An Application of Evolutionary Algorithms to Study the Extent of SLHF Anomaly Associated with Coastal Earthquakes

Guido Cervone, Liviu Panait, Ramesh Singh, and Sean Luke

George Mason University, Fairfax, VA 22030
gcervone@gmu.edu, lpanait@cs.gmu.edu, rsingh3@gmu.edu, and sean@cs.gmu.edu

Abstract. Multi sensor remote sensing provides real time high resolution data that can be used to study anomalous changes on land, in the ocean, and in the atmosphere associated with an impending earthquake. Anomalous behaviour in Surface Latent Heat Flux (SLHF) prior to large coastal earthquakes has been recently found. However, an SLHF time series usually contains several sharp peaks that may be associated either with earthquakes or with atmospheric perturbations. In this paper we have used evolutionary algorithms to perform a search in a large space bounded by longitude, latitude and time, to distinguish between signals associated with earthquakes and those associated with atmospheric phenomena. The algorithm finds paths which delimit the extent of the detected anomalies by optimizing an objective function that takes into consideration several aspects, such as spatial and time continuity, the magnitude of the anomalies, and the distance to the continental boundary. This search strategy is crucial for the development of a fully automated early warning system for providing information about impending earthquakes in a seismically active coastal region.
Experiments have been performed over a 2000 km^2 area comprising a part of the continental boundary between the African and Eurasian plate, roughly corresponding to Italy and Greece, one of the most seismically active regions. Using a 365-days-long time series, we identified three signals associated with seismic events. Additionally, it was possible to establish that the extent of the signal does not propagate further than 600 km from the epicenter of the earthquake.

1 Introduction

Earthquakes every year cost lives and property. Recently, Surface Latent Heat Flux (SLHF) has been found to be a potential earthquake precursor. Recent analysis of SLHF data has shown that its routine measurements can provide early warning information about an impending coastal earthquake [1]. SLHF is a parameter directly correlated with the amount of evaporation from the epicentral and adjoining regions. The magnitude of the SLHF peaks are found to be variable and to depend on location, the distance from the ocean, and the ocean depth. In general, SLHF is higher over the ocean and lower over the land.

We have recently introduced a new data mining methodology, which relies on SLHF data, wavelet maxima curves, and statistical analysis to detect precursory signals associated with earthquakes [2]. We validated this methodology on two large earthquakes in Greece using SLHF remote sensing data from the NCAR/NCEP reanalysis project. Our research identified prominent anomalies about two weeks prior to both events, suggesting the possibility of their use in an early warning system for impending earthquakes.

We have also identified the spatial constraint by selecting grids immediately above the continental boundary, because such a region is likely to undergo the highest stress prior to an impending earthquake [2]. However, since SLHF is related to changes in atmospheric and oceanic parameters, the shape of the anomalies may be considerably affected, thus making the selection of grid paths difficult. The problem of space continuity is very complex because it requires finding an optimal path in a three dimensional space, where the first two dimensions are, respectively, latitude and longitude, and the third dimension is time. To solve this problem, we employed an evolutionary algorithm to search the space for paths which show both time and space continuity.

The large size of the search space (the number of possible combinations in absence of constraints is roughly 2^{29200}) makes the problem very challenging, and also renders numerical approaches almost impractical. Additionally, we able to define a measure quality of a grid path (a fitness function). Due to these reasons, we considered evolutionary computation as a first approach to automating the process of discovering such grid paths. In this paper we used EAs over seismically active regions of Italy and Greece. Three signals found by the EA are associated with coastal earthquakes that occurred on 14 August 2003 off the coast of the island of Lefkada, with magnitude 6.7; on 1 March 2004 in the Peloponnese, with magnitude 5.7; and on 17 March 2004 south of the island of Crete; with magnitude 6.1. Using automatically generated paths it is possible to determine the extent of the anomalies, and the regions to be monitored for an impending earthquake.

The results suggest that this method could be used for the development of a fully automated operational system to provide early warning information for large coastal earthquakes. An operational system called CQuake, has been developed to automatically analyze and obtain early information about disastrous earthquakes. This paper discusses the automated selection of grid paths using evolutionary algorithms in CQuake.

2 Methodology

We employed an evolutionary algorithm to search for grid paths. In our approach, an individual is a variable length vector describing a contiguous sequence of adjacent grid cells. A pair of grid cells is considered adjacent if one cell is a horizontal, vertical or diagonal neighbor of the other one (therefore, each grid cell has eight neighbors). We randomly generated initial grid paths of lengths between 3 and 20. The algorithm ran with a fixed population size of 200 individuals, and parents were selected via lexicographic parsimony tournament of size 5 [3]. The best individual in the population automatically survived to the next generation.

The fitness function takes into consideration four different aspects of a path, such as its length, its space and type continuity, the magnitude of the anomalies, and its distance to the continental boundary. Because the evaluation function is made of four distinct criterion, this type of optimization problem is also called multi-objective [4]. There are different techniques to combine together the results of a multi objective evaluation function, the simpler being the use of a parametric function, where the criteria are weighted differently according to their importance. For the research in this paper, the fitness function was defined (according to the results of many empirical experiments) as follows:

$$F = .4 * C1 + .2 * C2 + .2 * C3 + .2 * C4 \tag{1}$$

where

- *C1: Length of the path* This first criterion is probably the most important. Given a sequence of grids, the system computes in how many of them anomalies where detected. An additional constraint stipulates that anomalies in two consecutive grids must occur within one or two days of each other. The problem is effectively to discover anomalies which are spaced close to each other, and which occur within a short time frame. There are only few of such anomalies within a year, and this technique can effectively be used to discriminate between anomalies associated with earthquakes and anomalies due to other events. A discontinuity in a single grid is allowed, but heavily penalized. Small discontinuities may occur because of local atmospheric phenomena which cancel the anomaly, or by severe discontinuities underground.

- *C2: Distance of the grids from the continental boundary* This second criterion gives a slight advantage to paths which follow the continental boundary. This was introduced because the anomalies associated with earthquakes are thought to follow the tectonic geometry of the region. However because SLHF is an atmospheric parameter, the problem is not constrained to the continental boundary.
- *C3: Spread in time of the anomalies* This third condition penalized anomalies which spread in time. In condition one, anomalies in two contiguous grids are constrained to occur within one or two days of each other. However, the overall spread in time for all the anomalies could potentially grow very large. Let us assume a grid path of 10 grids, and each grid is spaced within 2 days of each other. Then the overall spread of the grid path is 20 days.
- *C4: Magnitude of the anomalies* This fourth and last criterion takes into consideration the magnitude of the detected anomaly. The magnitude is computed by eliminating the seasonal trend from the time-series. This criterion is important because it estimates the statistical significance of the generated grid path, which is proportional to the magnitude of the anomalies along the grid path.

Another important component of an evolutionary computation system is the breeding process. Our search uses individuals represented as lists of grid cells. For the initial experiments presented in this paper, we employed two types of mutation (*one-point* and *two-point*) and no crossover operators to create children. One-point mutation picks a consecutive sequence of grid cells (at either end of the grid path with equal probability), and replaces them with another random series of grid cells such that the result is a valid grid path. The length of the list of cells selected is smaller than half of the size of the genome. The length of the list to replace it is randomly selected between zero and twice the size of the selected list of cells. Two-point mutation picks two cells inside the individual's genome, and replaces the grid cells between them with another list of cells such that the resulting individual is a valid grid path as defined earlier. Either of the two mutation operators are used with equal probability to create children.

3 Results and Discussion

We downloaded SLHF data from 1 January 1998 to 28 March 2004 for the region bounded by latitudes 33N to 45N and longitudes 14E to 28E from the website of the Scientific Computing Division of the National

Center for Atmospheric Research (NCAR)[1]. Kalnay et al. discussed the validation and detailed description of the reanalysis of NCEP SLHF data [5].

Experiments used a 365-day long time series from 30 March 2003 to 28 March 2004. The generated grid paths are illustrated in Figures 1-3. SLHF anomalies are usually not found over the grids at the border of the map, implying that the extent of the signals is confined to within approximately 600 km of the epicenter, and thus is a promising predictor for impending earthquakes. The resolution of the data is approximately 200 km, therefore it is not possible to get a more accurate estimate of the extent of the signal, but this limitation can be overcome in the future when higher resolution data becomes available.

The grid paths generated do not precisely follow the continental boundary. This could be due to the fact that the signals are heavily perturbed by atmospheric and oceanic disturbances, but also may be an artifact of the algorithm trying to generate "long" paths. Paths which zigzag achieve a higher fitness than straight lines do because they are longer. In all the generated grid paths, the epicentral area is always either included or adjacent to the grid path. This important result is consistent with the hypothesis that the maximum stress, which is responsible for higher anomalies, is located in the proximity of the epicenter.

In all three cases the grid paths include anomalies occurring about two weeks prior to the corresponding earthquake events, further confirming the published results [2], where a single grid path was used to find anomalies associated with both the earthquakes of 14 August 2003 and 1 March 2004. This grid path was generated manually by selecting the grids comprising the continental boundary, and it was much shorter (8 grids), compared to longer paths (20 grids) found by the evolutionary algorithm. We believe that the longer paths are better suited to understand the extent of the SLHF anomalies. We also believe that such longer paths are additionally more unique to each specific earthquake.

Figure 4 shows a generated grid path for a time where no earthquake occurred. The grids are very close to each other, and very far away from the continental boundary, resulting in a low fitness. It is also possible to differentiate this path from those associated with earthquakes because of its geometry. After detailed analysis we have found that grids over the ocean tend to be selected more than grids over the land. This is due to the fact that evaporation on the ocean is much higher than over the land, leading to larger fluctuations of the SLHF time-series. Although the values

[1] http://ingrid.ldeo.columbia.edu/ SOURCES/ NOAA/ NCEP-NCAR/

are normalized, anomalies over the ocean still tend to be higher compared to those over the land, and this is due both to the physical properties and because estimation of SLHF over the ocean is always more accurate compared over the land. Another factor that helps in the selection of grids over the ocean is that the continental boundary lies over the ocean, and therefore those grids are most likely to be chosen.

4 Conclusion

We introduced a method based on evolutionary algorithms to distinguish paths associated with earthquakes from those associated to other phenomena. We then experimented with a year long SLHF time-series, and we found continuous series of significant anomalies about two weeks prior to the three earthquakes (magnitudes 6.7, 5.7, 6.1) occurred in West and Southern Greece. In all cases the extent of the signal is constrained within 600 km of the epicenter of the earthquakes, which justifies a local analysis of the data. The proposed methodology could be used to develop a fully automated system to study and monitor SLHF signals associated with large coastal earthquake events.

References

1. Dey, S., Singh, R.P.: Surface Latent Heat Flux as an earthquake precursor. Natural Hazards and Earth System Science **3** (2003) 749–755
2. Cervone, G., Kafatos, M., Napoletani, D., Singh, R.P.: Wavelet maxima curves associated with two recent Greek earthquakes. Natural Hazards and Earth System Sciences (2004) to appear.
3. Luke, S., Panait, L.A.: Lexicographic parsimony pressure. In: GECCO 2002: Proceedings of the Genetic and Evolutionary Computation Conference, Morgan Kaufmann Publishers (2002) 829–836
4. Coello Coello, C.A., Christiansen, A.D.: Multiobjective optimization of trusses using genetic algorithms. Computers and Structures **75** (2000) 647–660
5. Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K.C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., Joseph, D.: The NCEP/NCAR 40-year reanalysis project. Bulletin American Meteorological Society **77** (1996) 437–471

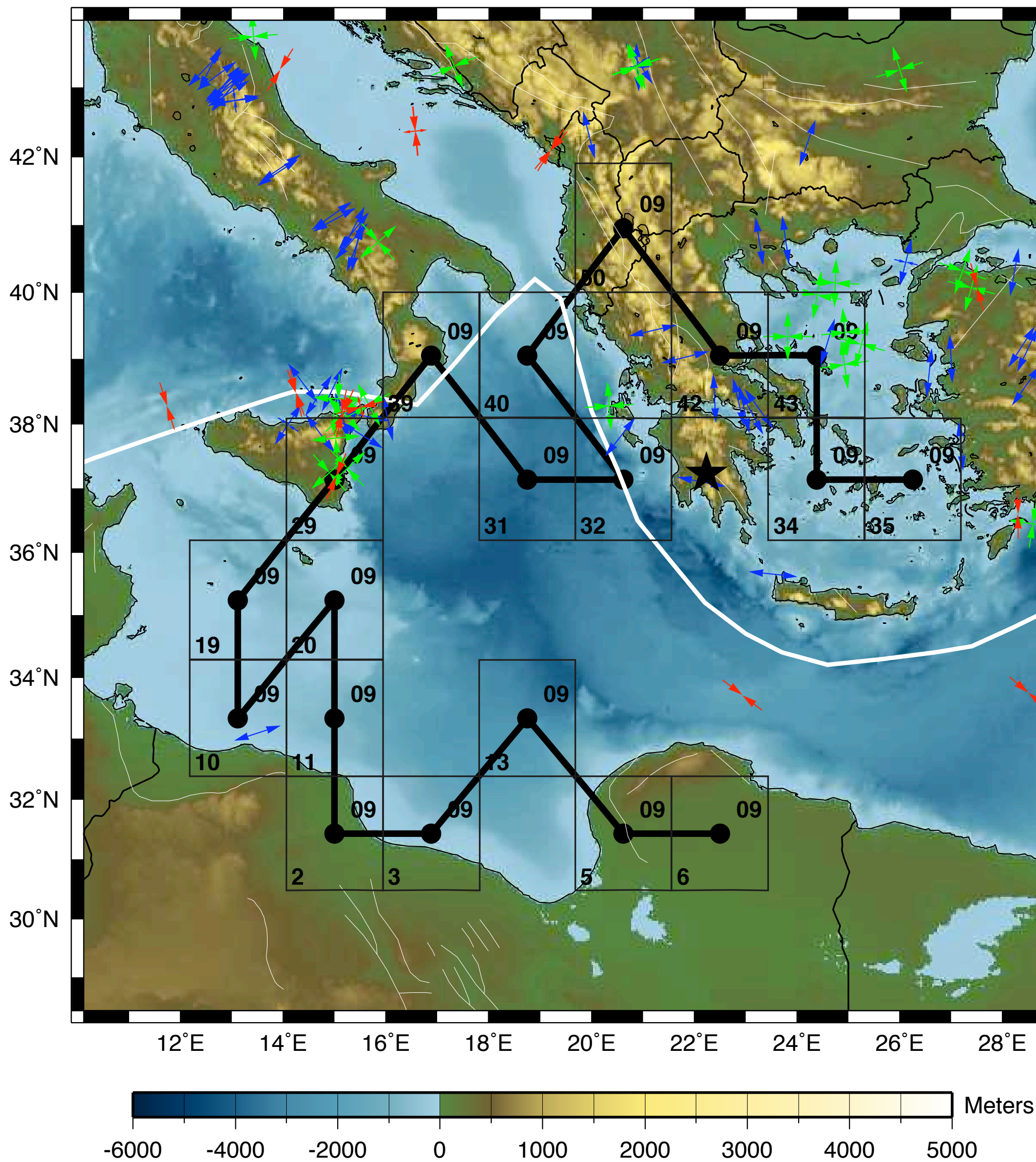

Fig. 1. Grid path associated with the earthquake of 17 March 2004. The epicenter is located at 34.58N and 23.47E

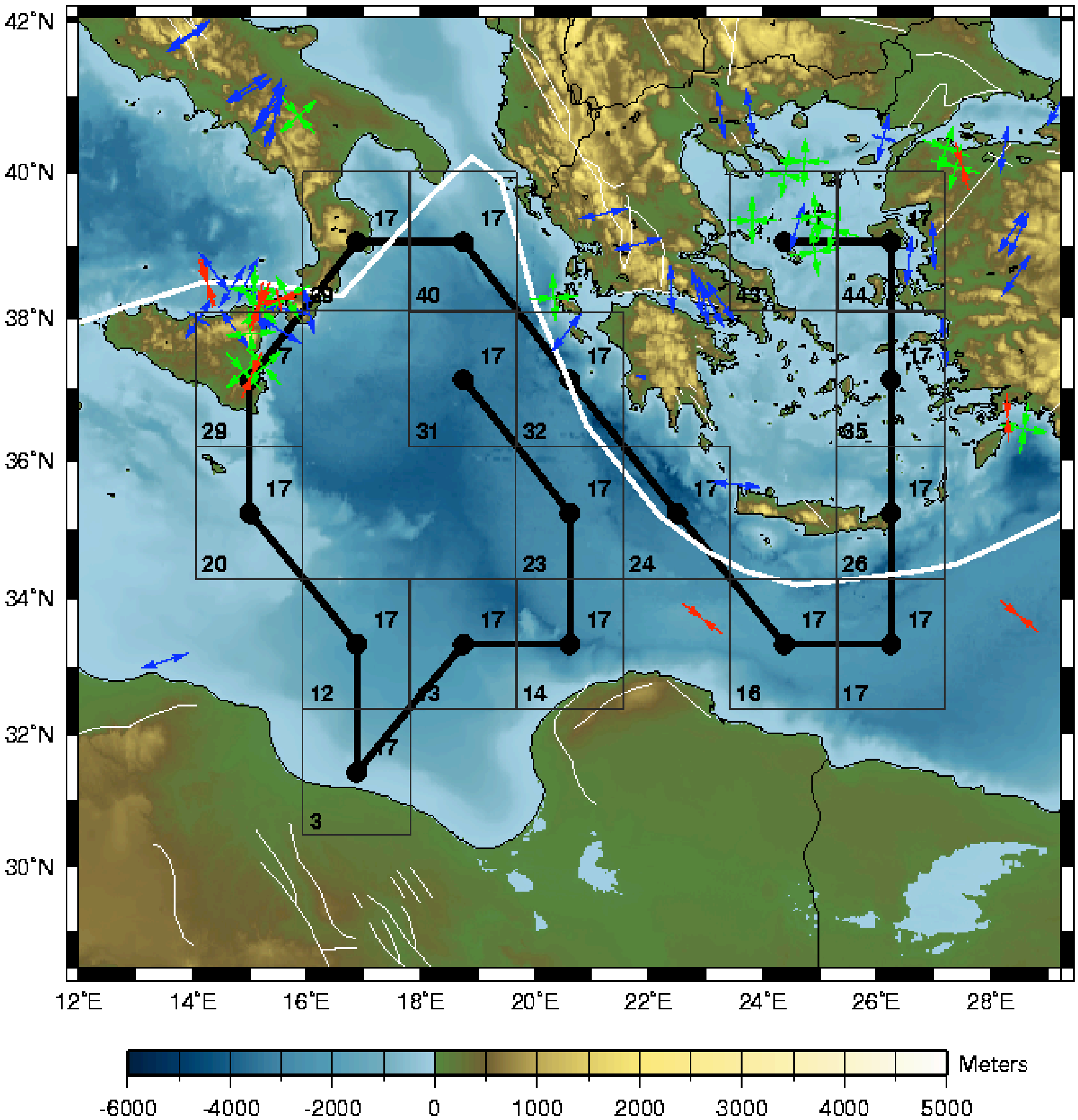

Fig. 2. Grid path associated with the earthquake of 1 March 2004. The epicenter is located at 39.19N and 20.74E

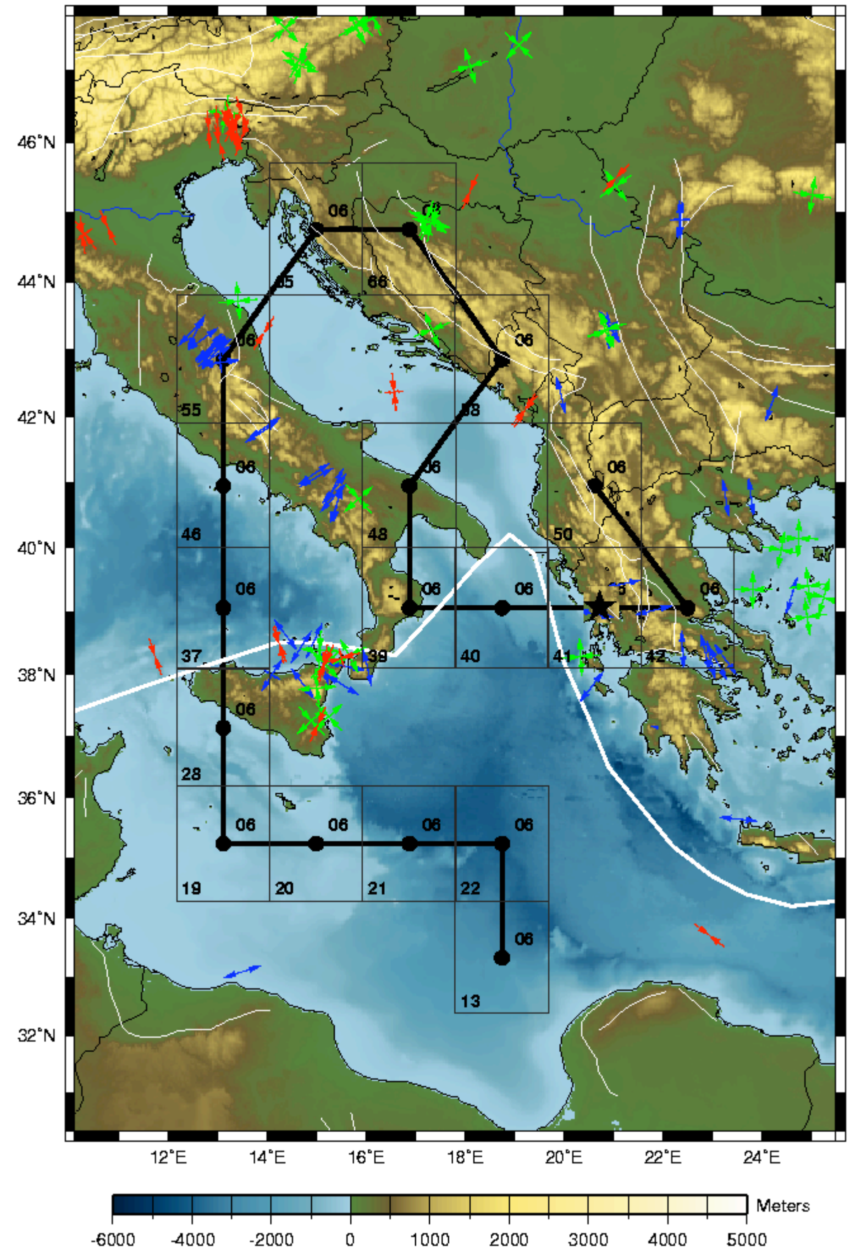

Fig. 3. Grid path associated with the earthquake of 14 August 2003. The epicenter is located at 37.23N and 22.24E

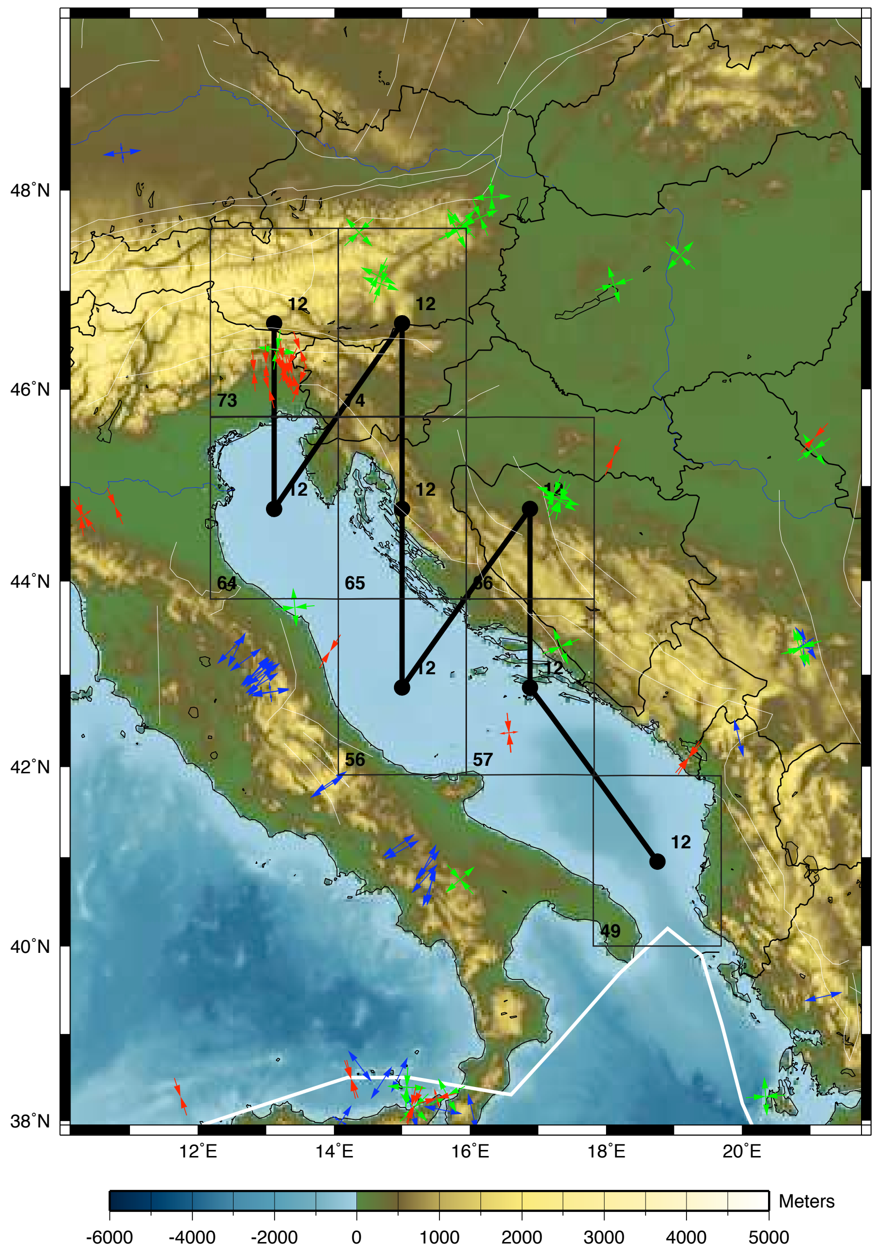

Fig. 4. Grid path not associated with any earthquakes