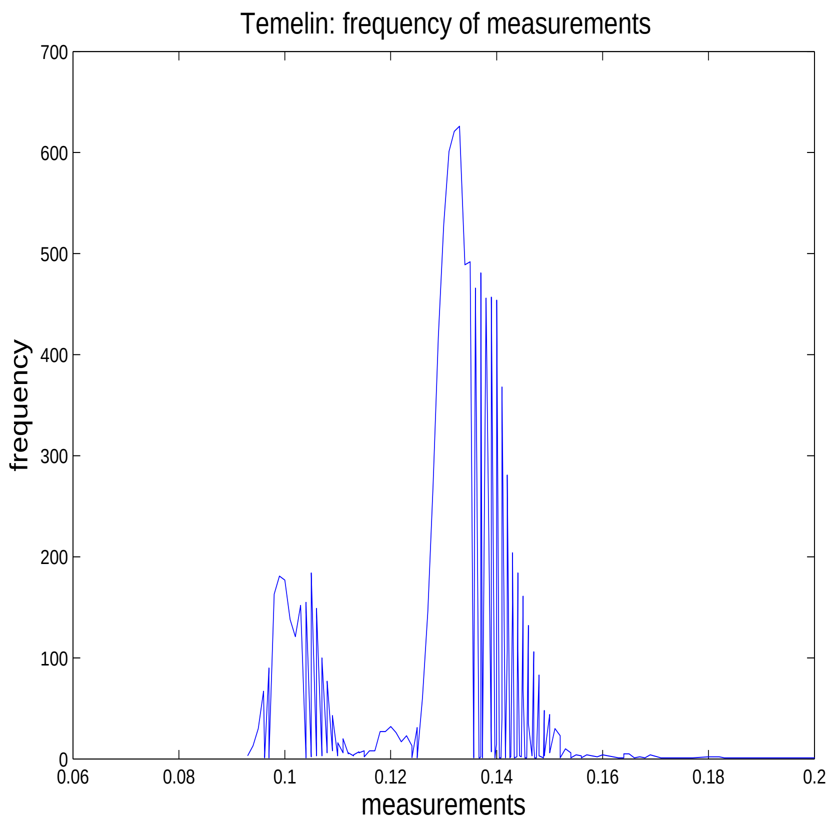

Figure 3: Histogram of data, site Temelín

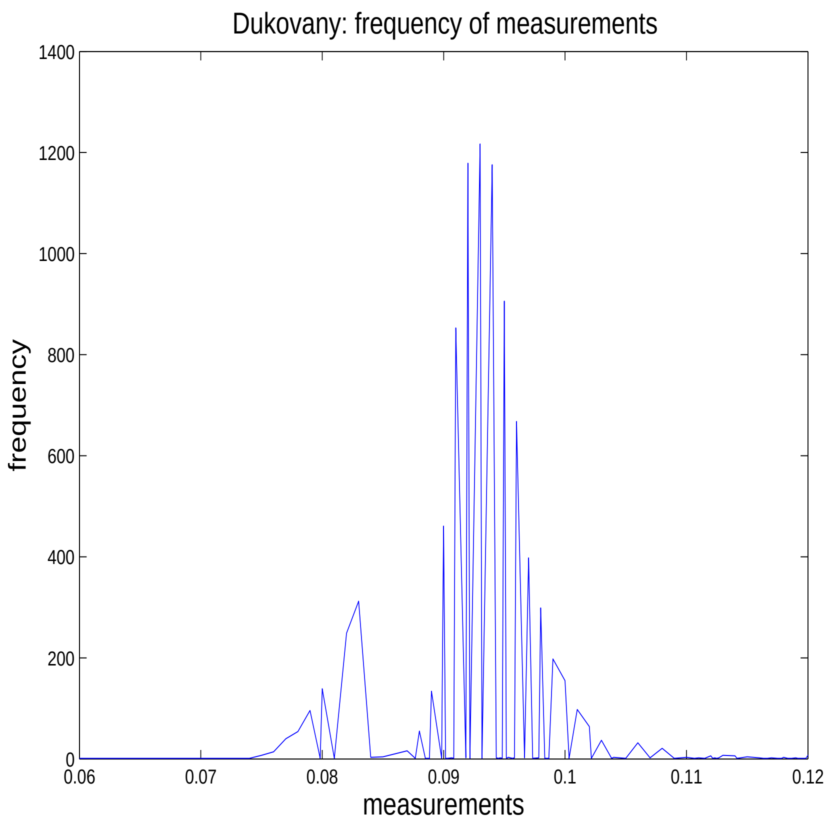

Figure 4: Histogram of data, site Dukovany

In this study, predictions of scalar measurements originating from a single site are treated. Consequently, parametric components coincide with parametric factors. They factors are normal autoregressive (AR) models. Their conditional expectations are assumed to be a linear combination of an unknown number of delayed measurements. The auto-regression coefficients and variances are assumed to be constant but factor specific.

The theory connected to mixtures and estimation of their structure and parameters can be found in [4]. Relevant software and case studies can be found in supplement of this book. Within this report, the results of the following particular subtasks are presented:

- *Mixture initialization* It finds the mixture structure (number of components $\mathring{c}$, structure of respective factors forming the component, construction of statistics describing initial estimates of parameters)
- *Prediction with mixture* It predict the modelled data up to using the initialized and estimated mixture.

5 Results of mixture initialization

The initialization starts with an initial mixture of one component of the order 6. The resulting mixtures are displayed in the Fig. 5. The pdf of resulting components is denoted by dash-dotted lines. The full line corresponds with the marginal pdf.

The following table describes structures of mixtures for respective measurement sites. The column "dfcs" contains an effective number of data described by the corresponding component. Thus, two (three) numbers in this column mean that the mixture has two (three) components. The last column, labelled delays, lists delayed data that were found significant for predicting of the current value of the modelled signals. The numbers 1 4 6 mean that d_t is the best predicted using linear combination of d_{t-1}, d_{t-4} and d_{t-6}. Other possibilities should be read similarly.

site	dfcs	delays
Churáňov	4278	2 3 4 5 6
	295	1
Temelín	153	1 4 6
	6488	1 2 4
	4352	3 4 5 6
Dukovany	164	1 2 3 6
	5708	1
	5121	2 3 4 5

The components found in the mixtures correspond to individual components of the background radiation The peaks in measurements are produced by radon daughter products in aerosol form washed out by rain, the other components are caused by year seasons and whether and/or reflect changes of terrestrial and cosmic radiation.

6 Results of prediction with mixtures

Prediction means estimation of future values of the measurements from their history. Its quality is usually judge according to the prediction error, i.e., the difference between the predicted and the real value of the measurement. Root-mean square error normalized by standard deviation of data is used here as a numerical indicator of prediction quality. This *prediction error norm* "epn" shows what part of variance of measurements can be "explained" by the model.

Typical results of recursive prediction of Churáňov measurements by the initialized mixture are in the Fig. 6. The measurements are displayed by dotted line (lower curve) and the prediction by full line.

The prediction error norms for all three sites are in the following table

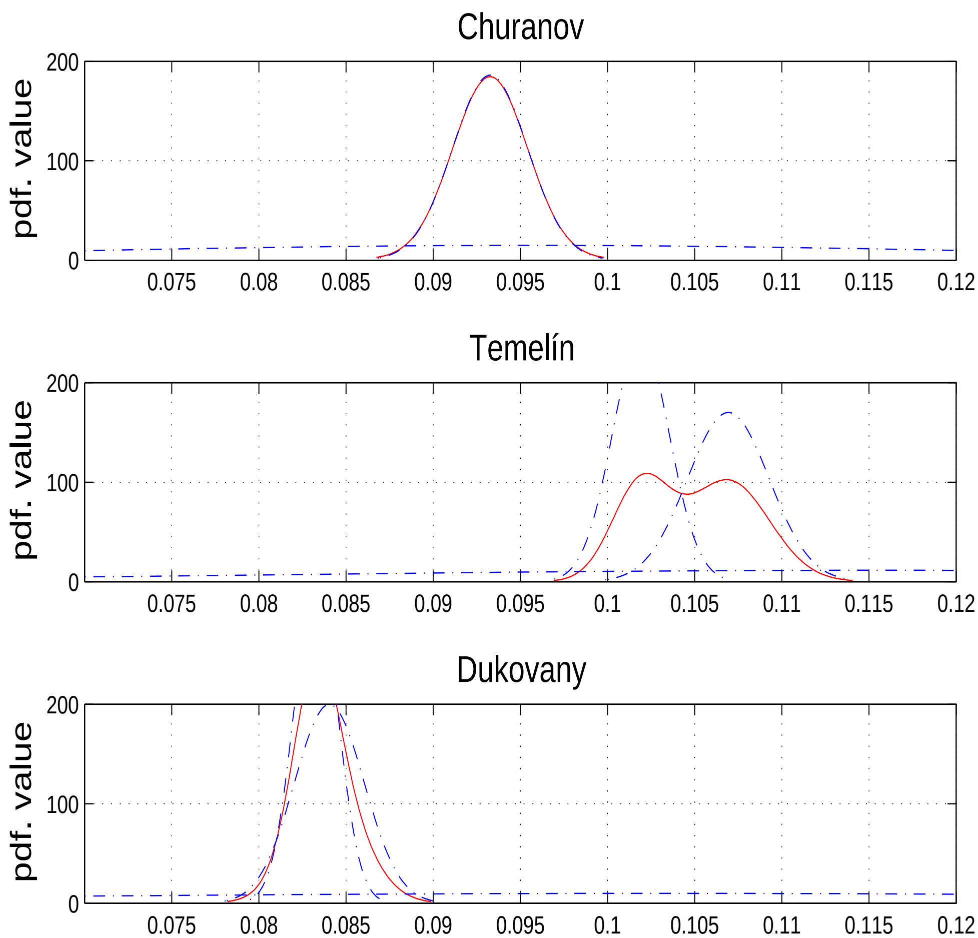

Figure 5: Initialized mixtures

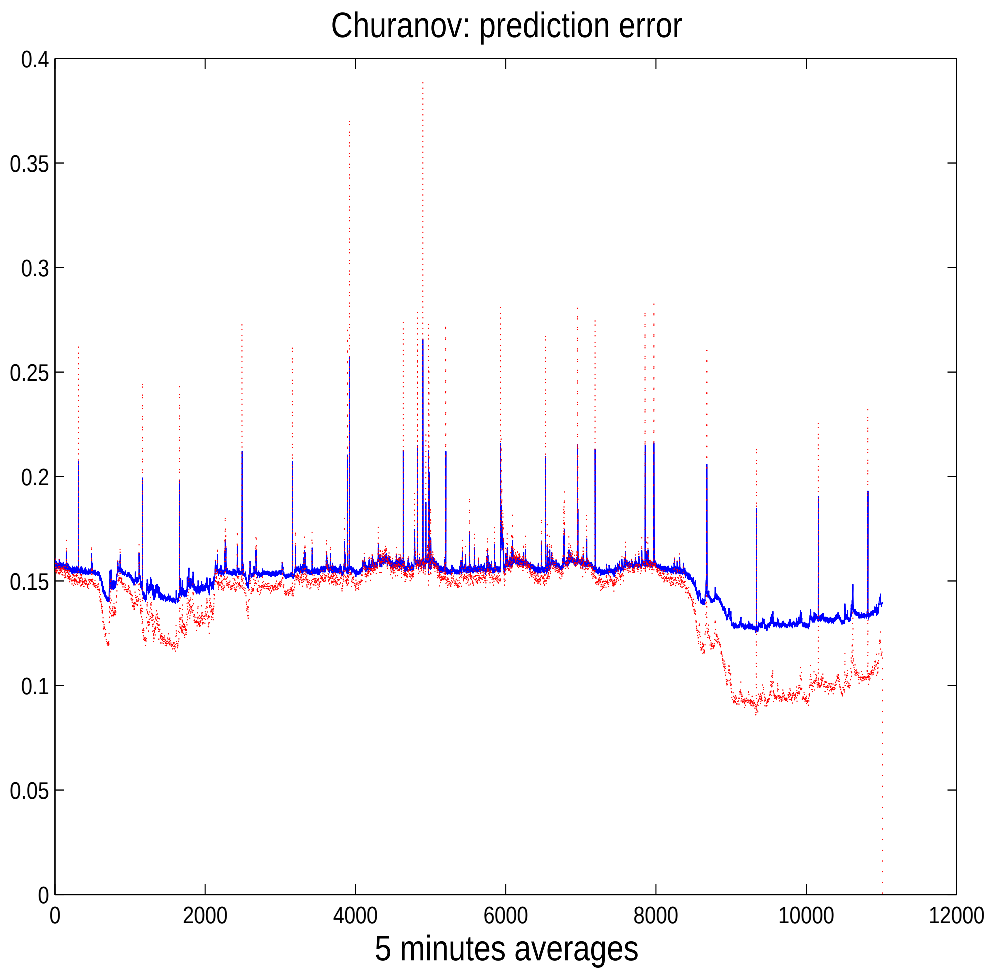

Figure 6: Predictions

site	epn
Churáňov	0.34
Temelín	0.39
Dukovany	0.78

To get a feeling how much we gained by using complex mixture model, we have performed another case study with auto-regressive model only with estimated offset (marked by 0 in column of delays). The following table shows the results for Churáňov site.

delays	epn
1	0.39
1 0	0.38
1 1 0	0.60

7 Conclusions

The presented results indicate that mixtures of auto-regressive models applied to individual sites improve predictions comparing to plain auto-regression. Their individual components can be well interpreted as respective constituents of natural radiation. The results promise that this interpretation and application to multivariate date will provide a very reliable warning on non-natural increase of background radiation.

References

[1] V. Peterka, "Bayesian system identification", in *Trends and Progress in System Identification*, P. Eykhoff, Ed., pp. 239–304. Pergamon Press, Oxford, 1981.

[2] J.O. Berger, *Statistical Decision Theory and Bayesian Analysis*, Springer-Verlag, New York, 1985.

[3] The Matlab Inc., *Matlab documentation*, vol. MA 01760-2098, Natick, 2000.

[4] M. Kárný, J. Böhm, T. V. Guy, L. Jirsa, I. Nagy, P. Nedoma, and L. Tesař, *Optimized Bayesian Dynamic Advising: Theory and Algorithms*, Springer, London, 2005.