Geophysical Research Abstracts, Vol. 9, 01084, 2007
SRef-ID: 1607-7962/gra/EGU2007-A-01084

K-Nearest-Neighbour classifiers predict seismic precursors by hydrogeochemical data

L. Castellana (1), T. Maggipinto (2), P.F. Biagi (2,3)

(1) Department of Biomedical Sciences, University of Foggia, Italy, (2) Department of Physics, University of Bari, Italy, (3) Inter-Department Centre for the Evaluation and Mitigation of the Volcanic and Seismic Risk, University of Bari, Italy, (castellana@fisica.uniba.it / Fax: +390805442434 / Phone: +390805443245)

In this study we address the problem of detecting seismic precursors by analyzing time series of hydrogeochemical data. Classical approaches are mainly based on a *qualitative* analysis of the time series and do not address the problem of estimating the number of false positives. In particular, they examine the whole time series X=$\{x_i\}_{i=1,..,N}$ to search anomalous signal shapes that can be related to earthquakes. This makes precursor detection still empirical and earthquake forecasting far away a quantitative evaluation. The method we propose analyses short temporal windows, $\mathrm{x} = (x_j^{\eta}, x_j^{\eta-1}, ..., x_j^1)^T \in \Re^{\eta}$, of size η and establishes if **x** is a seismic precursor on the basis of observations of seismic precursors previously detected and characterized by a human expert. This base of knowledge takes the form of a sample $S = \{(\mathrm{x}_1, y_1), (\mathrm{x}_2, y_2), ..., (\mathrm{x}_\ell, y_\ell)\}$ composed of ℓexamples ($\mathbf{x}_i$, y_i) where $\mathrm{x}_i = (x_i^{\eta}, x_i^{\eta-1}, ..., x_i^1)^T \in \Re^{\eta}$, and y_i =1 if $\mathbf{x}_i$ is a seismic precursor event, y_i=0 otherwise. In this setting, the problem of seismic precursor detection or classification can be seen as a supervised learning problem, or a *learning from examples* problem in which the goal is to determine a separating surface which is able to discriminate seismic precursors from no seismic ones, or to distinguish among different types of events. Before introducing the main aspects of our work, it is worth to point out that the ultimate goal of any classifier, and in general of any predictor, is *to generalize*, that is to predict the correct output y relative to never seen before input patterns **x**, by using a sample S, composed of a *finite* number of data. Thus the central problem is not classifying the training data in S. The crucial problem is to design classifiers having low error rate on new data.

In this paper we use K Nearest Neighbour (K-NN) classifiers for discriminating "no-seismic signal" (y_i=0), "precursor-signal" (y_i=1) and "co-post seismic signal" (y_i=2) in time series, relative to 13 different hydrogeochemical parameters collected in water samples from a natural spring in Kamchachta (Russia) peninsula in the last 27 years, daily sampled. In particular we analyze Na^+, Cl^-, Ca^{++}, HCO_3 and H_3BO_3 ions; pH, Q and T parameters; N_2, CO_2, CH_4, O_2 and Ar gases. Experiments have been carried out by varying the number ℓ of training examples and the order η of the model. The prediction error was measured by Leave-K-Out-Cross-Validation (LKOCV) procedure, a statistically well founded method for estimating the accuracy of predictors by using a finite number of observations. We measured a prediction accuracy of 85% by using Na^+ time series data and a temporal window of size η=100. This shows that information collected some months before the event under analysis are necessary to improve the classification accuracy. Moreover, the results show that ions are more effective than parameters and gases to discover hydrogeochemical seismic precursors. A complete analysis and discussion of the experimental results obtained with the other data will be provided in the final version of the paper.