

# Wavelet maxima curves of surface latent heat flux associated with two recent Greek earthquakes

G. Cervone<sup>1</sup>, M. Kafatos<sup>1</sup>, D. Napolitano<sup>2</sup>, and R. P. Singh<sup>1,3</sup>

<sup>1</sup>Center for Earth Observing and Space Research, School of Computational Sciences, George Mason University, Fairfax, VA 22039, USA

<sup>2</sup>School of Computational Sciences, George Mason University, Fairfax, VA, 22030, USA

<sup>3</sup>Department of Civil Engineering, Indian Institute of Technology, Kanpur 208 016, India

Received: 14 January 2004 – Revised: 5 April 2004 – Accepted: 6 April 2004 – Published: 28 May 2004

**Abstract.** Multi sensor data available through remote sensing satellites provide information about changes in the state of the oceans, land and atmosphere. Recent studies have shown anomalous changes in oceans, land, atmospheric and ionospheric parameters prior to earthquakes events. This paper introduces an innovative data mining technique to identify precursory signals associated with earthquakes. The proposed methodology is a multi strategy approach which employs one dimensional wavelet transformations to identify singularities in the data, and an analysis of the continuity of the wavelet maxima in time and space to identify the singularities associated with earthquakes. The proposed methodology has been employed using Surface Latent Heat Flux (SLHF) data to study the earthquakes which occurred on 14 August 2003 and on 1 March 2004 in Greece. A single prominent SLHF anomaly has been found about two weeks prior to each of the earthquakes.

## 1 Introduction

Multi sensor data available from airborne and spaceborne platforms have become widely used to study the changes of land, ocean, atmospheric and ionospheric parameters, and their relation to various natural hazards. Significant changes prior to earthquake events have been observed in the Surface Latent Heat Flux (SLHF), Sea Surface Temperature (SST), Water Vapor and Chlorophyll Concentrations using multi sensor data (Qiang, 1997; Thanassoulas and Klentos, 2001; Singh et al., 2001a, b, 2004; Dey and Singh, 2003; Dey et al., 2003; Ouzonov and Freund, 2004). These changes suggest the existence of interaction between the lithosphere and the atmosphere, and have opened up new possibilities to the use of satellite-based observations to identify and study earthquakes precursors.

Dey and Singh (2003) and Singh et al. (2004) have recently shown the consistent occurrence of anomalous SLHF peaks a few days prior to the main earthquake event in the case of coastal earthquakes. They have concluded that the routine measurements of SLHF can provide an early warning information about an impending coastal earthquake. The magnitudes of the SLHF peaks are found to be variable, while SLHF tends to be higher over oceans and lower over land. The origin of the anomalous SLHF is likely to be related with the manifestations of surface temperature in the epicentral region which is associated with the build up of stress and movements along the faults. It is believed that the temperature increases prior to an earthquake (Qiang, 1997). The change in temperature for the Gujarat earthquake of 26 January 2001 has been mapped using Infrared (IR) wavelength observations by the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor and is likely due to frictional mechanism along the fault or due to movements of the fluids at depth (Ouzonov and Freund, 2004). Due to the heat conduction, the sea surface temperature increases which in turn is likely responsible to increase the ocean evaporation giving rise anomalous SLHF prior to the main earthquake events (Dey and Singh, 2003).

The yearly time series of the SLHF contains a large number of maxima peaks, several of which are more than 1 or 2 times above the standard deviation. These peaks are due to atmospheric phenomena, earthquakes, or ocean disturbances, and therefore the main challenging task is to identify the SLHF peaks that are precursors of an impending earthquake.

In this paper, a general methodology is presented which employs spatial and temporal analysis of wavelet maxima to identify signals associated with earthquakes. The application of the proposed methodology has been found to identify SLHF anomalies associated with two large recent coastal earthquakes which occurred in Greece on 14 August 2003 and on 1 March 2004, respectively. These two earthquakes were chosen because they provide an excellent case study to validate the methodology, since they are both strong coastal earthquakes occurring in the same region and within a short

span of time. The earthquake of 14 August 2003 occurred off the Ionian island of Lefkada in Western Greece, while the earthquake of 1 March 2004 occurred just outside the southern city of Kalamata.

The proposed methodology uses wavelet transformations as data mining tools by computing the wavelet maxima that propagate from coarser to finer scales. Those maxima are used to identify strong anomalies in the data, however only those anomalies that show continuity in both time and space are assumed as possible precursors for earthquakes. Time continuity means that the detected anomalies occur at the same time or with a short delay of each other, while space continuity means that the detected anomalies are distributed in space according to a precise geometry conforming to the geological settings of the region. The proposed methodology can be applied to different types of spatial and temporal data, and it is not bound to a particular resolution or time sampling.

The results show strong SLHF anomaly about two weeks prior to the earthquake with a precise continuity in both space and time. This methodology has also been validated using several other earthquakes, and in all of the coastal earthquakes, prominent SLHF precursory signals are found to be associated with coastal earthquakes.

## 2 Seismicity of Greece

Greece is located in the most seismically active region of the Mediterranean and of the West Eurasian plate. This region is a part of the collision zone between the Eurasian and the African plates, due to which earthquakes and volcanic eruptions are common in this region. There are two main different tectonic boundaries in the South-West and in the East of Greece, known as the Hellenic Trench and the Hellenic Arc. The Hellenic Trench is the largest area of subduction, in which the denser African plate goes under the less dense Eurasian plate. In this region, the magma rises from under the Earth's crust resulting in a large number of volcanoes. The Hellenic Arc is a transform boundary, in which African and Eurasian plates slide side by side, causing many fault lines which are responsible for the large number of earthquakes, but no volcanoes (Barnes et al., 2003).

Papazachos and Kiratzi (1996) have studied in detail the kinematics of Greece. They have found different types of tectonic regions, namely extensional, compressional and strike-slip motions (Fig. 1). Earthquakes are common in the areas of highest tectonic activities, tend to be of the largest magnitude in the compression and strike-slip zone, and of smaller magnitude in the extension zone.

Several commercial and research models have been developed for an early warning system, mainly using past historical earthquake data which in some case can be traced back as far back as the V century B.C. The model proposed in Thanassoulas and Klentos (2001), (<http://users.otenet.gr/~thandin>), uses past historical data, fracture zones calculated using gravity fields, changes in the electromagnetic field and tidal cycles to determine the occurrence of earthquakes. The

model uses ground based data, but only one ground monitoring station is available, and this precludes the possibility to use the model for real time prediction.

## 3 Earthquakes of 14 August 2003 and of 1 March 2004

On 14 August 2003, an earthquake of magnitude 6.4 occurred about 40 km off the Ionian island of Lefkada in Western Greece. Fifty people were injured, and the region suffered heavy damages amounting to several hundred thousand Euros. The epicenter of this earthquake was at  $39.18^{\circ}$  N– $20.74^{\circ}$  E in the Ionian sea, with focal depth 10 km. The main earthquake was followed by two strong, and a series of smaller aftershocks ([http://neic.usgs.gov/neis/bulletin/neic\\\_xlaf\\\_1.html](http://neic.usgs.gov/neis/bulletin/neic\_xlaf\_1.html)).

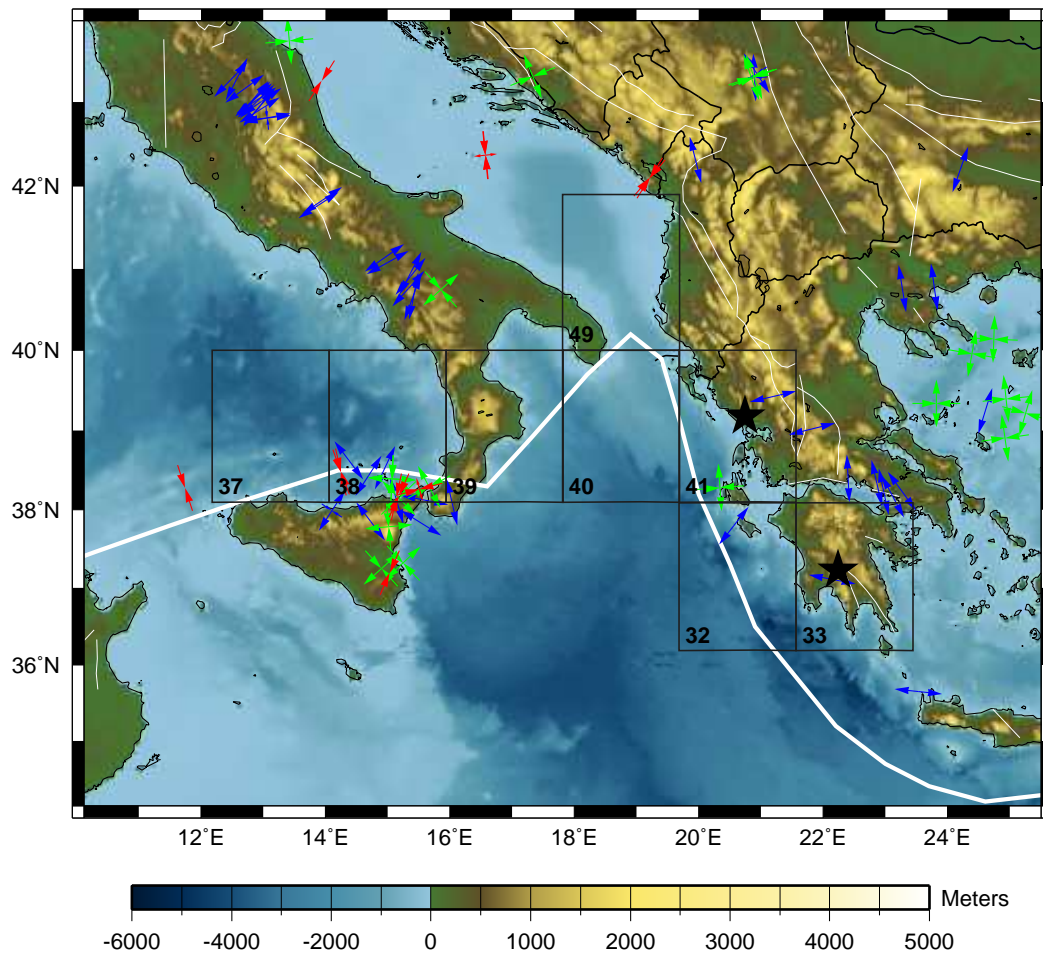
On 1 March 2004, an earthquake of magnitude 5.7 occurred about 7 km NE of the city of Kalamata. The area suffered light to moderate structural damages, and no casualties were directly associated with this seismic event. The epicenter of this earthquake was at  $37.23^{\circ}$  N– $22.24^{\circ}$  E in the SW Peloponnesos, with focal depth of 7 km. The epicenter for this earthquake is located 13 km North of the epicenter of the 6.0 earthquake that struck this region on 13 September 1986.

These two earthquakes occurred in a highly seismic and rapidly deforming region of Greece. Earthquakes of magnitude 5 or larger have been observed to usually occur every two years within 60 km of the epicenters of recent earthquakes. Figure 1 shows a map of northern Greece, the location of the two epicenters (marked with stars), the location of the plate boundary, fault lines and their type (compressional, extensional, transform).

## 4 Data

The data used consist of SLHF data from 1 January 1998 to 28 March 2004 for the region bounded by latitudes  $33^{\circ}$  N to  $45^{\circ}$  N and longitudes  $14^{\circ}$  E to  $28^{\circ}$  E. The SLHF data have been downloaded from the website of the Scientific Computing Division of the National Center for Atmospheric Research (NCAR) <http://ingrid.ldeo.columbia.edu/SOURCES/NOAA/NCEP-NCAR/>.

The data set is represented by a Gaussian grid of 94 lines from equator to pole with a regular  $1.8^{\circ}$  longitudinal spacing and projected into  $2^{\circ}$  latitude by  $2^{\circ}$  longitude in a rectangular grid. The global database of various meteorological parameters is maintained by NCEP. This database is generated taking into consideration the measured values at various worldwide stations and also retrieved from satellite data. The fluxes used in the operational weather forecast models incorporate in-situ observations through the assimilation process. The main drawback of the data source is the frequent change of assimilation methodology and of model resolution, which has been solved by the re-analysis procedure by NCEP incorporating the whole archived data set into single frozen data assimilation system. The validation and detailed description



**Fig. 1.** Map of the region of the earthquake. The epicenters are marked with a star. The grids used in the experiment are marked and labeled. The plate boundary, major fault lines and the type of faulting are also indicated.

of the reanalysis of NCEP SLHF data have been discussed by Kalnay et al. (1996).

The plate boundary data have been downloaded from <http://www-geology.ucdavis.edu/GEL102/demets.html>, and they consist of the best fitting Euler vectors, closure fitting Euler vectors and the global model NUVEL-1 to describe geologically current plate motions between 12 assumed rigid plates. The details of this dataset are described by DeMets et al. (1990).

The major fault lines and their type (compressional, extensional, transform, undetermined) have been downloaded from <ftp.ig.utexas.edu>. This dataset was developed as part of the PLATES project, by the Department of Geophysics at the University of Texas (UTIG). This project is supported by a consortium of oil companies, and it focuses on research into plate tectonic and geologic reconstructions.

## 5 Methodology

The proposed method is a multi-strategy approach which employs a one-dimensional real continuous wavelet transformation to discover singularities in the time series for a par-

ticular grid, and a geometrical analysis of the continuity in time and space of the detected singularities across several grids adjacent to the epicenter chosen according to the tectonics of the region, such as continental boundaries of fault lines, etc. The methodology has been tested using SLHF data, but it is not constrained to a specific type of data. The method has been implemented in a software called CQuake which was written using C, Java, Perl, the statistical package R ([www.r-project.org](http://www.r-project.org)), the mathematical package octave developed at University of Wisconsin ([www.octave.org](http://www.octave.org)), in combination with the WaveLab library developed at Stanford University ([www-stat.stanford.edu/wavelab](http://www-stat.stanford.edu/wavelab)), and the Generic Mapping Tool (GMT) collection of functions developed at University of Hawaii (<http://gmt.soest.hawaii.edu/>) used to plot geographical maps.

### 5.1 Wavelet transformation

For any given real valued function  $\phi$  with zero average  $\int_{-\infty}^{\infty} \phi(t) dt = 0$ , let

$$Wf(u, s) = \int \frac{f(t)}{\sqrt{|s|}} \phi\left(\frac{t-u}{s}\right) dt \quad (1)$$

be the real continuous wavelet transform of a function  $f$ . Since  $\phi$  has zero mean, the previous integral measures the variation of  $f$  in a neighborhood of time (position)  $u$  of size proportional to the so called scale factor  $s > 0$ .  $(u_0, s_0)$  is defined to be a modulus maximum if  $|Wf(u, s_0)|$  is a local maximum, i.e. if

$$\frac{\partial Wf(u_0, s_0)}{\partial u} = 0 \quad (2)$$

and if  $Wf(u, s_0)$  is strictly increasing to the left of  $u_0$  or strictly decreasing to the right of  $u_0$ . In other words, the goal is to identify the isolated local maxima (along the position coordinate  $u$ ) of the wavelet transform  $Wf(u, s)$  for each scale  $s > 0$ . A connected curve  $\gamma$  in the scale-time plane is called “a maxima line” if  $(u, s) \in \gamma$  implies  $(u, s)$  is a modulus maximum. Since the modulus maxima are isolated at each scale  $s_0$ , for each  $(u_0, s_0)$  there is at most one curve  $\gamma$  that locally intersects the point  $(u_0, s_0)$  (but there could be none in general).

Let us define now a singularity  $t_0$  of the function  $f$  as a point for which the derivative  $\frac{df}{dt}$  is not defined. Modulus maxima carry a significant degree of information about the position of singularities, more particularly, it is possible to prove that for every singularity  $t_0$  of  $f$  one finds “a sequence” of modulus maxima  $(u_i, s_i)$  such that  $s_i \rightarrow 0$  and  $u_i \rightarrow t_0$  (Mallat and Hwang, 1992). The word “sequence” is used, and not maxima line: maxima lines could break up and stop before reaching the smallest scales, and this would be a problem in using maxima lines to identify singularities. Fortunately, there is a choice of  $\phi$  that is immune to this problem, indeed, if  $\phi$  is a derivative of a Gaussian, all modulus maxima belonging to maxima lines propagate to fine scales (Yuille and Poggio, 1986).

For sake of completeness, even though sharp isolated singularities of  $f$  are expected to be located at the abscissa of end points of significant maxima lines, this may not always be the case and there can be maxima lines that converge to perfectly smooth points of  $f$ .

A rigorous detection of the singularities requires also the computation of the Lipschitz regularity coefficient of  $f$  at the abscissa of the end point of the maxima lines (Mallat and Hwang, 1992). However, in the present methodology only the detection of significant maxima line as a possible indication of the presence of sharp singularities is used.

The example in Fig. 2 shows a function with a non smooth point at  $t = 717$  and the corresponding set of maxima lines. Three of the maxima lines clearly converge for  $s \rightarrow 0$  to 717, but there are some maxima curves converging to smooth points of the curve.

When dealing with signals with significantly large noise, the number of maxima lines may also be very large. In order to select only the most significant maxima lines, the parameter “propagation factor”, identified with  $\xi$ , has been introduced, with legal values  $0 \leq \xi \leq 1$ . This parameter determines the fraction of the total number of scales  $S$  that a

maxima line  $\gamma$  must intersect in order for this to be considered significant, i.e.

$$\gamma = \begin{cases} \text{significant} & \text{length}(\gamma) \geq \xi S \\ \text{not significant} & \text{length}(\gamma) < \xi S \end{cases} \quad (3)$$

The length of  $\gamma$  is computed from finer scales (high frequency) to coarser scales (low frequencies) and therefore only the maxima lines which propagate at least up to scale  $\approx \xi S$  are considered to be significant (note  $S$ , and therefore  $\xi S$ , is dependent on the specific discretization of the real continuous wavelet transform; there is of course an unavoidable degree of ambiguity in the choice of discretization parameters, but this does not affect significantly the quality or robustness of the results).

The discussion of maxima lines was in the context of continuous scale and continuous position variation, however in practice the identification of maxima lines is performed in a discrete setting, and therefore errors could occur that prevent the very last point of the maxima line to converge to the time-series singularity. In order to have a more robust estimate of the time at which the singularity occurs, the following equation has been defined

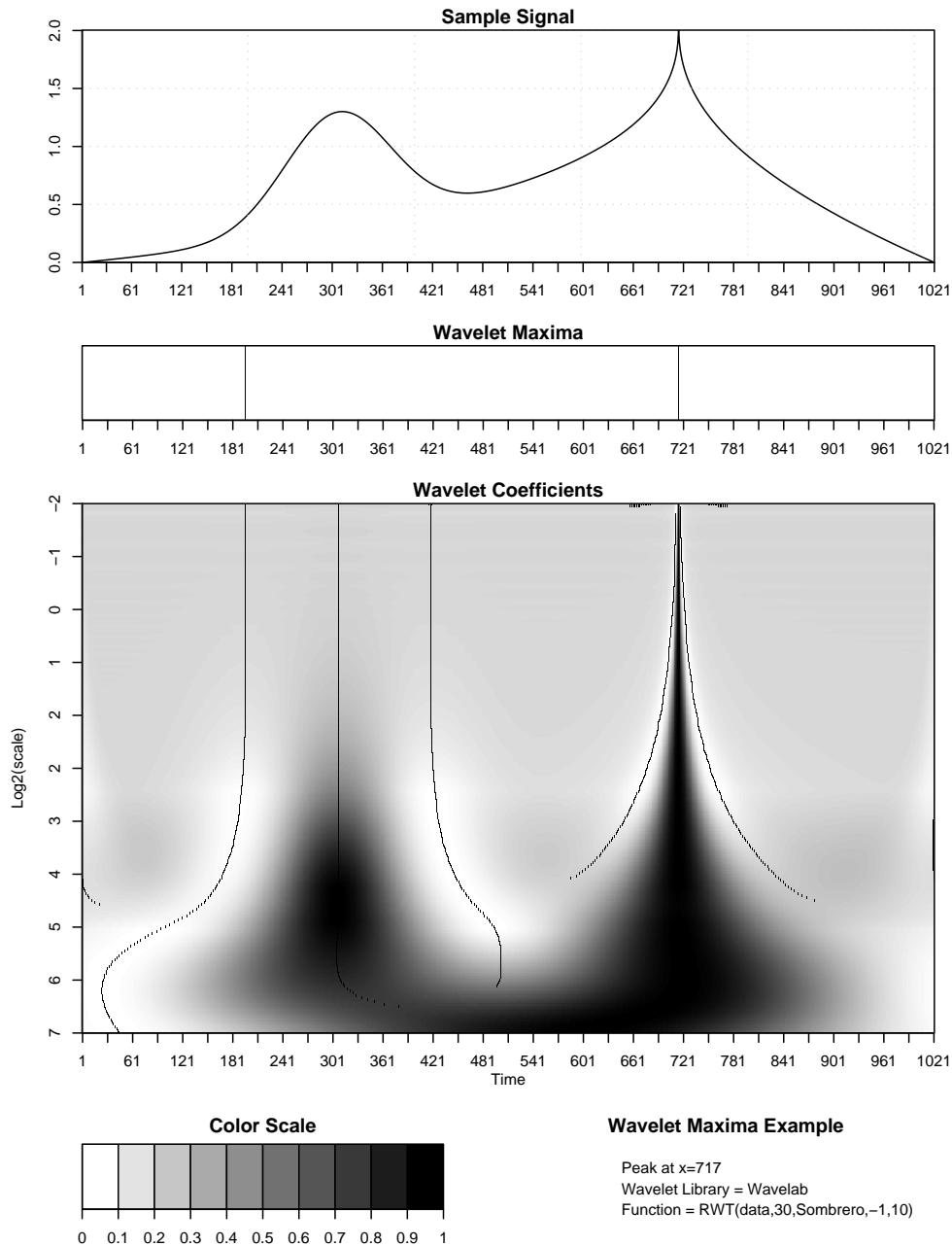
$$g(\gamma, n) = \frac{\sum_{i=1}^n \gamma_i}{n} \quad (4)$$

which corresponds to the average of the values at the abscissa of the last  $n$  end points of the maxima line, where  $n$  is a parameter either fixed or proportional to the number of scales, and  $\gamma_i$  is the abscissa value of  $\gamma$  at scale  $i$ . In case of daily data, as in the present study, the calculated value  $g(\gamma, n)$  corresponds to the approximate day of when the anomaly occurred. Initial experiments have shown that the average of the last 10 values ( $n=10$ ) yields a good accuracy in determining the days corresponding to the detected singularities. Smaller and larger values of  $n$  were affected by the overall trend or by noise introduced by the high frequency components of the signal.

The final result of our wavelet analysis is a characteristic vector of the same length as the original signal, where all values are set to zero except for those corresponding to the calculated values of  $g(\gamma, n)$  for all significant maxima lines  $\gamma$ . Such points are set equal to the length of  $\gamma$ , or in other words to the propagation of the maxima line. The resulting characteristic vector indicates the time when a singularity was detected, and its weight defined by the propagation of the corresponding maxima line.

In this methodology one dimensional, (1-D) wavelet transformations were used instead two dimensional (2-D) ones because the latter analysis cannot detect the exact geometry of the signal, and some similar one dimensional construction along the continental boundary or fault line may still be required. Therefore the algorithm was restricted to the 1-D wavelet analysis to avoid adding unnecessary complications to the present methodology.

This wavelet based approach has also the advantage of effectively filtering out the peaks due to small high frequency variations, because those are associated only with very short



**Fig. 2.** Example of wavelet analysis and corresponding maxima curves.

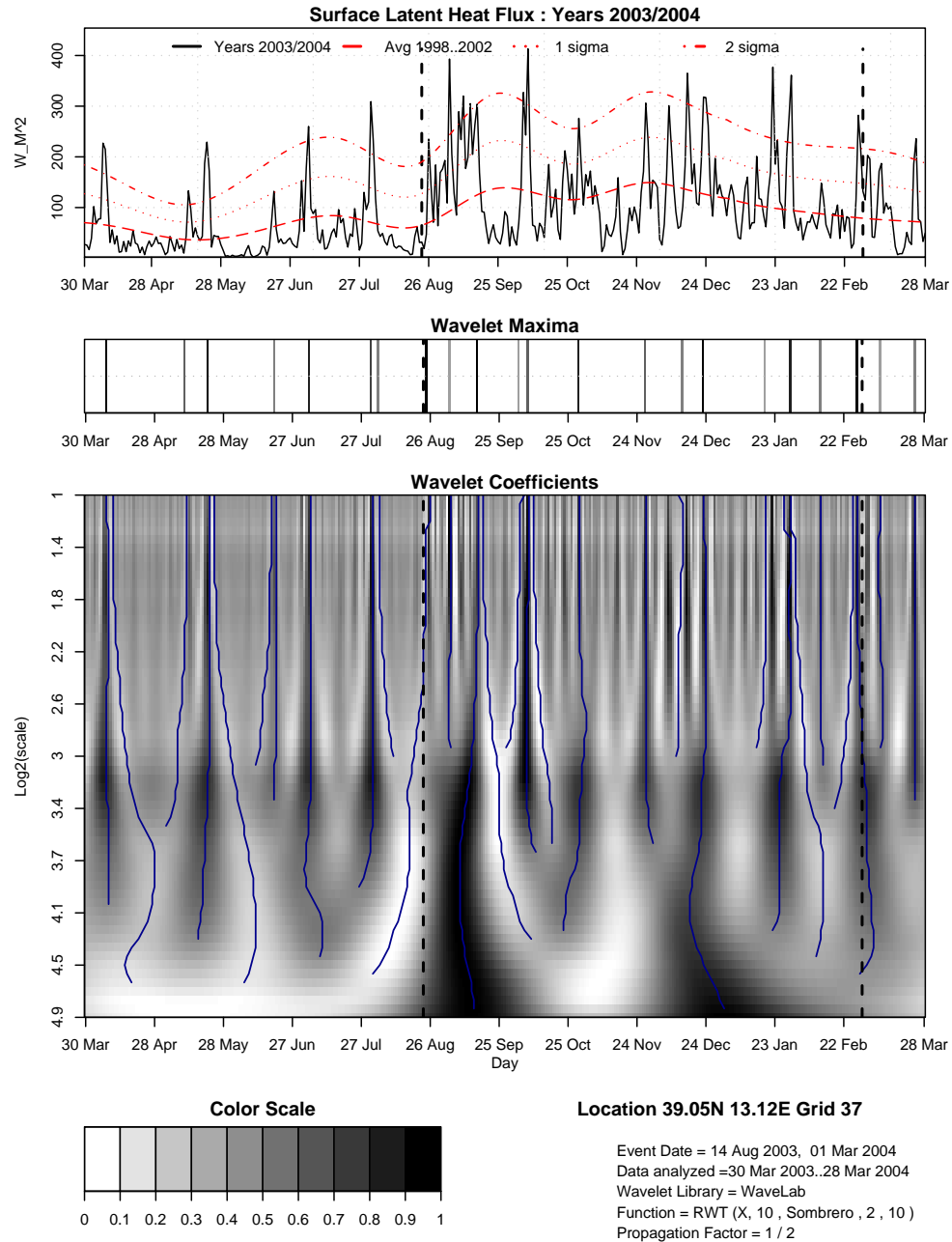
lines  $\gamma$ . Peaks caused by the seasonal trends are also disregarded because the length of  $\gamma$  is computed from finer to coarser scales, or higher to lower frequencies. Therefore, depending on the propagation factor  $\xi$ , it is not important if a line  $\gamma$  propagates all the way to the coarser scales, or lowest frequencies, which are affected primarily by seasonal trends.

### 5.2 Continuity of detected singularities in space and time

The wavelet analysis discussed in Sect. 5.1 has been applied to the time series of several grids for the epicentral region. The result of the analysis may identify several anoma-

lous peaks, which may be caused by earthquakes or by atmospheric or oceanic perturbations. For example, SLHF is an atmospheric parameter which is directly correlated to the evaporation of water on the surface. SLHF is particularly affected by changes in land surface temperature over the land, and sea surface temperature over the oceans. The origin of the increase or decrease of the surface temperature can be due to atmospheric perturbations, such as strong winds, precipitation, intense cloud cover, or due to geological phenomena which cause a drastic increase in surface temperature..

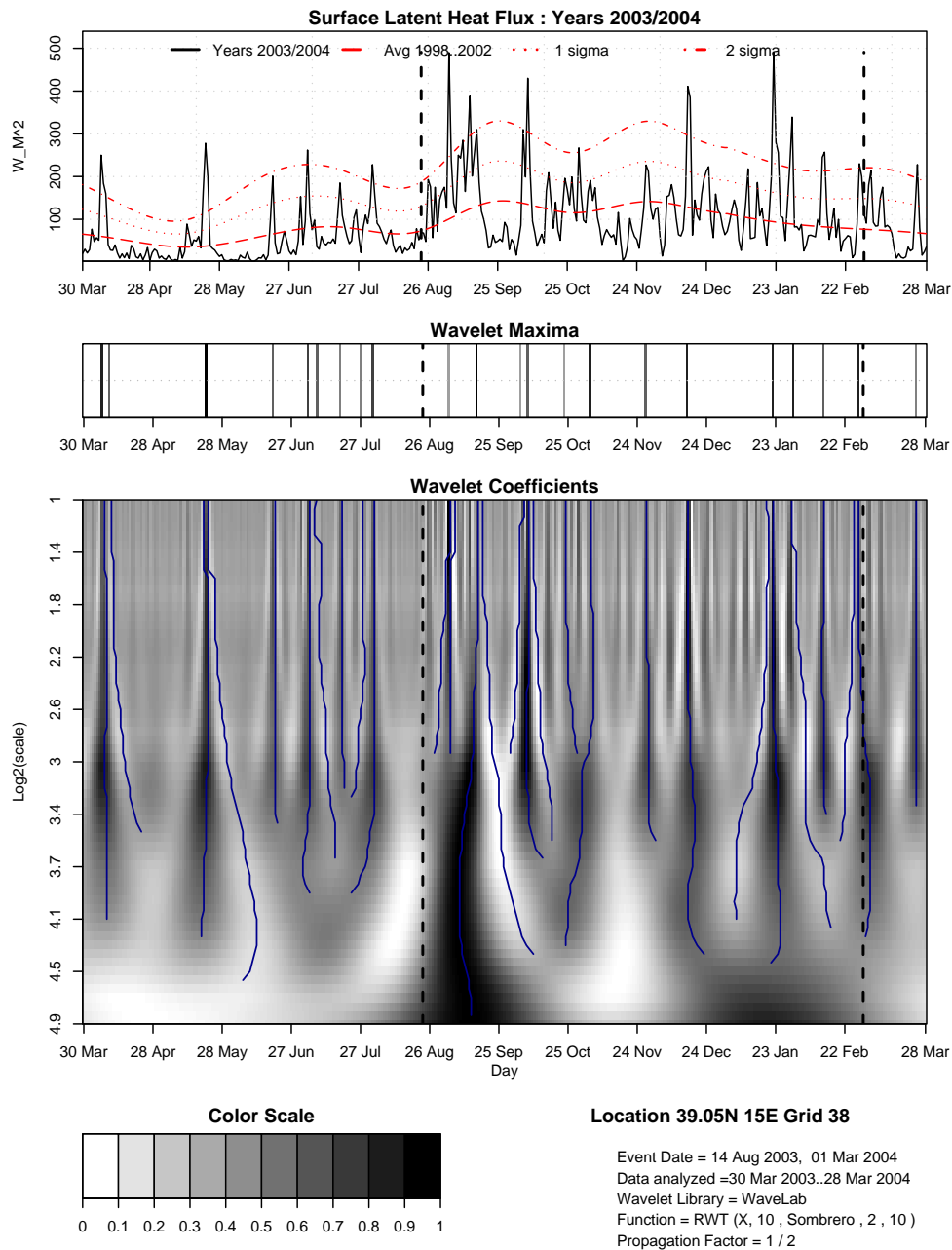
In the present methodology, such disturbances are filtered starting from the assumption that seismic activity is



**Fig. 3.** Wavelet analysis for SLHF data over Grid 37 from 30 March 2003 to 28 March 2004.

a large scale phenomena which manifest with a precise geometry conforming to the geological properties of the region, peaks caused by earthquakes can be discriminated from peaks caused by atmospheric perturbations by using the concept of geometrical continuity under space and time constraints. A detailed analysis of the detected anomalies has shown that peaks associated with earthquake events appear over a large area, and also follow a precise geometry, determined primarily by the geological conditions of the region. This means that a significant anomaly in one grid is related to significant anomalies in other grids if they all follow a precise geometrical path conforming to geological settings such

as continental boundaries, fault lines, orography, sea depth, and all the other characteristics that play a role on how the anomalies spread around the epicentral region. A collection of grids that satisfy these geometrical constraints is called a “grid path”. Furthermore, anomalies over the chosen grid path must appear within a very short time range (usually 1 or 2 days), adding an additional constraint of continuity in time. All these spatial and temporal constraints can be used to “link” in a suitable way the singularity detected in each single grid to discriminate signals which are most likely due to an impending earthquake, from those due to other phenomena.

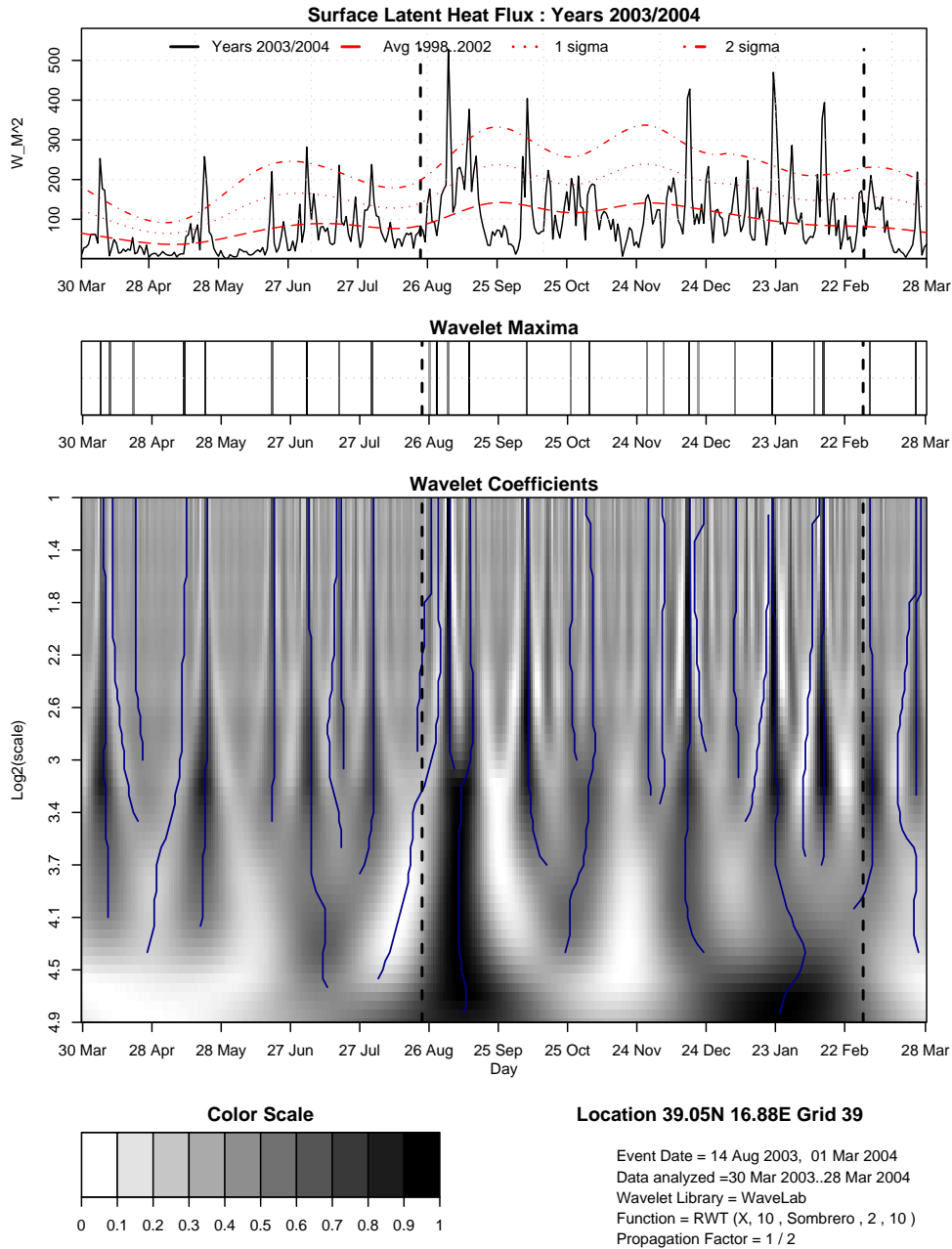


**Fig. 4.** Wavelet analysis for SLHF data over Grid 38 from 30 March 2003 to 28 March 2004.

However, the behavior of the various parameters may not follow exactly the grid path defined by the geological properties of the region, primarily due to atmospheric disturbances or discontinuity in the sub-surface. Therefore alternative grid paths must be chosen taking into account the effect of such factors. The selection of grid paths can be viewed as a search in a two dimensional space, where the dimensions are width and height (or longitude and latitude). This selection can be done using a specialized search technique such as evolutionary algorithms, or heuristics that take into account different factors which may affect the shape of the signal. Initial experiments suggest that grids can be selected simply by taking

into account the continental boundaries or fault lines. Moreover the time-series measurements are sufficiently smooth in space and time to make arguably the selection of the grids robust under small perturbations of the grid path; this would be the case particularly if higher resolution data were used. However, the SLHF data used in the present study is available only at a coarse resolution, roughly  $1.9^\circ$  by  $1.9^\circ$ .

The results of the wavelet analysis for the different grids in the grid path are combined in a  $n$  by  $m$  matrix  $N(n \times m)$ , in which the rows  $n$  correspond to the grids of the grid path at which the wavelet analysis has been performed, the columns  $m$  correspond to time, and the values are either the



**Fig. 5.** Wavelet analysis for SLHF data over Grid 39 from 30 March 2003 to 28 March 2004.

propagation length of a significant maxima lines or zero if none has been detected at this particular point in space/time. A continuity line  $\chi$  is a signal which propagates in space and time across all non-zero cells of  $N$  for at least  $minLength$  rows and within  $maxDisc$  columns, where  $minLength$  and  $maxDisc$  are user defined parameters. The number of valid continuity lines could be potentially very large, and therefore it is necessary to introduce the following constraints to limit the search space:

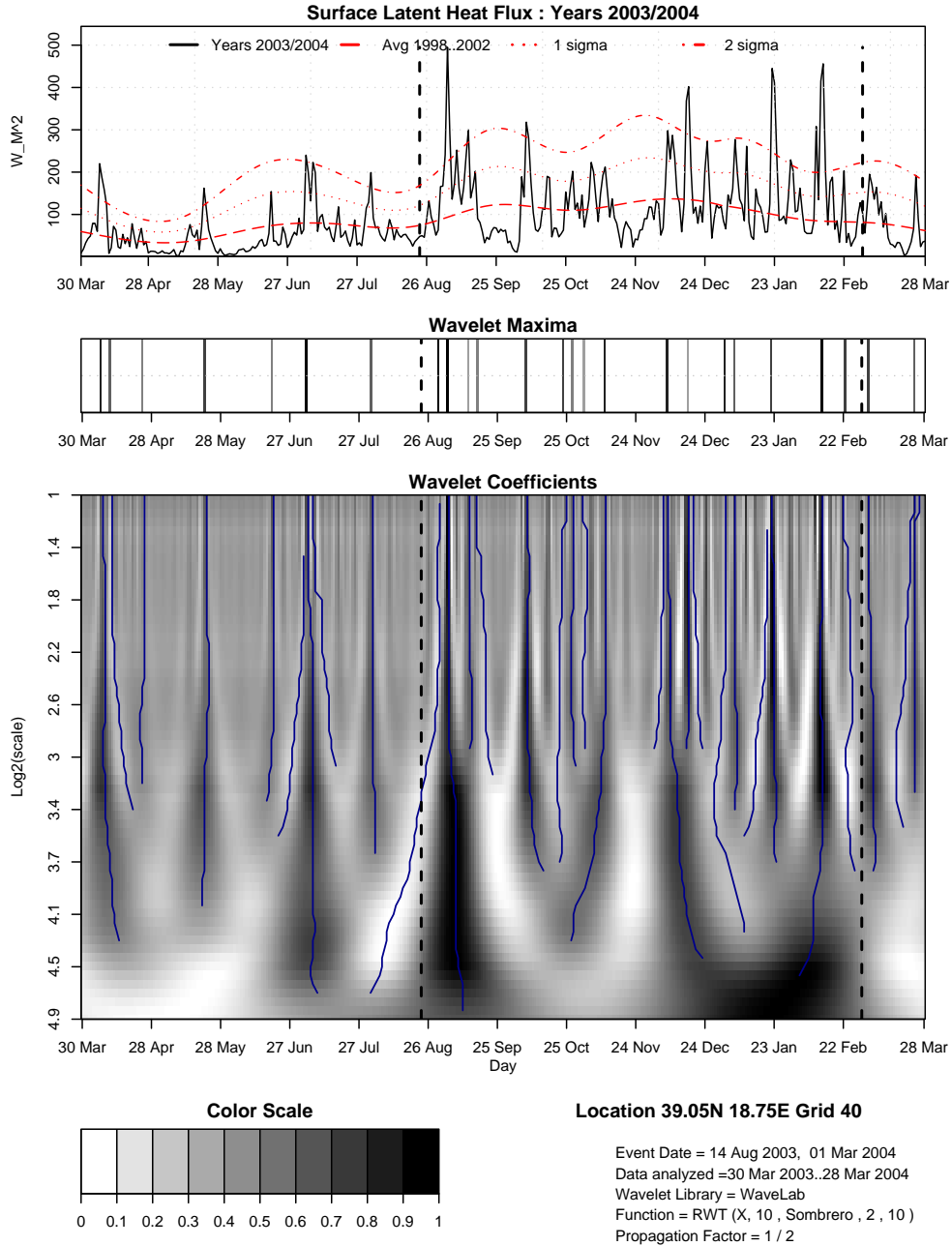
- C1: The minimum length of  $\chi$  must be larger or equal than the parameter  $minLength$ .

- C2: The time discontinuity can be at most  $\pm maxDisc$ .
- C3: The space discontinuity can be at most one consecutive grid, but cannot be at either the first or the last grid.

These constraints and their formulas of derivation are formalized in Eq. (5)–(8).

$$\min(|\chi_c|) \geq minLength \tag{5}$$





**Fig. 6.** Wavelet analysis for SLHF data over Grid 40 from 30 March 2003 to 28 March 2004.

specifying that the length of the continuity line  $\chi$  at time  $c$  must be greater than or equal to  $minLength$ , where

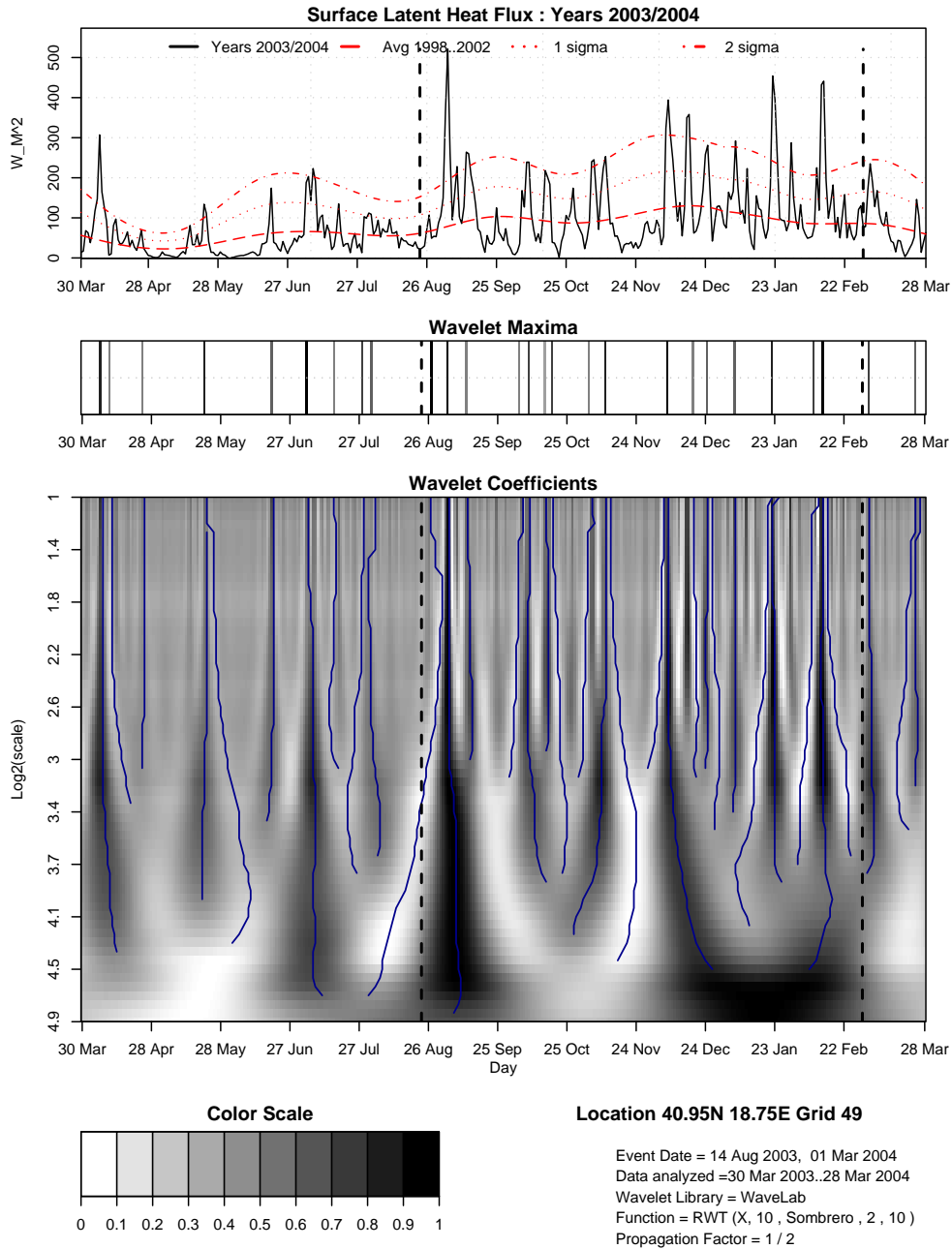
$$|\chi_c| = \sum_{r=1}^n (A_{r,c} - \sum_{r=1}^{n-2} (D_{r,c} * PC)). \quad (6)$$

This length corresponds to the number of detected anomalies  $A$  minus the number of discontinuities for the chosen grid path  $D$ . Discontinuities are penalized by the parameter  $PC$ , which indicates a value that is subtracted for each discontinuity. For example, if  $PC$  is equal to 0.5 a path of length 5 with no discontinuity is equivalent to a path of length 6 with two discontinuity. Alternatively, it is possible

to add an increasingly larger penalization coefficient to each new anomaly, such that paths with many discontinuities are penalized more heavily.

$$A_{r,c} = \begin{cases} 1 & N(r,k) > 0 \\ & c - maxDisc < k < c + maxDisc \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

An anomaly  $A$  for grid  $r$  and time  $c$  is detected if a wavelet maximum exist for grid  $r$  and time  $k$ . Time  $k$  is defined as the interval of the original time  $c \pm MaxDisc$ , and it represents



**Fig. 7.** Wavelet analysis for SLHF data over Grid 49 from 30 March 2003 to 28 March 2004.

the maximum time discontinuity allowed.

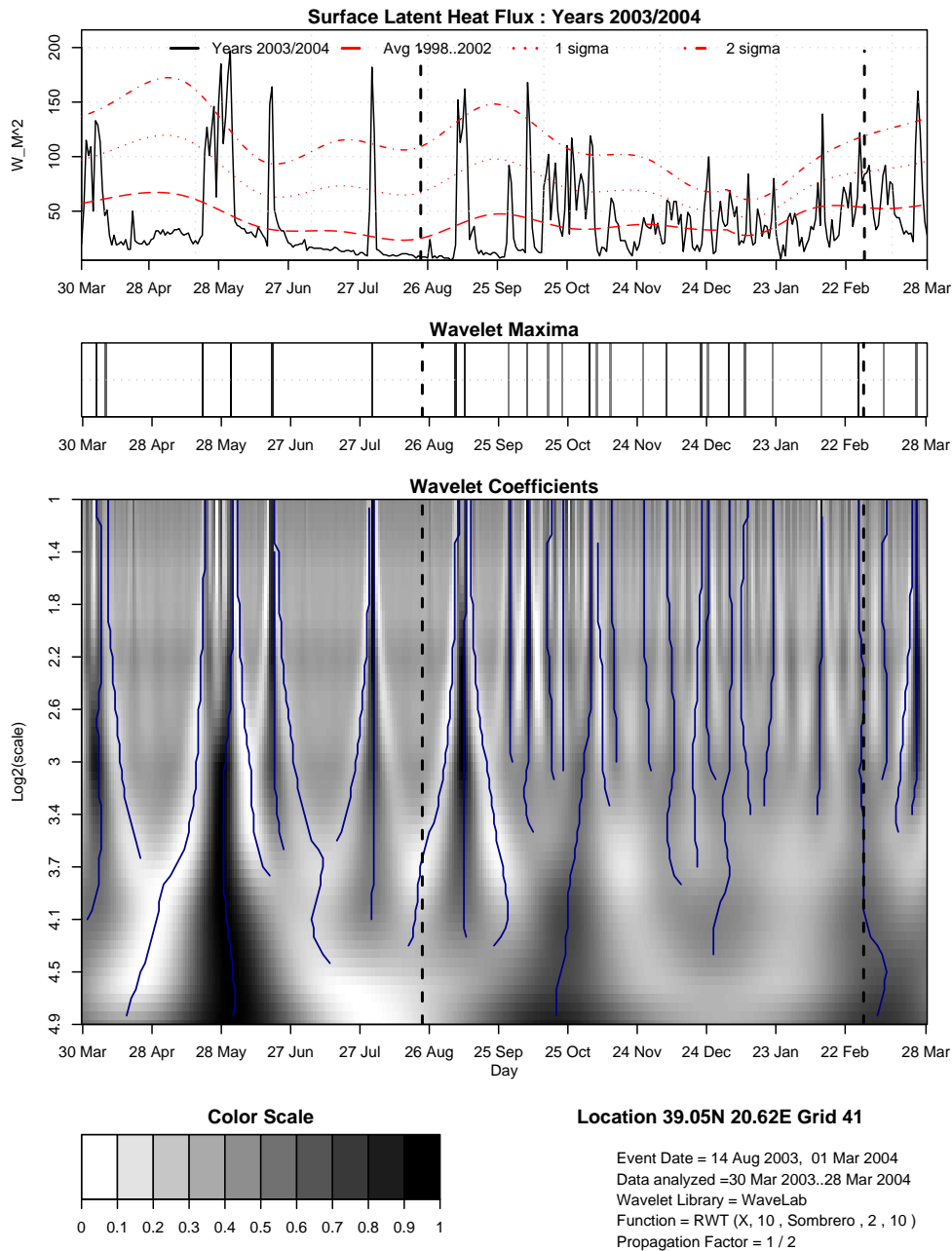
$$D_{r,c} = \begin{cases} 1 & \begin{cases} N_{r,k} > 0 \\ N_{r+1,k} = 0 \\ N_{r+2,k} > 0 \\ c - \text{maxDisc} < k < c + \text{maxDisc} \end{cases} \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

A discontinuity for grid  $r$  and time  $c$  is detected if a wavelet maximum exists for grid  $r$  and grid  $r + 2$ , but does not exist for grid  $r + 1$ . The time  $k$  is defined as given in

Eq. (8), and there can be at most a discontinuity of a single grid.

The discontinuity in space and time has been introduced to compensate for rounding errors in the wavelet transformation, the approximation due to the averaging of the end points of the significant maxima lines, and also for the possible geological anomalies and atmospheric perturbations previously mentioned.

Each path generated is evaluated according to different criteria in order to determine the most significant continuity lines. The evaluation is computed using pairs of the form  $K\tau$ , where  $K$  is a criterion, and  $\tau$  is a tolerance expressed in



**Fig. 8.** Wavelet analysis for SLHF data over Grid 41 from 30 March 2003 to 28 March 2004.

percentage. For each criterion  $K$ , it computes the maximum value and selects only the paths which are within a given tolerance of the maximum value. Only the paths which satisfy all criteria are considered to be significant. Three different criteria  $K$  have been used:

1. Maximum length

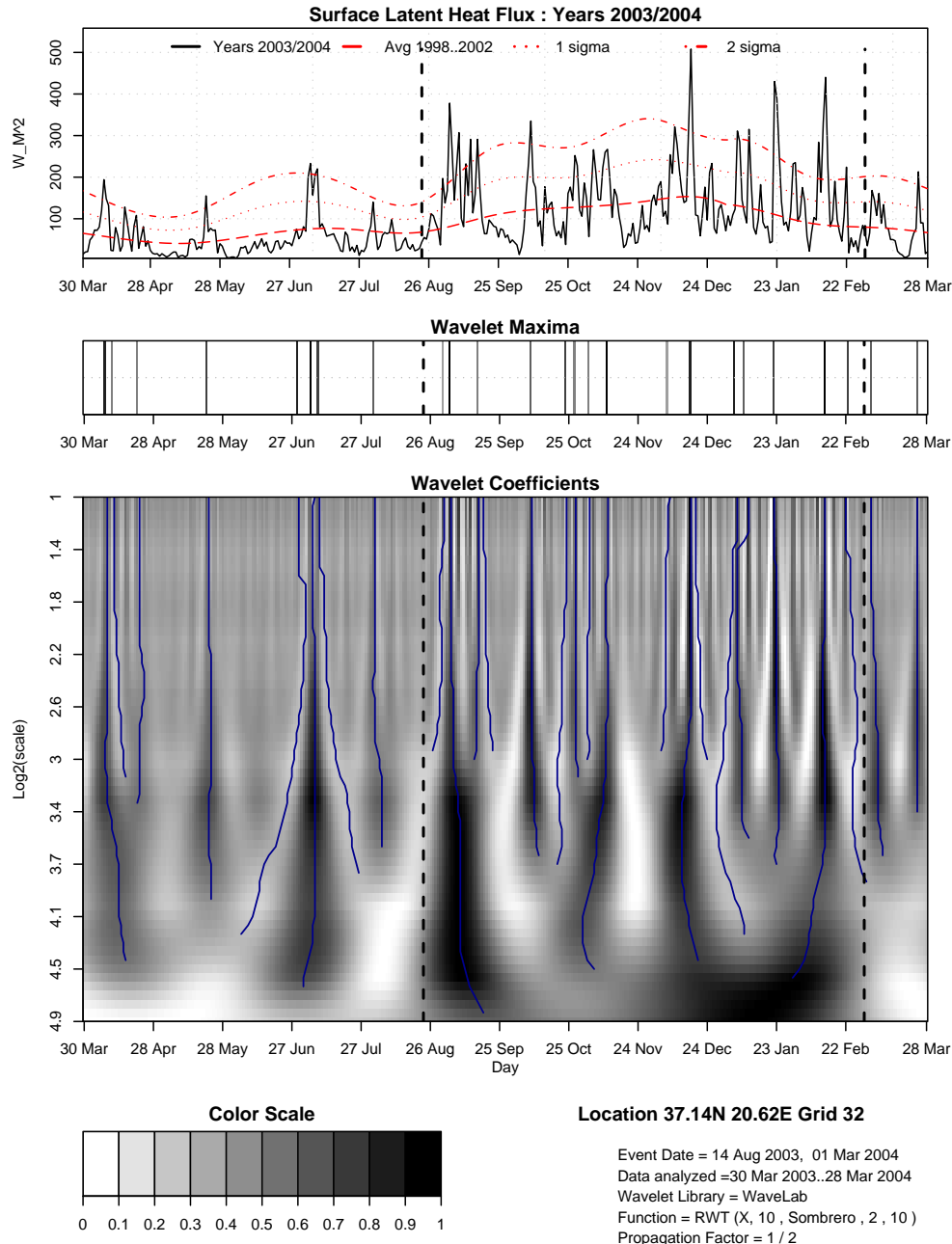
This criterion evaluates the length of a continuity line  $\chi$  by summing the number of grids where a propagating signal is present. As described earlier, a discontinuity in a single grid is allowed to compensate errors which could be due to anomalies along the selected grid path due to rounding errors or inhomogeneity in the subsurface.

2. Minimum spread

This criterion evaluates the spread in time of a signal, penalizing the signals that propagate across the fault line with a rather large discontinuity in time. For example, let us assume a  $minDisc=2$  (days), a signal could be always moving two days forward in time across the fault line, appearing in grid 1 at time 100, grid 2 at time 102, grid 2 time 104 and so on. This criterion evaluates the standard deviation (sigma) for all paths, and keeps only the paths with smaller values.

3. Maximum anomalies

This criterion takes into consideration the statistical significance of the anomalies associated with earthquake



**Fig. 9.** Wavelet analysis for SLHF data over Grid 32 from 30 March 2003 to 28 March 2004.

events. Each wavelet maxima corresponds to a peak of the original signal, and its statistical significance is proportional to the size of the anomaly. Initial experiments have shown that large SLHF anomalies are usually one to two sigma above the mean value.

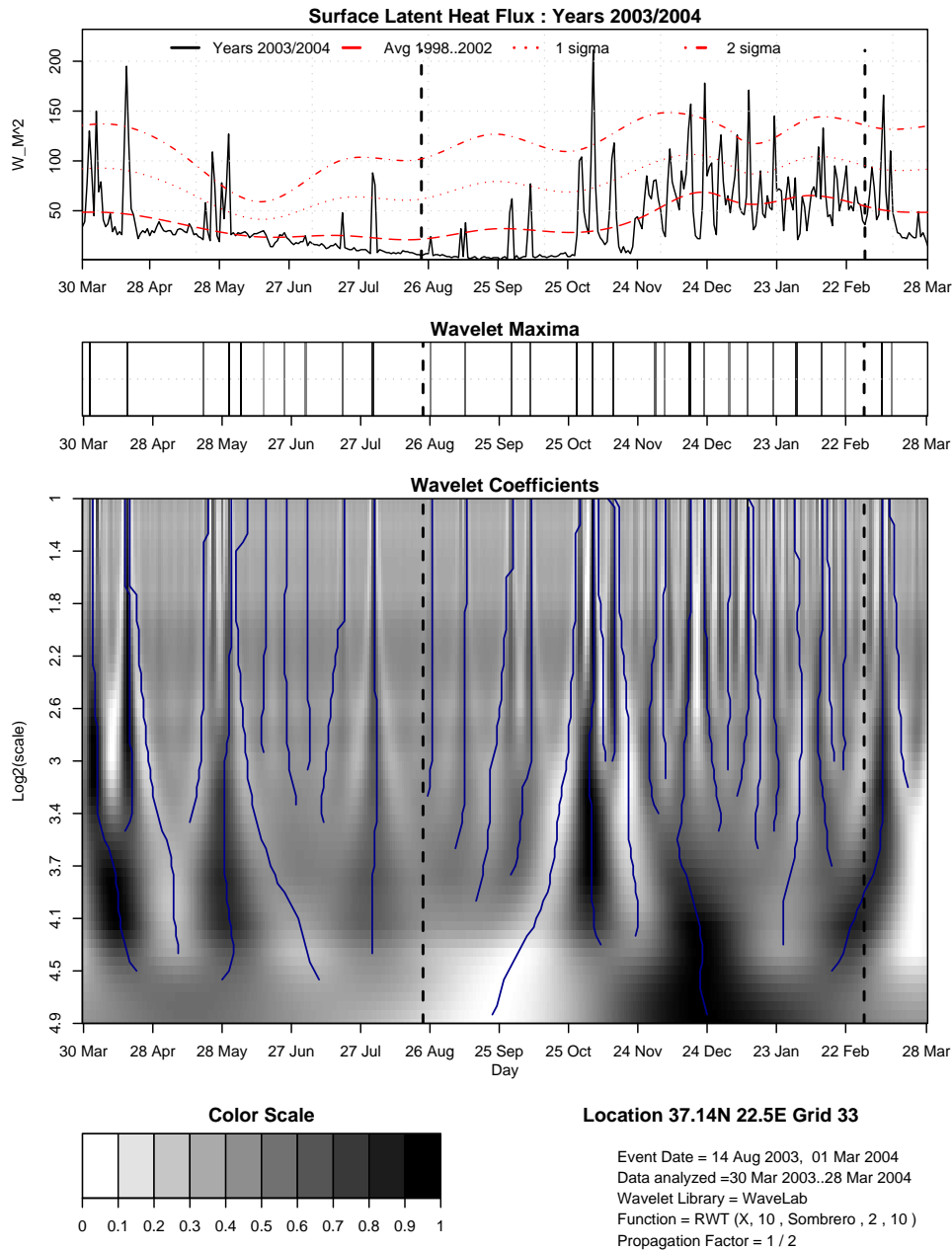
Finally, in order to compare different earthquakes, the concept of anomaly  $A$  is introduced here and used throughout our analysis, and is defined as:

$$A = \frac{X_i - M}{STD}, \quad (9)$$

where  $X_i$  is the SLHF value at time  $i$ ,  $M$  is the 30-days average for the previous years, and  $STD$  is the standard deviation of  $M$ .

## 6 Results and discussion

Experiments have been performed to establish the validity of the original hypothesis that SLHF shows a detectable anomaly prior to an earthquake. Detailed analysis of SLHF has been carried out for the Greek earthquakes of 14 August 2003 of magnitude 6.7 and of 1 March 2004 of magnitude



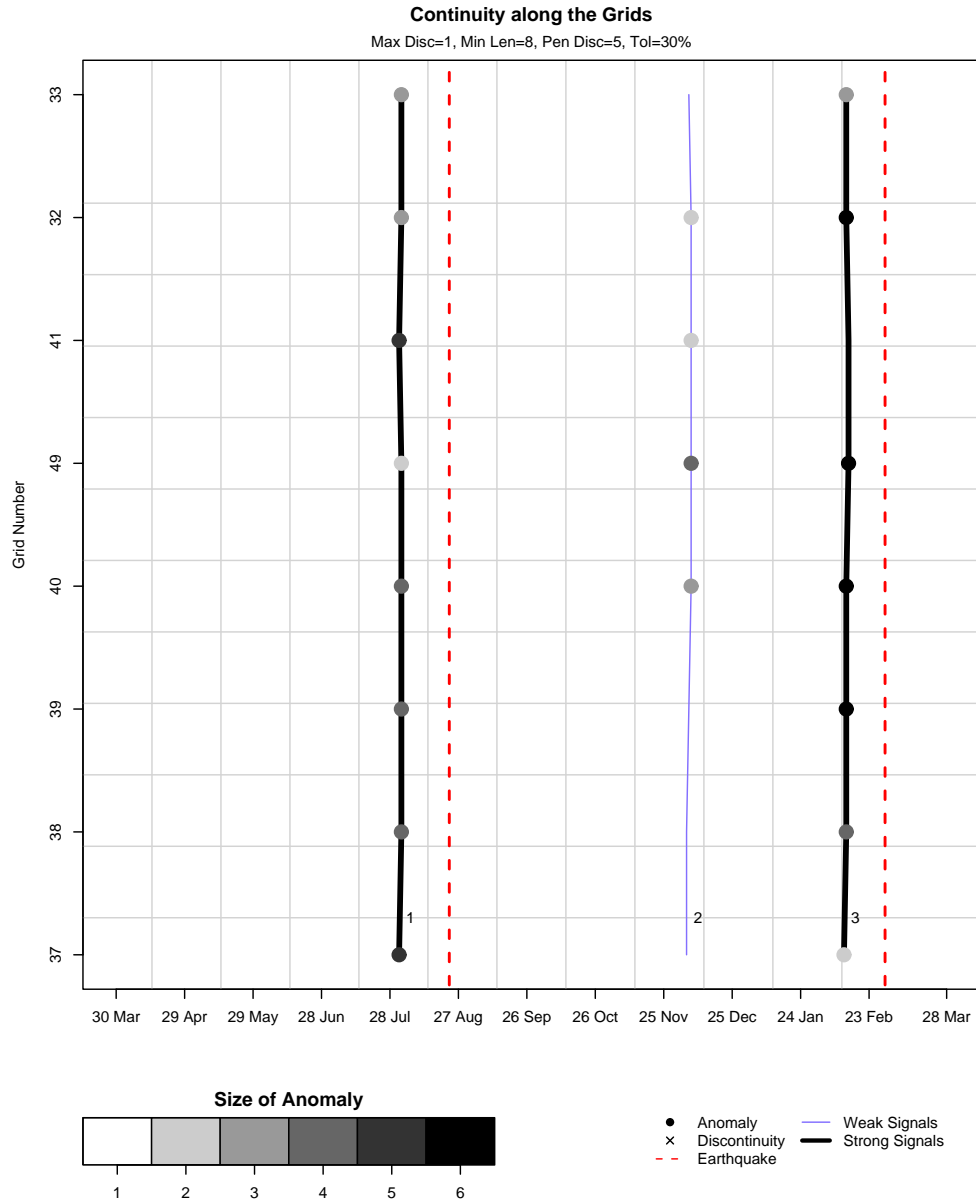
**Fig. 10.** Wavelet analysis for SLHF data over Grid 33 from 30 March 2003 to 28 March 2004.

5.7. Experiments have been carried out over the stronger 2003 earthquake, and the results validated using the same grid path for the weaker 2004 earthquake. The grid path covering grids 37, 38, 39, 40, 49, 41, 32 and 33 (Fig. 1), has been used to determine the existence of geometrical continuity of anomalies over the continental boundary. The epicenters of the two earthquakes are located in grids 41 and 33, respectively, and are shown with stars in (Fig. 1).

The wavelet analysis has been performed using 365 SLHF data points, from 30 March 2003 to 28 March 2004, for each of the grids in the grid path. The results of the wavelet analysis are shown in Figs. 3 to 10, and the parameters used are summarized in Table 1. Each of these figures contains three parts:

**Table 1.** Parameters of the numerical real wavelet transformation (RWT) used in the experiments.

Function	RWT
Mother wavelet	Sombrero
Nvoice	10
Scale	10
Octave	2
Propagation factor	$\frac{1}{2}$

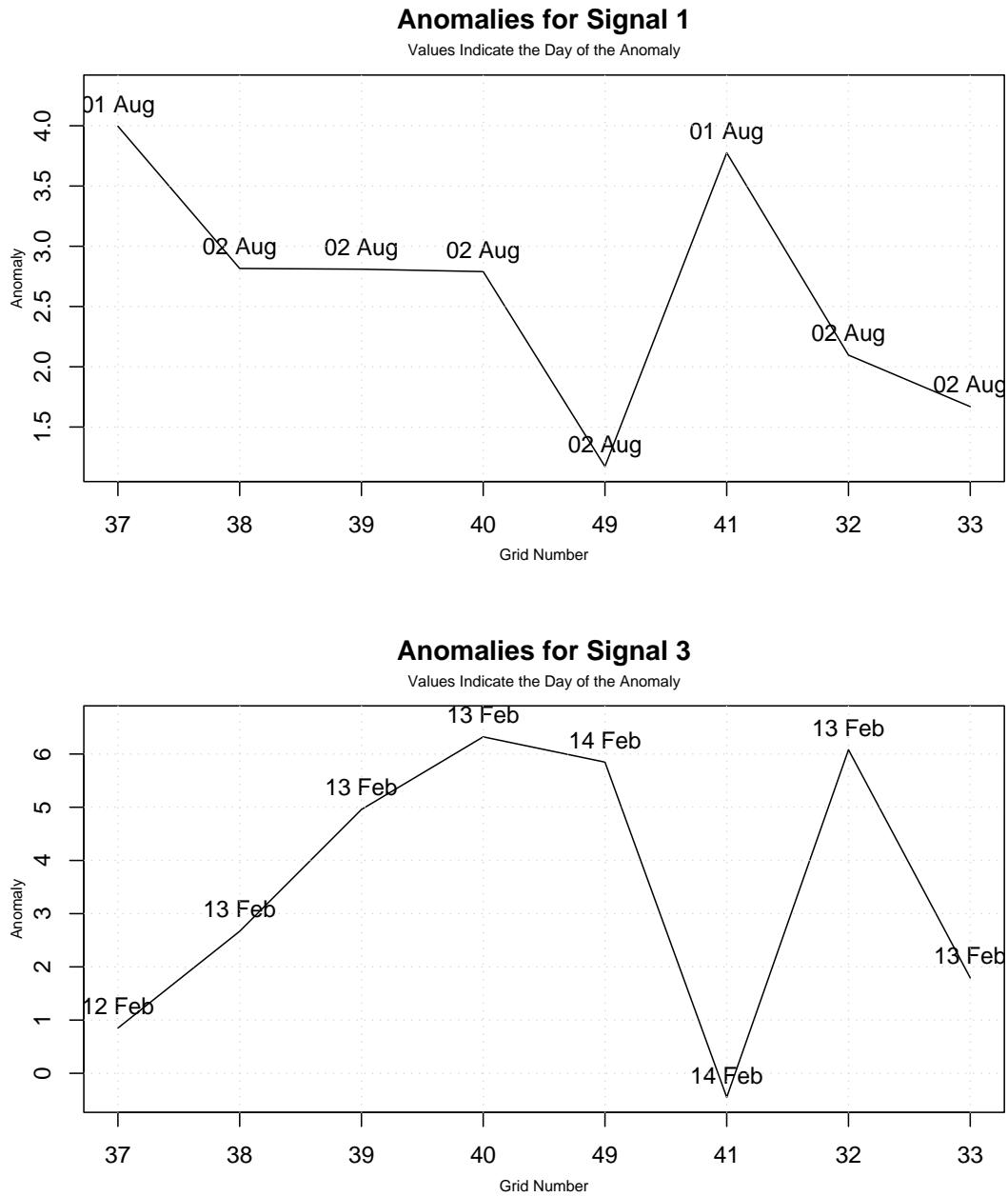


**Fig. 11.** Results showing all the possible paths of the wavelet maxima computed in the wavelet transformations.

1. The first part shows the time series for the original signal, the 30-days average for the previous five years 1998–2002, and the 1 and 2 sigma line for the 30-days average.
2. The second part is a graphical representation of the characteristic vector described in Sect. 5.1, which shows the time when significant wavelet maxima are detected. The color indicates the propagation depth in order to emphasize those maxima which propagate to the finer scales.
3. The third and last part consists of a graphical representation of the wavelet coefficients, and the corresponding maxima lines. It is possible to notice how the maxima lines converge to the lines indicated in the previous part.

Figure 8 shows the wavelet transformation for the epicentral region of the 2003 earthquake. A sharp SLHF peak well above two sigma (dashed line) is seen about two weeks prior to the earthquake. Similarly Fig. 10 shows the wavelet transformation for the epicentral region of the 2004 earthquake. A sharp SLHF peak above one sigma (dotted line) is also seen about two days weeks prior to the earthquake. The lesser intensity of this peak is probably due to the smaller magnitude of the 2004 earthquake.

The analysis of spatial and temporal continuity (Fig. 11) illustrates the significant paths generated using the results of the wavelet analysis, under the specified time and space constraints. In the figure the x-axis represents time expressed in days, and the y-axis represents the grid of the grid path. The strength of the anomaly is color coded.



**Fig. 12.** Size of the anomalies and their time of occurrence for the two precursory signals of the 14 August 2003 earthquake (signal 1) and of the 1 March 2004 earthquake (signal 3). The epicenters of the two earthquakes are found in grid 41 and grid 32, respectively.

Out of a 365 days period, only three continuous paths are found which satisfy the space and time constraints. One of these paths occurred about 12 days prior to the earthquake of 14 August 2003, and another occurred about 14 days prior to the earthquake of 1 March 2003. Although the third signal satisfies the given constraints of time and space continuity, it appears to be much weaker than the other two, especially because the detected singularities are not very strong. Both the stronger signals can be considered as strong precursors of the two earthquakes. The parameters used to compute the spatial and temporal continuity are given in Table 2.

**Table 2.** Parameters used to construct the continuity lines described in the experiments.

Max discontinuity	1 day
Min length	8 grids
PC	2
Tolerance	20%

Figure 12 shows the SLHF anomalies for the two significant signals. The values on the graph indicate the day when each of the anomalies was registered. It is possible to notice that in the two cases the largest anomaly occurred either at the epicenter (earthquake of 14 August 2003) or in the grid immediately next to the epicenter (earthquake of 1 March 2004).

## 7 Conclusion

The methodology discussed in the present paper uses data mining techniques, including wavelet transformations and spatial/temporal continuity analysis of the wavelet maxima to identify earthquake precursory signals. The methodology has been applied using SLHF data to study two earthquakes occurred in Greece on 14 August 2003 and on 1 March 2004. These two earthquakes are chosen since they occurred in the same region and within a short time span. In both earthquake events significant SLHF anomalies are found to be associated prior to the earthquake.

The use of 1-D real continuous wavelet transformation combined with the study of spatial and temporal continuity of the anomaly over the continental boundary have shown SLHF anomalies as earthquake precursors. For both earthquakes the observed precursory signals are seen about two weeks prior to the main events. The precursory signals associated with the earthquake events follow a continuity in time and space which can be used to discriminate from other signals due to different weather and atmospheric processes.

*Acknowledgements.* We wish to thank L. Panait for his comments and suggestions on how to select grid paths, J. Guillory and D. Struppa for the many discussions and support. We wish to thank J. Kalsmith and S. Mirza for their comments on how to improve the visualization of the wavelet analysis. We also wish to thank the two anonymous reviewers for their fruitful comments and suggestions. This research was partially supported by NASA's office of Earth Science Enterprise under grant NAG12-01009, NAG13-02054 and NAG13-03019, VAccess/MAGIC projects.

Edited by: M. Contadakis

Reviewed by: C. Kotsakis and another referee

## References

- Barnes, R., Watkins, K., Gairdner, R. A., and W., T.: A Geologic Interpretation of the African/Eurasian plate boundary near Greece, <http://www.owl.net.rice.edu/~uloveme/plate.htm>, 2003.
- DeMets, C., Gordon, R., Argus, D., and Stein, S.: Current plate motions, *Int. J. Geophys.*, 101, 425–478, 1990.
- Dey, S. and Singh, R. P.: Surface latent heat flux as an earthquake precursor, *Nat. Haz. Earth Sys. Sc.*, 3, 749–755, 2003.
- Dey, S., Sarkar, S., and Singh, R. P.: Anomalous changes in column water vapor after Gujarat earthquake, *Adv. Space Res.*, 33, 274–278, 2003.
- 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., and Joseph, D.: The NCEP/NCAR 40-year reanalysis project, *Bulletin American Met. Soc.*, 77, 437–471, 1996.
- Mallat, S. and Hwang, W. L.: Singularity Detection And Processing With Wavelets, *IEEE Transactions on Information Theory*, 38, 617–643, 1992.
- Ouzonov, D. and Freund, D.: Mid-Infrared emission prior to strong earthquakes analyzed by remote sensing data, *Adv. Space Res.*, 33, 268–273, 2004.
- Papazachos, B. C. and Kiratzi, A.: A detailed study of the active crustal deformation in the Aegean and surrounding area, *Tectonophys.*, 253, 129–153, 1996.
- Qiang, Z.: Thermal Infrared Anomaly Precursor of Impending Earthquakes, *Pure Pur. A. Geoph.*, 149, 159–171, 1997.
- Singh, R. P., Bhoi, S., and Sahoo, A. K.: Significant changes in the ocean parameters after the Gujarat earthquake, *Curr. Sci.*, 80, 1376–1377, 2001a.
- Singh, R. P., Sahoo, A. K., Bhoi, S., Kumar, M. G., and Bhuiyan, C. S.: Ground deformation of the Gujarat earthquake of 26 January 2001, *J. Geol. Soc. India*, 58, 209–214, 2001b.
- Singh, R. P., Dey, S., Singh, V. P., Cervone, G., Sarkar, S., and Kafatos, M.: Prediction of Coastal Earthquakes Using Surface Latent Heat Flux Retrieved from Satellite Data, in *Proceedings of the World Congress on Natural Disaster Mitigation*, World Federation of Engineering Organisation, 2, 129–134, 2004.
- Thanassoulas, C. and Klentos, V.: The “energy-flow model” of the Earth's lithosphere. Its application on the prediction of the “magnitude” of an imminent large Earthquake, *Institute of Geology and Mineral Exploration (IGME), Athens, Greece, Open File Report A*, 1–20, 2001.
- Yuille, A. L. and Poggio, T.: Scaling theorems for zero crossings, *IEEE Trans. Pattern Anal. Machine Intell.*, 8, 15–25, 1986.