Geophysical Research Abstracts, Vol. 8, 02525, 2006
SRef-ID: 1607-7962/gra/EGU06-A-02525

Automatic detection of earthquake precursors in groundwater Chlorine content

L. Castellana, T. MAggipinto, P.F. Biagi
Physics Department, University of Bari, Italy (castellana@fisica.uniba.it)

Earthquake prediction is strictly related to the study of empirical precursors. However, the use of precursors in earthquake prediction is a large extent still empirical since the relationship linking earthquakes and premonitory anomalies is very complex and might vary among different seismogenetic zones. Rather than simply to observe, by visual inspection, the concomitantly occurring of changes in a groundwater ion content and earthquakes, we face the problem of the automatic detection of hydrogeochemical earthquake precursors. The problem is to assign a label y - "normal signal", "precursor event", "co-post seismic event" - to a new signal $\mathrm{x} \in \Re^m$, starting from the knowledge of ℓ *examples* $S = \left\{ \left(\mathrm{x}_{1,y_1} \right), \left(\mathrm{x}_{2,y_2} \right), ..., (\mathrm{x}_\ell, y_\ell) \right\}$ whose association between the signal pattern $\mathrm{x}_i \in \Re^m$ and its relative class label y_i is known in advance. In this study, the signal x is composed by the groundwater chlorine content recorded from a natural spring in the Petropavlovsk area, Kamchatka. Under this perspective, the problem of seismic event 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, optimal under certain conditions, which is able to discriminate normal from seismic events, or to distinguish among different types of seismic events. It is worth to point out that the ultimate goal of any classifier, and in general of any learning machine, is *to generalize*, that is to predict the correct output y relative to never seen before input pattern x, by using a training set S composed of a *finite* number of examples. Thus the central problem is not classifying the training data in S, because any sufficiently complex learning machine could separate S without errors. The crucial problem is to design classifiers having low error rate on new data. For measuring the generalization abilities of the learning machines trained by using a finite amount of data we use the Leave-K-Out-Cross-Validation (LKOCV) procedure. In this general framework, experiments have been carried out for discriminating precursor events in chlorine measurements. To this

end, we have used a dataset composed by 10224 groundwater Chlorine ion measurements. We have analysed the behaviour of the generalization error measured by using the LKOCV error of K Nearest Neighbour (K-NN) classifiers by varying the number ℓ of training examples and the order m of the model. We find that, for a fixed order m of the model, the generalization error decreases increasing the number ℓ of training data, reaching a plateau of 20% with $\ell = 210$. This shows that the size of our data set is more than enough for training accurate classifiers of seismic events. Concerning the order m of the model, we find that the generalization error decreases by varying m in the range [10, 120]. This show that information collected some months before the event under analysis are necessary to improve the classification accuracy. The assessment of the methodology aims to use other components of the signal like the spectral content, the derivative information etc.