TIME SERIES ANALYSIS OF REMOTELY-SENSED TIR EMISSION PRECEDING STRONG EARTHQUAKES

EFTHYMIA PAVLIDOU
March, 2013

SUPERVISORS:
Mark van der Meijde, Associate Professor, Dpt of Earth Systems Analysis.
Chris Hecker, Lecturer/Researcher, Dpt of Earth Systems Analysis

TIME SERIES ANALYSIS OF REMOTELY-SENSED TIR EMISSION PRECEDING STRONG EARTHQUAKES

EFTHYMIA PAVLIDOU
Enschede, The Netherlands, March, 2013

Thesis submitted to the Faculty of Geo-information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Applied Earth Sciences— Natural Hazards and Disaster Risk Management

SUPERVISORS:
Mark van der Meijde, Associate Professor, Dpt of Earth Systems Analysis.
Chris Hecker, Lecturer/Researcher, Dpt of Earth Systems Analysis

THESIS ASSESSMENT BOARD:

Prof.Dr.Victor G. Jetten, ESA, ITC UTwente (chair)

Mark van der Meijde (ESA, ITC UTwente), Chris Hecker (ESA, ITC UTwente),

Prof.Dr.Steven M. de Jong (Universiteit Utrecht)

ABSTRACT

Many non-seismic phenomena have been reported in literature to precede earthquakes and have been considered as potential precursors. Unusual increases of ground temperatures and recorded thermal infrared emission belong to this category. However, it is challenging to define thermal anomalies and distinguish them from normal temperature fluctuations or signal variations. Recent advances in the field include a detection scheme which handles meteorological influences to isolate thermal anomalies by normalizing pixel values with the mean value of neighbouring pixels. The present study employed long-length time series analysis of hypertemporal thermal images, obtained from geostationary first- and second-generation meteorological satellites, to study the thermal field of two earthquake-struck areas: Bam, Iran and Van, Turkey. The performance of the existing detection scheme was assessed for the first time by means of imposing synthetic anomalies in the dataset and trying to retrieve them. Improvements in the existing scheme were applied. Furthermore, the STL method for decomposition of time series was investigated as a means of removing diurnal, seasonal and inter-annual patterns from the dataset in order to facilitate anomaly detection. The method was examined both as a self-standing approach and in combination with the existing scheme, also with the application of synthetic anomalies. The attempts to retrieve artificial anomalies, with either of the two approaches, revealed that the natural, pre-existing variability of the image (due to many potential factors, like altitude and terrain morphology) played a key role in the performance of the detection scheme. The results showed that the signal contained variability which could not be attributed to influences of diurnal, seasonal or inter-annual frequency. This variability sometimes obscured and sometimes highlighted the imposed anomalies. The decomposition was successful in defining diurnal and seasonal patterns without affecting the imposed anomalies, but more modeling is needed to extract non-earthquake-related signal variations and facilitate detection. A combination of STL and the existing methodology was applied in original datasets from Bam and Van, as it appeared that the decomposition improved the performance of the detection scheme. In both of the areas, the largest anomalous counts appeared prior to the earthquake in pixels located close to faults. An attempt was made to link, qualitatively, the appearance of non-earthquake-related anomalies with other factors (elevation, terrain and geological substrate). Although the findings indicate a possible relation between the occurrence of thermal anomalies and earthquakes, the methodological approach needs to be further refined in order to confirm and quantify this relation.

Keywords

Earthquake precursors, Thermal anomaly, Time series

ACKNOWLEDGEMENTS

I am deeply grateful for my supervisors, Mark van der Meijde and Chris Hecker. For the chance to work on such a fascinating topic, for all the challenges, for their patience, their guidance, for always being there for me, and for the immense amount of things they taught me. They are real coaches, and working with them has been a great experience.

I am greatly indebted to Harald van der Werf for his help, patience and contribution in the parts of IDL programming.

This study would have never been fully accomplished if the author of the R-packages for STL implementation, Ryan Hafen, did not have the courtesy to not only provide them, but also build them for me in the most recent version of R.

I consider myself very lucky for having had the chance to study in the ESA department and the whole ITC. This is a unique environment, I have met and worked with wonderful people, and there is no single moment of my MSc that I do not cherish.

Special thanks to David Rossiter for his guidance and advice, Janneke Ettema for the support and insight, Ben Maathuis, Petra Budda, Bas Retsios and Ali Bagislayici for technical assistance and their welcoming willingness to advice and help. To Tolga Gorum for providing geological data for Van. To Giovanni Buzzo for his introduction to technical aspects I would encounter. To my colleagues Rory Nealon, Manuel Guilberto Garcia Alvarez and Rodrigo Alejandro Lopez Rangel for providing valuable material. To Mila Luleva, for the best freshman reception I could ask for. To Bezaye Tesfaye, Adriana Luz Garcia Guatame and Nuno Cesar De Sa for their feedback and support. To all my AES colleagues and my ITC family, for this amazing one-and-a-half year and all the unforgettable times together, in and out of the clusters.

It is customary to also thank the family, but in my case this is not a mere formality. Without them I just wouldn't be here.

I cannot help believing that we shall gradually find out the physical mechanism by which these [relationships] are maintained...
Sir Gilbert T.Walker (1918), on the Southern Oscillation Index

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1
Introduction

Tectonic earthquakes occur as the result of a process of gradual build-up of elastic strain energy and stress, leading to failure and sudden release of energy with often catastrophic results. Although the main mechanisms of seismic activity have been studied in detail, the complexity and the diversity of the contributing processes have prevented so far the development of an earthquake forecasting system.

There have been many approaches to earthquake prediction. Tiampo and Shcherbakov (2012) mention two kinds of forecasting models. In one category fall models which implement techniques based on the identification of physical processes underlying seismicity. In the other, models which treat seismicity in a more mathematical and/or probabilistic manner. As suggested by the authors, the first type of models may be affected by the complexity of the analyses, by oversimplifications and/or by real-life environmental heterogeneities. On the other hand, smoothed seismicity models rely heavily on availability and accuracy of earthquake data; additionally, they cannot account for regions which may have earthquake potential on longer time scales but have shown no significant recent activity. In this sense, approaches based on physical processes can be considered to have more potential for development, as it can be argued that tackling with technical issues might be, under conditions, more realistic than waiting for a considerable amount of years to improve the quality of available catalogues and databases.

An ideal solution would combine the two approaches, by constant monitoring and statistical evaluation of physical processes related to earthquake occurrence. This would imply:

(a) the existence of a measurable phenomenon, related to earthquake occurrence,

(b) an appropriate technical solution for monitoring and measuring, and

(c) a suitable statistical approach for assessing the accumulated data.

Concerning point (a), a series of phenomena have been directly or indirectly connected to the occurrence of earthquakes. Potential precursors include, among others, “earth deformation, surface temperature alterations, gas and aerosol content, electromagnetic disturbances in the ionosphere” (Tronin, 2006). Of those, increases of Land Surface Temperature (LST) have been repeatedly documented to precede earthquakes of large magnitude (Yao and Qiang, 2012) , and the theoretical background behind the connection of LST with earthquakes is being extensively studied.

The applicability of Remote Sensing in monitoring temperature changes has increased the potential of their use as a precursor, starting with Gorny et al (in Saraf et al. (2008)) and continuing with constant efforts of enhancing the available satellite data analysis techniques (Tramutoli et al., 2001). Indeed, thermal infrared sensors are present in many operational satellites and a wealth of observations is increasingly available, supporting monitoring attempts at different scales of temporal and spatial resolution. Thus, Remote Sensing and the use of Thermal Infrared (TIR) sensors can play an important role in addressing point (b).

In respect to point (c), time series analysis is highly recommended for manipulating climatic data, because they are characterized by the existence of “trends and seasonal variations that can be modeled deterministically with mathematical functions of time and, additionally, observa-

tions close together in time tend to be correlated (serially dependent)"(Cowpertwait and Metcalfe, 2009). The application of time series analysis, after removal of seasonal trends and variations, can reveal the existence of temperature values which can be considered unusual when compared to reference data.

Thus, the use of satellite-recorded thermal data to monitor temperature increases seems very appealing as a potential earthquake forecasting approach. However, earthquake prediction on the basis of thermal increases is not possible yet. The most prominent challenges in the field include (i) the establishment of a sufficient definition of "thermal anomaly", which should exclude spatial and temporal climatic variations affecting TIR signals, and (ii) surmounting the difficulties of cloud and noise detection in TIR signal recording. In this framework, the present study attempted to explore and exploit the potential of satellite recorded temperature increases as earthquake precursor. The main focus was to assess and further develop a scheme for anomaly detection, and then investigate the relation between the occurrence of earthquakes and thermal anomalies on the basis of this scheme.

1.1 POTENTIAL UNDERLYING PROCESSES RELATED TO THERMAL ALTERATIONS PRIOR TO EARTHQUAKES

Temperature increases prior to earthquakes are being reported in literature for more than 20 years (Genzano et al., 2007; Ouzounov and Freund, 2004; Saraf et al., 2008, 2012; Tramutoli et al., 2005, 2001; Tronin, 2009a, 2006; Yao and Qiang, 2012; Aliano et al., 2008; Tronin, 2009b; Nicola Pergola, 2004; Pulinets et al., 2006). Researchers have used different approaches, different sensors, land surface and top-of-the atmosphere measurements, case studies in many places around the world, to describe abnormal temperature rises preceding earthquakes. Thermal anomalies have been reported for larger time scales (long-term thermal anomalies, Pulinets et al. (2006)) and considered as long-term precursor of strong earthquakes (Yao and Qiang, 2012); however, the largest part of the literature describes short-lived anomalies appearing 1-15 days prior to earthquakes with magnitude$>$4.7, lasting for a few days and sometimes reappearing for a few days after the main shock. The temperature increases range from 2-13^oC, affect areas of tenths to hundredths of sq.km and appear in the proximity of large faults on land or ocean almost everywhere in the world: China, India, Turkey, Iran, Italy, Greece, USA and Mexico among others. The reports are theoretically backed by the proposition that not all accumulated mechanical energy leading to the earthquake is released instantly during the event: part of the stress is channeled to other energy forms (e.g. electricity, thermal energy) already before the earthquake (Freund, 2003). However, more often than not, the research settings have been questioned and the results have been greeted with skepticism. Even though no-one has proven beyond dispute the relation between thermal increases and earthquakes, on the other hand no-one has proven the absence of any relation either. The complexity and variety of the interacting processes leading up to an earthquake make it challenging to standardize their study and this explains the diversity of available descriptions- or, actually, the absence of a single "typical" anomaly profile. As Freund notes (2003), 'some earthquakes produce recognizable non-seismic precursory phenomena' while others don't, and there has been no solid proof of a theory to provide physical explanation behind pre-earthquake events; but this doesn't defy the existence of the phenomena. It rather highlights the lack of sufficient knowledge of the underlying processes.

There exist different theories attempting to explain pre-earthquake thermal alterations; some of them can be seen as complementing rather than competing with each other.

-The *Earth degassing theory*, proposed by Quiang et al, 1991 (in Tramutoli et al. (2001)), is based on the assumption of increased **release of warm gas** in the proximity of faults, which is

intensified prior to an earthquake due to the formation of microcracks under the accumulating tectonic stress. Tramutoli et al. (2001); Tronin (2009a) further discuss the formation of a localized **greenhouse effect** due to the optically active content of the released gas (CO_2, CH_4, H_2).

-Ziqi et al. (2002) take the issue of heated flux further, to the **water content** of the Earth surface. They describe alterations in the hydraulic balance under the effect of mechanical stress (acoustic waves) and note that temperature abnormalities also include temporary decreases, implying that an earthquake-related thermal rise might be masked by a preceding surface temperature decrease connected to groundwater attributes.

-Pulinets et al. (2006) propose a mechanism of increasing air temperature and relative humidity (as registered by meteorological stations) resulting from **air ionization by radon gas**; the latter is considered as a catalyst for "attachment of gaseous aerosols with water molecules", water condensation and release of latent heat. Their suggestion concludes that the reason for the observed increase in thermal signal prior to earthquakes is the air close to the surface rather than the earth surface itself, and is backed by the similarity in the fluctuations of air temperature, humidity and radon concentrations around the earthquake areas they studied.

However, it could be argued that radon is not present at sufficient quantities in every earthquake struck region; underground water levels vary, and the effect of warm gas and greenhouse phenomena is highly susceptible to local weather conditions like prevailing winds (Ouzounov and Freund, 2004; Ziqi et al., 2002).

On the other hand,Freund (2002); Freund et al. (2007); Freund (2011) and Ouzounov and Freund (2004) argue that heat conductivity data do not correspond to the rapid increase and decrease of the thermal signal described to appear prior to earthquakes: it is highly unlikely that the thermal rises are a result of heat transfered from large depths. They describe a theory based on the **activation of electric charges** in mechanically stressed rocks. The so-called P-holes are charge carriers, abundant in dormant state as defect peroxy links ($O_3X - OO - YO_3$, with X,Y= Si, Al, etc) in most crustal rocks; this supports the geographic universality of the theory. When the rocks undergo stress, as is the case of the period leading to an earthquake, they are activated and propagate rapidly being increasingly present in boundaries and surfaces of increased curvature. Electric potentials and IR emission generated by dry rock under application of deviatoric stress have been measured in laboratory conditions. The p-hole activation theory, besides not being limited geographically, attempts to provide a unifying explanation on the existence of many non-seismic phenomena which have been proposed as potential precursors and yet, at first sight, they seemed too versatile to be connected by a common underlying physical process. According to Freund the theory could address the issue of pre-earthquake ionospheric perturbations, emissions in different wavelengths (mainly low-frequency emissions), and geoelectric and electromagnetic anomalies. In the case of thermal rise, it is suggested that the observed increases in TIR signal stem primarily from IR emission caused by the recombination of activated p-holes at the surface of stressed rocks.

As Ziqi et al. (2002) point out, there are many factors that can cause abnormalities in remotely sensed images. So it is possible that all the above theories, backed as they are by preliminary laboratory settings, hold truth and can explain (to different extents according to local situations) the temperature increases which are frequently reported to precede earthquakes. It is however evident that,

(a) there is a variety of processes which can contribute to pre-earthquake changes in the thermal signal

(b) there is no precise quantification available for the contribution of each process

(c) some of the processes imply a potential presence of factors (i.e.water vapour) which might interfere with the recording of the TIR signal by the satellite sensors.

The above pose challenges in the effort to record and model normal, expected thermal behaviour and consequently distinguish the potential anomalies and the processes related to them.

1.2 THERMAL REMOTE SENSING

In Remote Sensing, temperature estimations are obtained from the Thermal Infra-Red (TIR) region of the electromagnetic spectrum. Thermal sensors record radiance emitted from surfaces and measured at the top of the atmosphere. Different calibration methods are followed depending on the sensor, for ensuring proper quantification of the signal; depending on the sensor type, the data can be delivered readily calibrated or require further pre-processing. The recorded signal is converted to different forms, according to the desired application. There is a growing number of available spaceborne sensors with thermal bands and their technical specifications vary. The list of operational sensors includes NOAA AVHRR, ASTER, MODIS, OLS, TM and ETM+, VHRR (onboard polar-orbiting satelites) and MVIRI, SEVIRI and GOES Imager along with the Indian VHRR of Kalpana-1 and Insat satellites (onboard geostationary platforms). The spatial resolution for polar orbiting sensors varies between 60m-2.7km and is ranging between 3-8km for geostationary sensors. The smallest temperature increase recordable by the sensor, can be down to 0.05^oC. The same area of the earth's thermal field is recorded up to once every to 12 hours in the case of polar-orbiting, and once every 15-30 minutes in the case of geostationary platforms. Technological advances continuously improve the performance of the sensors, and more thermal bands are available nowadays allowing for combined use in different products and applications (cloud-removal, ozone monitoring, dust-cloud tracking, meteorological applications etc) (EUMETSAT, 2012; NOAA, 2012; ISRO, 2012).

The above make clear the diversity of options for application of thermal remote sensing. In order to make effective use of the available products, it is important to make sound choices in terms of type of measurement, type of sensor and processing steps, according to the specific needs of each project. In the present study the key factors involved in making these choices were:
(a) The need of a temperature-estimating measure least affected by calculations.
(b) The need to maintain a valid time series of recorded values. Proper co-location of the images is vital, to ensure that the values of the time series all come from the same pixel; otherwise comparisons and change detection would lose their meaning.
(c) The need for a frequent registration of thermal signal. In this way, even in the case of a short-lived anomaly and a temporary sensor malfunction, there would still be enough data available to detect anomalies.

Concerning the temperature estimator, observed radiances can be converted to brightness temperature values (BT) otherwise called equivalent blackbody temperatures (EBBT) to facilitate interpretation of the observations (Tjemkes, 2005). This term stands for the temperature a blackbody should have in order to duplicate the observed specific intensity. Conversions can also be done to Land Surface Temperature (LST). However, this would require atmospheric corrections. Atmospheric effects like scattering can have great influence on the recorded signal. Spectral transmittance depends on temperature and humidity vertical profiles (Tramutoli et al., 2005). Ground signals can be attenuated or enhanced by the presence of gas aerosols and suspended particles in the atmosphere, which in their turn vary with site, time, altitude and local prevailing conditions. There exist different methods for performing atmospheric corrections, like air-to-ground correlation or various algorithms (single-IR channel, split-window) (Zhengming and Zhao-Liang, 1997; Wan, 2008) ; however, the calculations included in pre-processing could introduce errors, which can be acceptable or not according to the application. Wan and Dozier (1996) specifically note that, "because of the difficulties in correcting for atmospheric absorption and emission, and sur-

face emissivity, the development of accurate LST algorithms is not easy... the accuracy is limited by radiative transfer methods, molecular and aerosol absorption/scattering coefficients and uncertainties in atmospheric profiles". As stated in relevant literature (Tramutoli et al., 2005) variations of the estimate of LST from $1 - 3^oC$ (enough to alter anomaly detection results) are expected solely as a consequence of an uncertainty in emissivity calculations of only 0.01; this would essentially imply that the use of BT values would be preferable for the scope of the present study. The choice can be further backed up by the underlying processes potentially included in the expected thermal alterations, which might not necessarily relate to the surface itself but to the near-surface air or to IR radiation from stressed materials. It is important to note that it is not possible to precisely distinguish which portion of the measured signal is related to which (thermogenic or radiative) process.

Choice of sensor type (polar-orbiting or geostationary) affects three major issues: temporal resolution, spatial resolution and image co-location. Polar-orbiting sensors provide better spatial resolution while geostationary offer better temporal resolution. The imaging swaths of the polar-orbiting spaceborn sensors are critical for the thermal signatures aquired. Wide swaths create greater viewing angles for the edges of an image, compared to central image areas; thus atmospheric effects on the signal will be stronger. Atmospheric path lengths for polar orbiters also change at each visit to the same location, because the satellite zenithal angle differs at each revisit. The exact local time of recording signal of the same location also varies; the resulting variability in observation conditions can cause inconsistencies of more than 10^oC. The constant changes in polar Field Of View (FOV) and times of imaging pose serious difficulties for radiometric and spatial preprocessing. Tramutolli (2001) mentions potential variations of 5^oK in NOAA-AVHRR recorded TIR signal due only to the afforementioned effects; based on the ranges reported as anomalous in literature, this could easily mask out or highlight an earthquake-related anomaly. Aliano et al. (2008) assessed the performance of a thermal anomaly index (named "RETIRA") based on different data sources and noted the prevalence of geostationary TIR sensors; they further proposed a future use of passive microwave sensors. Thus, it can be argued that the use of thermal sensors aboard geostationary satellites has several advantages: despite their lower spatial resolution, they maintain relatively constant position over the same location, allowing for a more precise colocation of the images and minimizing signal-to-noise ratio and signal variability due to observational conditions (EUMETSAT, 2012; NOAA, 2012; ISRO, 2012; Tramutoli et al., 2005; Aliano et al., 2008). Additionally, the frequent registration of images provides more detailed and well-structured time series datasets and can compensate for potential restriction of data availability due to cloud cover or technical issues.

Overall, available geostationary thermal sensors have proven to be very useful in monitoring the earth's thermal field because they provide high sampling frequencies, sensitivity in temperature measurements and improving spatial resolution. BT values are prefered in the context of the present study to avoid potential introduction of errors during atmospheric corrections, given that the main focus is oriented to relative thermal changes (considering past history and surroundings) rather than absolute temperature values. Expected effects that should be taken into consideration, are caused by the differences in terrain and topography, geological substrate, variations in water content, local prevailing condition variability (e.g. winds) and of course clouds, which cannot be penetrated by TIR signal and their diversity (in shape, height, constituents) may affect the signal in different ways.

1.3 TIME SERIES ANALYSIS AND DETECTION OF ANOMALY: THEORETICAL FRAMEWORK

Even if a considerable temperature rise takes place and is duly recorded prior to an earthquake, its value as a precursor practically vanishes if we cannot effectively distinguish it within the available data. There are many factors affecting temperature and its fluctuations:

- Movement of the earth in relation to the sun, ergo seasonality and daily fluctuations, in global scale;

-Climatic, atmospheric and physical factors, (e.g. suspended matter, water vapour, carbon cycle) can account for low-frequency fluctuations;

-Terrain/topography, geological substrates, presence of water and vegetation.

Regular, and recurring temperature patterns can be related to specific processes and determine an expected, "normal" temperature profile of an area. The application of pre-earthquake temperature increases as precursors lies in the assumption that the occurrence of earthquakes is not regular, contrary to the majority of the processes that affect temperature. Thus, the temperature increase caused by the earthquake should be discernible as a deviation from the expected, normal profile— it should be distinguished as an "anomaly". It is therefore of fundamental importance to define the characteristics of the "anomalous" temperature rise preceding an earthquake, and be able to extract it from the "normal", expected background. Inversely, a first step could be a sufficient description of the "normality" of the data.

Chandola, Banerjee and Kumar (2009) describe anomalies as "patterns in data that do not conform to a well defined notion of normal behavior". Thus, in the context of earthquake monitoring, the definition of normality is based on relating observed changes with the underlying procedure(s), considering that temperature can be affected by more or less persistent processes. According to Anscombe and Guttman (1960, in Chandola et al.), "an anomaly is an observation which is suspected of being partially or wholly irrelevant because it is not generated by the stochastic model assumed", or, as phrased by Hawkins (1980), "it deviates from others as to arouse suspicions of being generated by a different mechanism". So the key issue is to define the expected temperature changes caused by recognized factors, exclude fluctuations that can be linked to factors not related to tectonic processes, and examine the remaining as "anomalous" instances.

Different types of anomalies have been described in literature (see Table 1.1). In time series, we can further distinguish single anomalous events, anomalous subsequences and whole time series anomalous compared to other series. Consequently, different approaches have been developed to identify them (Table 1.1). Anomaly detection is the procedure of tracing parts of the data which deviate from the expected behaviour. Defining the expected, normal behaviour is not a straightforward task though. Firstly, it may depend on the context. The boundaries between normal and abnormal behaviour are not always clear. Normal patterns are not always static and may evolve, imposing a need for constant adjustment of the notion of normality. Finally, the processes that affect the behaviour, and thus its abnormality, may be interrelated. Taking into account the vast variety of application domains, the existing anomaly detection techniques are very diverse and rather address "specific formulations of a problem" (after (Chandola et al., 2009)). The output of anomaly detection can be either a label or a score assigned to the value. The label is a fixed threshold or description, based on which an instance is declared anomalous or not. The score is an indicator of the degree of abnormality, allowing flexibility in the decision of 'how much anomalous' does the value have to be in order for the user to bother with it.

Table 1.1: Categorization of anomalies and main anomaly detection techniques, after (Chandola et al., 2009; Cheboli, 2010)

Type of Anomaly	*Description*	*Anomaly detection technique*	*Description*
Point Anomalies	Individual data instances considered as anomalous with respect to the rest of data	Classification techniques	distinguish normal/anomalous classes
Contextual Anomalies	Instances considered anomalous in a specific context (but not otherwise) Contexts: Sequential(11^oC on a July afternoon in Greece), Spatial(pixel very different to its neighbours)	Proximity-based techniques (Nearest-Neighbour analysis, Clustering)	Normal data instances lie close to each other, anomalies are at a larger distance from their closest neighbour. These techniques identify the existence of an anomaly but do not locate it exactly in the time series
Collective Anomalies	Related data instances, anomalous with respect to the entire data set (e.g., same temperatures throughout day and night)	Information-theoretic techniques	Anomaly=irregularity in the information content of the dataset
		Spectral techniques	Combinations of data attributes describe the variability in a lower dimensional subspace which highlights anomalies
		Statistical techniques	Normal data fall in high probability regions of a stochastic model; anomalies are deviations from the model which captures normal behaviour

Anomalous sequences in time series can be further identified by the application of **window-based techniques** (extracting fixed-size subsequences to locate anomalous patterns), **hidden—Markov sequences** (assuming a hidden markovian process generating the normal data), **segmentation** (which investigates if all the transitions from one subsequence to the other follow the same patterns) and **hybrid techniques** combining parts of all the aforementioned. Anomalous sequences can be traced either in a single time series or by comparing a test time series to another, which is considered as "normal" (training data). Before applying any technique, transformations of the time series can take place, if it is deemed that they will improve the detection performance. Such transformations include *aggregation* (usually replacing sets of consecutive values with their mean), *discretization* (treating small subsequences as discrete values) or *signal processing*, namely Fourier and Wavelet transforms (to reduce the dimensional representation of the series to a different space of coefficients, where the potential anomalies would be more obvious) (after (Cheboli, 2010)).

Taking into consideration the potential loss of information resulting from transformation, the lack of sufficient knowledge of the parameters affecting the dataset and the subsequent limitations in controlling the transformation results and their causes, no transformation was attempted for the scope of the present study. Furthermore, an examination of the available temperature datasets can reveal all kinds of anomalies. However, while point anomalies and collective anomalies would be related rather to sensor failure or technical problems, the focus of detection lies to contextual anomalies: subsequent, unusually increased temperature values, temporally and spatially related to earthquake epicenters. Given the characteristics of the available techniques and the dataset (there is no indication of the existence of a hidden—Markov process; proximity-based techniques do not provide the location of the anomaly, which is vital for the connection with earthquakes; nothing is known about the transition between discrete parts of the time series; and some of the techniques, like information-based, are irrelevant to the specific nature of the study), the main interest was focused on statistical techniques. A further reason is that, as described above, the majority of the detection techniques can be used in a straight-forward manner in the case of point anomalies, but their application in contextual anomalies would imply the loss of valuable information or impose undesired limitations in context. The use of statistical techniques is an exception. These techniques, along with providing statistically justifiable results, can be used in combination either with fixed anomaly thresholds or with scores associated with confidence intervals, offering a range of flexibility when deciding on the abnormality of a data instance. They allow for the development of unsupervised settings, depending on the design. Their main drawback is the limitation of the assumption that the data follow a particular distribution (after Chandola et al. (2009)).

Statistically based techniques can be non-parametric, when they define the model from the data itself with the condition that at least part of the dataset is considered normal a priori. On the contrary, parametric techniques assume a parametric distribution and probability density function to characterize normal data; if an instance is not generated by the assumed distribution, it is considered anomalous (Chandola et al., 2009; Cheboli, 2010; Chawla and Sun, 2006). This is consistent to the definition of anomaly which is linking it to a different underlying process. Different types of distributions can be used for modeling. In the case of the Gaussian distribution, thresholds are applied to control the number of instances declared as anomalous; they can be equal to a number of standard deviations or a value taken by a t-distribution at a significance level based on the specific case. On the other hand, regression has been extensively investigated and proven very appropriate for anomaly detection in time-series data (Abraham and Chuang, 1989; Fox, 1972). Regression-based techniques fit a model on the data; the residuals represent the part of the instance value which is not explained by the model, and can be used as anomaly scores.

Many such modeling approaches exist, as robust regression (Rousseeuw and Leroy, 1987), autoregressive models (Fox, 1972), ARMA models (Abraham and Chuang, 1989), and ARIMA models (Tsay, 2000). These fall into the "prediction-based" category.

Another regression-based approach is the Seasonal-Trend Decomposition for time series (STL) based on LOcally weighted polynomial regrESSion (LOESS) and developed by Cleveland (1990). Although it can be applied also for prediction, its main aim is to detect patterns underlying the data, by distinguishing a trend and a seasonal component from the remainder. Low frequency fluctuations are intended to go to the trend, and higher frequency fluctuations to the seasonal component. The underlying assumption is that the majority of the data points are caused by the effect of the same factor, and their behaviour can be captured by describing this effect; by definition, outliers are values caused by a mechanism different than the one producing the majority of the data. The rationale of the procedure fits the scope of the present study and its characteristics further support its suitability: it provides flexibility in the tuning of the parameters which affect the modeling of normal behaviour, according to the intended application, and it is robust to outliers, retaining aberrant values in the remainder of the decomposition for further study Cleveland (1990). Unfortunately, the method was initially implemented in Fortran and its R version was not applicable until recently due to a constraint of missing values. However, recent work by P.Hafen (2010) implemented STL decomposition in R, plus adding improved functionalities. Although not yet publicly available through CRAN (the comprehensive R archive network, where normally R-related packages and documentation are available), the required packages were obtained courtesy of the author. The technique was applied in an attempt to define, and remove from the time series data, temperature fluctuations resulting from seasonality and inter—annual trends, with the assumption that the potential anomalies would be captured by the remainder and would be more easily detected there.

1.4 EXISTING RESEARCH AND RECENT ADVANCES

Research up to date is characterized by a great diversity in terms of choice of sensor, earthquake and methodological approach. For instance, Pulinets et al. (2006) use ground temeperature measurements, Yao and Qiang (2012); Genzano et al. (2007) use data from geostationary satellites, Ouzounov and Freund (2004); Saraf et al. (2012) from polar-orbiting sensors, Tramutoli et al. (2005); Aliano et al. (2008) have used both. Anomaly definition is based on change detection and anomaly indexes (RST, RETIRA, ALICE) as proposed by Tramutoli et al. (2001) onwards; or on the differences between maximum and minimum daily temperatures (Pulinets et al., 2006); or on user-defined normal temperature ranges (Saraf et al., 2008); or on the difference between the dataset and data from previous years (Ouzounov and Freund, 2004; Blackett, 2011). An indicative part of the existing literature is presented in Table 1.2, in order to underline implications of previous research in the design of the present study. Concerning the characteristics of thermal anomalies, researchers agree that they are developing fast, are related to earthquake magnitude and can be found in the proximity of faults (Yao and Qiang, 2012; Ouzounov and Freund, 2004; Saraf et al., 2008).

On one hand, the diversity in approaches can be used as evidence of the variety of thermal pre-earthquake phenomena; on the other hand, it makes it difficult to cross-check results and systematize the detection approaches. Each of the methodological suggestions is faced with specific limitations which, although not necessarily defying the validity of the research outcome, they do provide interesting insight on the development of more efficient designs in future work (see Table 1.2. Overall, existing literature underlines:

(a) the importance of the length of the time series studied for anomaly detection. A value

considered abnormal in a short dataset might be deemed normal if more observations are taken into account (Blackett, 2011).
(b) the use of appropriate thermal sensors. This might have implications in signal variation, as discussed in previous section 1.2, but also affects temporal resolution. When the sampling intervals are long (i.e. in the case of polar-orbiting sensors) the presence of clouds or a potential system malfunction would limit the data availability considerably and a 2—3 day lasting anomaly might be lost from the dataset.
(c) the necessity of a sufficient definition of a thermal anomaly. Most researchers follow different approaches and it is not easy to have comparable results. Additionally, in most of the cases meteorological effects and the spatial dimension are not systematically taken into consideration.

Table 1.2: Indicative summary of existing research

Researchers	*Research summary*	*Issues and concerns highlighted by the study*
Pulinets et al. (2006)	Air temperature and relative humidity from ground meteorogical stations. NOAA-derived observations. Earthquakes (M>6) in USA and Mexico anomaly definition: temporal variations in the difference between daily maximum and minimum. Apparent anomalies were observed in stations up to 60km from epicenter	
Yao and Qiang (2012)	Geostationary sensors earthquakes in Asia (>7)	(a) spatial comparison between TIR values of neighbouring areas and (b) location and development of anomaly (duration, movement, appearence and reappearence) no reference to the specific anomaly definition they applied
Saraf et al. (2008)	NOAA-AVHRR datasets Exclusion of clouded areas LST fluctuations in ten earthquakes (M>5.8) in Iran Anomaly definition: normal "user-specified temperature range" defined for each earthquake case Relation of anomalies and earth deformation	Shallower earthquakes of larger magnitudes produce more intense and spatially extended anomalies Potential release of energy in other forms may lessen the magnitude of upcoming earthquake No corrections or image co-location
Ouzounov and Freund (2004)	MODIS LST data Split-window algorithm corrections Anomaly definition: the residuals of differencing the earthquake-year and the previous-year temperatures	
Tramutolli (1998 and onwards), Aliano et al. (2008), Genzano et al. (2007)	Robust Satellite Technique (RST):estimation of anomaly scores based on change-detection Anomaly index: subtract the mean of a reference historical dataset from the difference of an observation with its spatial average, and divide by the standard deviation of the reference dataset Anomaly threshold :3.5σ of the enumerator of the index to flag anomalies	Statistically founded anomaly definition, exportable to different data sources Anomalous areas are sometimes found at distances not justifying relation to earthquake Index vulnerable to outliers. Choice of the size of the image affects scene variablity and anomaly detection
Blackett (2011)	Anomaly detection : (a)RST (b) differencing between earthquake year and one reference year (c) difference of the earthquake year from the mean LST of six previous years	When moving from a two-year to a six-year dataset for the same earthquake, LST values previously deemed abnormal are reconsidered as normal

A recent development in thermal anomaly detection is based on a methodological approach put forward by Khan (2010) and further developed by Buzzo (2012). The so-called normalization process takes place by dividing the BT of a pixel with the mean value of a square ring of neighbouring pixels, in order to diminish meteorological effects (diurnal and seasonal patterns). The new time series produced in this way is the basis for a statistical detection of anomalies based on the mean and standard deviation of the time series. This was the first systematic way of handling meteorological effects, and the researchers developed it further by introducing cloud masks and attempting a visualization to spatially and temporally link the occurrence of anomalies with earthquakes.

As can be derived from the issues discussed above, the challenges for the detection of anomalous thermal instances prior to earthquakes are multiple and diverse. To summarize them,

(a) a number of different potential physical processes affect the occurence of thermal increases prior to earthquakes,

(b) temperature fluctuations themselves are determined by numerous factors,

(c) the recording of these fluctuations by remote sensors can introduce extra variability; the choice of sensor and co-examination of other factors influencing the signal is important,

(d) the applied statistical approach for anomaly detection has to take into consideration, as much as possible, the peculiarities of the situation.

The objectives of the present study and the methodological approach followed, were determined on the basis of these issues and on the outcome of existing research.

1.5 OBJECTIVES OF PRESENT STUDY

The main objective of the proposed study was to apply time series analysis to hypertemporal thermal infrared images, in order to detect unusual temperature rises occurring prior to strong, shallow, landbased earthquakes. These limitations were applied following the results presented so far in literature. Longer datasets were used to allow for a more sturdy distinction of normal and unusual behaviour. Furthermore, an attempt was made to interprete the occurence of thermal anomalies not related to earthquakes. The analyses of the time series took into consideration:

1. the lack of a specific profile of expected anomaly
2. the need to retain the spatial and temporal characteristics of the anomaly, in order to assess relation to earthquakes at a later stage
3. the potential shown by the comparison of pixel values to their surroundings, as applied by Khan (2010) and Buzzo (2012), and the potential of the application of STL decomposition for distinguishing normal and abnormal fluctuations of the temperature values
4. the characteristics of the available sensors and TIR images
5. the prevailing conditions in the earthquake areas, to the extent they could be accounted for.

The specific objectives of the present study are as follows:

1. Improvement and standardization of detection scheme

   Working hypothesis:

   An optimal approach can highlight thermal peaks standing out from the pattern of normal temperature behaviour. Unusual temperature rises would be notably present in earthquake-affected pixels, but not in farther areas; they should appear only in the time before the

earthquake and not constitute a recurring event; their characteristics should be related to underlying mechanisms of tectonic activity.

Sub-objectives:

(i) Optimization of existing mehodological scheme. This refers to an assessment of the performance of the existing scheme (based on the normalization of pixel values by comparing them with their surroundings) and the application of potential improvements.

(ii) Improvement of cloud-masking procedures, to facilitate detection and eliminate false positives and/or loss of information.

(iii) Investigating different approaches to de—trending and de—seasonalizing extended time series of the available data. This step includes handling of missing data, and removal of the seasonal and daily patterns in a way potential extreme values remain unaffected and traceable.

(iv) Anomaly identification and flagging, preferably bypassing the use of standard deviation if possible.

(v) Improvements in image geo-referencing and use of DEM and other available information to reveal other factors affecting temperature fluctuations, in order to explain false positives and further exclude effect of non-earthquake related factors.

2. Application of methodology on different earthquakes

Working hypothesis:

Thermal peaks should be traced prior to earthquakes resulting from different processes and occurring in regions with different prevailing conditions.

Sub-objective: Application of proposed methodology on earthquakes of different magnitudes, resulting from different tectonic processes, occurring in different fault types, and in areas with different climatic conditions.

Overall, The present study investigates the potential of satellite monitoring of thermal increases as an earthquake precursor. Furthermore, it attempts the development of a methodological approach of possible applicability for similar research on other phenomena which are also based on time series and anomaly detection.

Chapter 2
Methodology

The methodological steps followed in the present study are summarized in Figure 2.1. The study was carried out in four stages:
(a)examination, assessment and review of the existing methodology
(b)investigation in the potential of the application of STL decomposition
(c)evaluation of different settings for incorporating STL in the existing scheme
(d)application of the new, standardized scheme to different earthquakes.
The evaluation of methodological approaches was based on testing synthetic anomalies. All statistical processing was carried out in R (R Core Team, 2012).

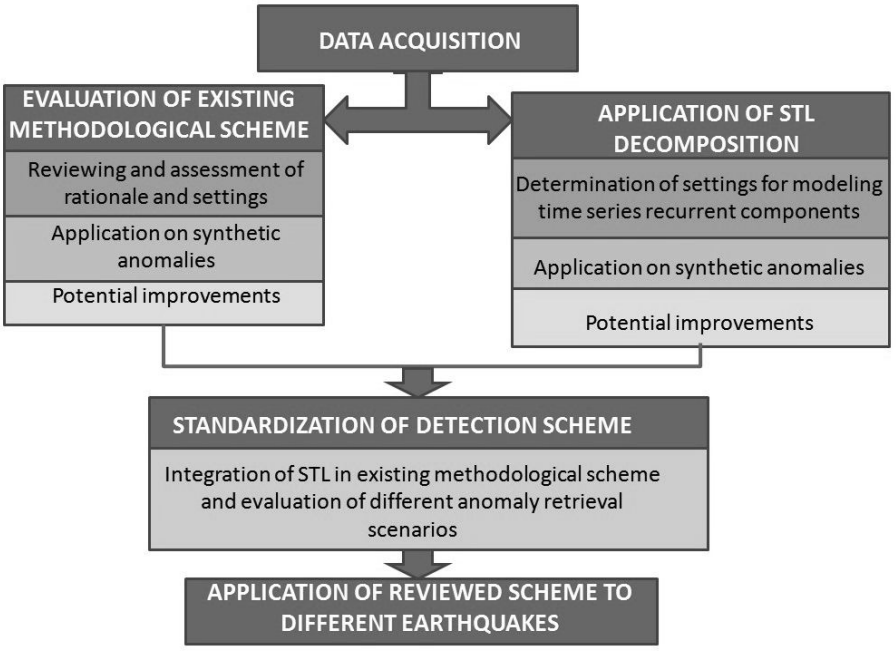


Figure 2.1: Flowchart

2.1 MATERIALS AND CHOICE OF DATASETS

The investigation was carried out on thermal data from different earthquake-struck areas. This was important in order to test the final detection scheme on earthquakes with different characteristics and describe potential required adjustments.

Based on the existing literature, the earthquakes considered had to be landbased, of a magnitude larger than 6.0, shallow (focal depth<20km) and preferably located in areas with scarce

vegetation, so as to ensure minimum interference in the TIR signal. Research for choosing suitable earthquakes was carried out via the USGS-NEIC database, which was the primary source for all relevant earthquake-related information (technical sheets, maps, tectonic information etc). Earthquakes were selected based on the following criteria:

- type of the tectonic process causing the earthquake
- magnitude and focal depth
- fault activity and frequency of earthquake occurrence in the area, also of smaller magnitude
- season that the earthquake occurred, in the sense that the proposed methodology should be able to detect thermal anomalies regardless of season characteristics
- sensor recording the TIR signal of the area in the period of the earthquake.

2.1.1 Study areas

A short description of the chosen earthquakes and the corresponding areas is given below. More information about each study area can be found in Appendix B.

1. **Bam, December 2003.** The earthquake of Bam took place on December 26th, 2003 at a focal depth of 10km, had a magnitude of 6.6 and was considered as the largest earthquake of the area in more than 2000 years. It caused the death of least 30,000 people and a 85% damage on buildings and infrastructure in the Bam area. According to the USGS, the earthquake was caused by right-lateral strikeslip motion on a north-south oriented fault, occurring as the result of stresses generated by the northward movement of the Arabian plate against the Eurasian plate at a rate of approximately 3 cm/yr. Surface faulting was observed on the Bam Fault between Bam and Baravat, and the maximum acceleration recorded at Bam was 0.98g (after USGS (2012).

   The Bam area is part of the Lut-e-Zangi Ahmad desert that has hot summers with temperatures up to 50^oC and winters with below freezing temperatures. The geomorphology of the region also includes a range of mountains to the N-NW and another one in the south, extending NW-SE. There exist seasonal rivers but the precipitation is not enough to sustain their flow throughout the year. According to the Geological Survey of Iran (GSI) the area around Bam is covered by coarse brown sandstone deposits; tuff and traciandesite rocks are found to the North while other parts are built on alluvial deposits. The Bam fault is the main tectonic feature in the area, separating older fans from younger sediments(Figure 3.13).

2. **Van, Turkey 2011.** The October 23, 2011 earthquake of Van had a magnitude of 7.1 and focal depth of 16kms. As a result of the earthquake, more than 600 people were killed and over 4000 were injured, while the damages in more than 11000 buildings left over 60000 people homeless (after USGS (2012)).

   As mentioned in the USGS report, the Van earthquake occurred in a broad region of convergence beyond the eastern extent of Anatolian strike-slip tectonics. The focal mechanism of the event is consistent with oblique-thrust faulting. The area is characterized by a tectonic complexity and a large number of small-scale faults with interlinked activity.Zahradnik and Sokos (2011) note that, “a possible spatio-temporal scenario of the earthquake consists of two or three major subevents, following each other towards south-west” while a diversity

of secondary events like aftershocks, landslides, rockfalls, liquefaction and lateral spreading were observed (Prime Ministry and Emergency Management Presidency, 2011).

The earthquake epicenter is located on the west shore of Lake Van, a tectonic depression situated at 1648 m above sea level in the eastern Taurus mountains in SE Turkey. The lake catchment consists of metamorphic rocks (South),conglomerates, carbonates and sandstones (East), and volcanic deposits (North and West) (Figure 3.15). Lake Van is a saline endorheic lake. The climate in the area is continental with long and cold winters and warm summers. Originally situated in the transitional zone between two vegetation types: an oak-forest belt (South-Southwest) and steppe to the (North-Northeast), the area has nowadays been stripped off the natural cover to a large extent and is used mostly for agriculture (after Wick (2003); Huguet (2012)).

2.1.2 Thermal imagery

For reasons explained in the Introduction 1.2, thermal images of interest for the present study came from geostationary satellites. The existing methodology had not yet been applied in second-generation sensors, so it was interesting to see if there are differences in the outcome of the detection depending on the sensor characteristics and thus any sensor-controlled limitations. The long-term aspiration is, that the detection methodology will be applicable to imagery coming from different kinds of geostationary sensors. This will support the its external validity but also allow for the study of earthquake regions which have been covered by older equipment.

The thermal images were acquired through the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT), which is operating the Meteosat series of geostationary spacecraft. The distinction between the first and second generation (MFG and MSG) designates the upgraded functionalities and services provided by the new series of satellites, which have enhanced sensors aboard. The Data Collection System (DCS) coverage is supported by two satellites: the main one is located in the operational position at 0^o longitude and the other is stand-by, a few degrees to the West. There is a Rapid Scanning Service (RSS) at 9.5^oE longitude, and a satellite placed at 57.5^oE, over the Indian Ocean (Indian Orbit Data Coverage, IODC)(EUMETSAT, 2012).

MFG satellites were equipped with the MVIRI sensor having one available IR band, (10.5-12.5) with a spatial resolution of 5km and providing observations every half hour. The MSG sensor, SEVIRI, has six TIR bands (channels 4, 7-11), a spatial resolution of 3km, and provides 4 images hourly. The area of Bam was studied with data derived from Meteosat-5 in IODC, in the years 1999-2004. The areaq covered had coordinates between 27-31^oN latitude, and 56.325-60.325^oE longitude. Van was covered by MSG-2 (Meteosat-9) in operational position for the years 2008-2012 (images provided by the RSG laboratory of ITC, via the MSG-data retriever). Area coverage was between 37-40^oN, 42-45^oE. In the case of Van, Channel 9 was chosen (IR centered at 10.8, ranging from 9.8-11.8 um). This is one of the so-called split-window IR channels (the other is channel 10, centered at 12.1) which can be used for reducing atmospheric effects and for cloud detection. The Meteosat-5 IR window used in the case of Bam has a range between 10.5 and 12.5um. The chosen image size in the case of Bam was around 400X360km and for Van 300X240km.

2.2 STRUCTURE OF THE RESEARCH AND EXPERIMENTAL SETTINGS

The following section gives a brief description of the methodological issues that were handled throughout the research.

2.2.1 Description of existing methodology

The existing methodology (Khan, 2010; Buzzo, 2012) is coded in an IDL based algorithm named SHAKY, which consists of sub-routines (programs) with different functionality. These programs are described below in order of execution, and the main steps are displayed in Figure 2.2.

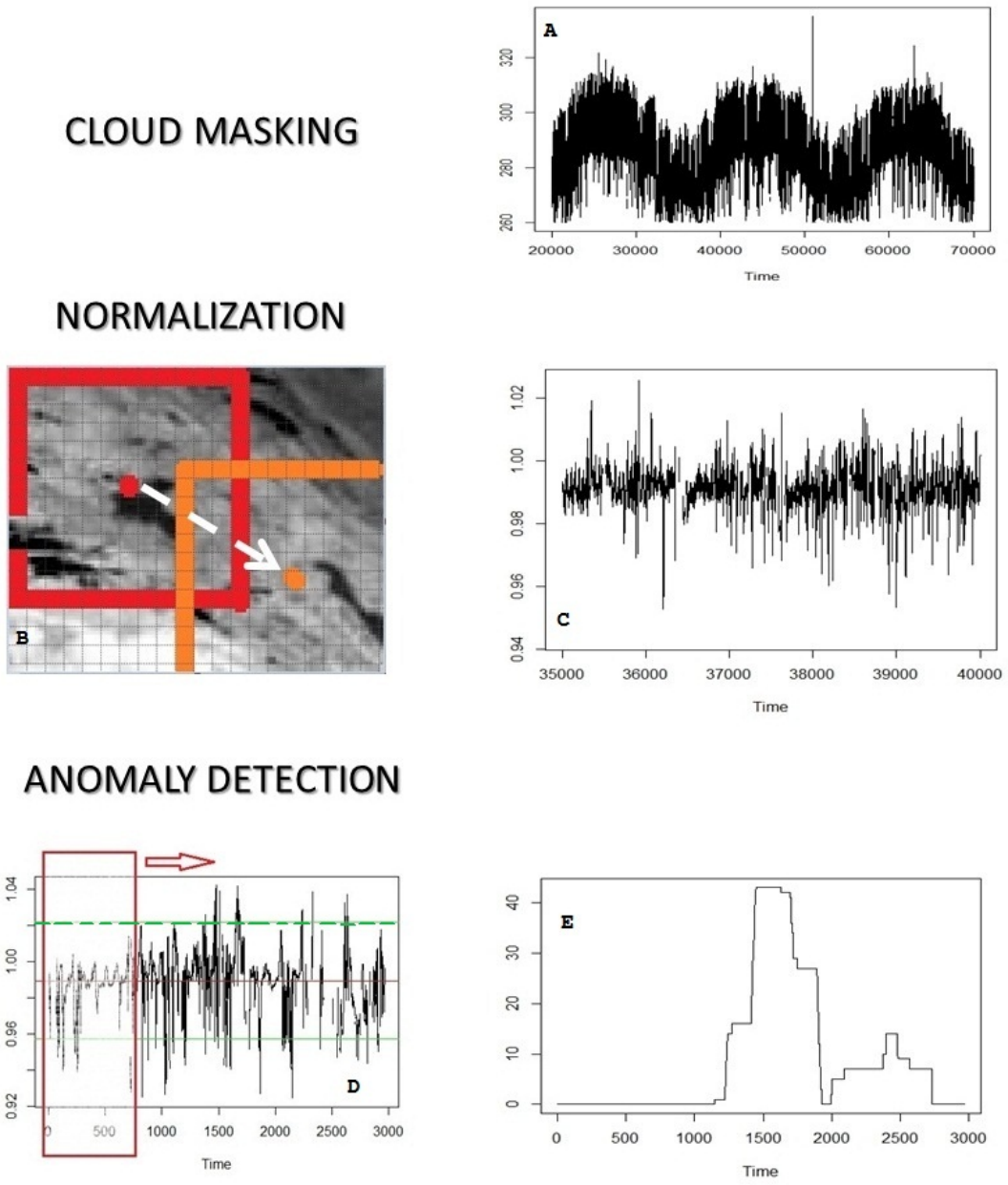


Figure 2.2: Existing detection scheme: main procedures (left) and outputs (right). A:Time series of one pixel after cloud removal. B: the available values of a square ring of pixels are summed and their mean is used to normalize the value of the central pixel. The procedure is repeated for every image and every pixel. C: normalized time series of the pixel in diagramm (A). D:temporal moving window which moves along the normalized time series and counts the values which are higher than the threshold(green dashed line). E: number of anomalous instances traced by the window (D). For example, timestamp 1500 corresponds to the number of anomalies found within the time period [1500-n, 1500+n], where window size=2n.

1. ***Image import***. Two programs exist, one for importing MFG images and one for importing MSG images. Aside from sensor-specific procedures (calibration and conversion) the programs read the image files, check for the existence of non-valid images in the dataset and stack the images sequentially in a single .bsq image file.

2. ***Cloud masking***. The algorithm implements a Relative and Absolute cloud Mask (RAM) consisting of two parts. The absolute mask discards all pixel values below a given cutoff threshold; this value is obtained from historical climatic archives of the study area. The relative mask calculates a second cut-off value as the 0.95% of the maximum observed temperature reduced by the natural temperature variation of the study area. Overall, the masks operate both on a band and a whole-dataset level, removing (a) values below an absolute historically-based minimum throughout the dataset, and (b) values that are considered low on the basis of the variation of each image.

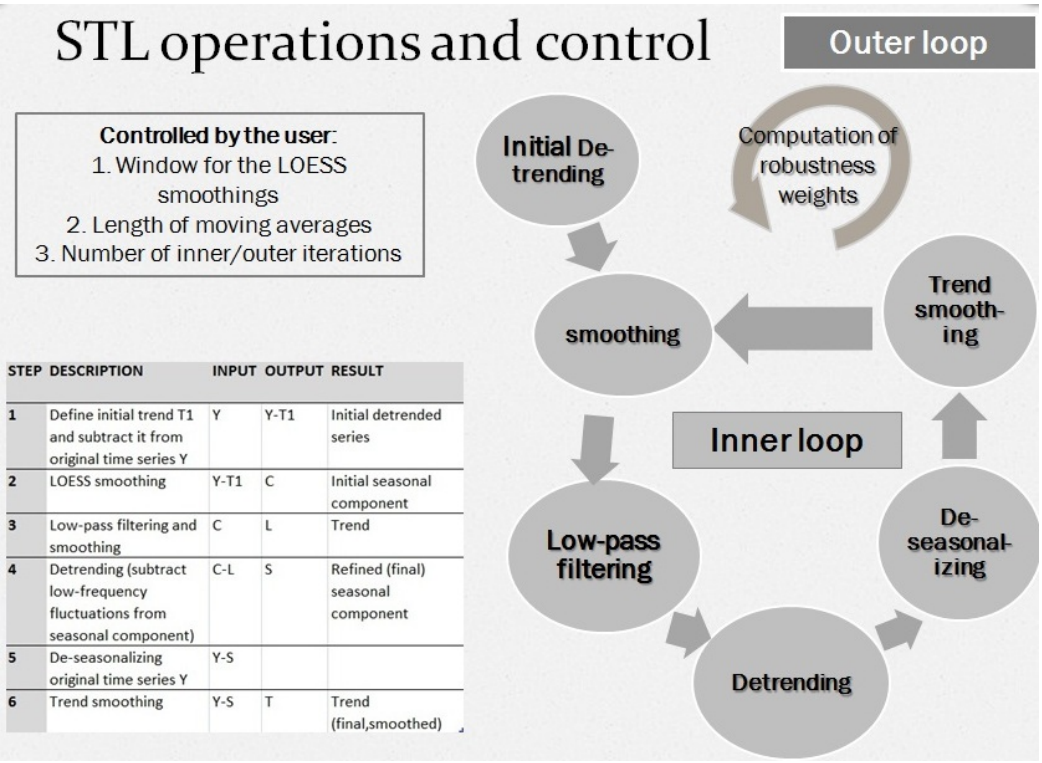
3. ***Normalization***. In order to neutralize climatic effects, each pixel value is divided by the average value of a square ring of pixels. The ring is located at a user-specified distance from the center-pixel. This distance has to be close enough to ensure that weather conditions are not completely different, but far enough so that a possible earthquake related anomaly is not included in the square ring. This procedure is repeated for all pixels and the result is stored in a new image, containing the normalized values of all the pixels. Effects of climatic conditions are suppressed. The normalized time series shows how different the pixel is compared to its neighbours at each moment in time.

4. ***Anomaly flagging***. The basis of anomaly detection is the time series of the normalized values. The first step in the procedure is the calculation of the mean and standard deviation of the normalized values of all observations in time for each pixel. A fixed threshold is defined as the mean of the series plus two standard deviations, and each value exceeding this threshold is flagged as anomalous. In a second step, a temporal window is specified by the user. The window moves through the observations (timestamps) of the time series. On each observation (timestamp), the number of anomalous incidences which fall within the time window is summed to give an anomaly count. This is stored at the timestamp of the center pixel of the window. The image produced in this way contains, in each location, the number of flagged instances that appear in the moving temporal window (2.2). Finally, a single-image output is created, containing for each location the maximum observed number of flagged values, throughout the whole dataset.

A review of the existing scheme was vital, to better understand what it is detecting and to test if it can be applied to other datasets. From a technical perspective, one goal was to ensure its functionality in larger datasets and make potential required adjustments. As mentioned in the Introduction, the length of the time series is crucial for anomaly detection. An examination of the rationale would bring up potential points for improvement; the previous researchers had already mentioned concerns about the normalization process (especially the choice of kernel size) or the effects of missing data. Furthermore, the existing methodology was tested for the first time with synthetic anomalies. This was very important: it was an effective way to establish the sensitivity of the method, assess what it detects and objectively evaluate its performance. The above were examined in both earthquake cases (Bam and Van) to make sure that any conclusions drawn were not case-specific.

2.2.2 STL decomposition

STL decomposition was examined as a potentially suitable method to model and remove signal fluctuations which happen due to the daily, seasonal and inter-annual behaviour of the data (Rossiter, 2012). It is structured as a sequence of two distinctive operation loops, based on weighed local regression smoothing. The inner loop consists of the following steps (after Cleveland (1990)):



STEP	DESCRIPTION	INPUT	OUTPUT	RESULT
1	Define initial trend T1 and subtract it from original time series Y	Y	Y-T1	Initial detrended series
2	LOESS smoothing	Y-T1	C	Initial seasonal component
3	Low-pass filtering and smoothing	C	L	Trend
4	Detrending (subtract low-frequency fluctuations from seasonal component)	C-L	S	Refined (final) seasonal component
5	De-seasonalizing original time series Y	Y-S		
6	Trend smoothing	Y-S	T	Trend (final,smoothed)

Figure 2.3: STL decomposition: description of process and settings (after Cleveland (1990))

STEP 1: Initial detrending(Figure 2.3). An initial trend is calculated and subtracted from the raw time series.

STEP 2: LOESS smoothing of the detrended series(Figure 2.3). LOESS-smoothed values are calculated for all time positions of a cycle subseries. The term cycle subseries refers to a time series of all values of the same time stamp in the dataset, for example, a series of all the Januaries. The January smoothed value of a year is computed on the basis of January values of a number of available years. This number of values used for the smoothing is defined by the user. The values which are closer in time to the one which is being smoothed, are given a higher "neighbourhood weight". This means that they are taken more into account for the smoothing.

STEP 3: Low-pass filtering(Figure 2.3). The series is filtered subsequently by three moving averages and then smoothed again by LOESS. The goal is to attenuate short-term fluctuations and highlight longer-term fluctuation, (definition of trend). The length of the moving averages and the smoothing range is determined by the user.

STEP 4: Detrending(Figure 2.3). The product of step (3) is subtracted from the smoothed series of step (2) refining the seasonal component.

STEP 5: Deseasonalizing(Figure 2.3). The seasonal component is subtracted from the original time series.

In the outer loop which follows, each value of the remainder R (where R=Y-T-S, see Figure 2.3) is given a weight calculated on the basis of the median of the remainder. This weight indicates

the level of "extremeness" of the value: outliers, or extreme values, have very small weights. This will be multiplied by the neighbourhood-weights used in the LOESS smoothing for the subsequent iteration of the inner loop(Figure 2.3).

The method allows for a choice of the settings of the decomposition by the user, based on the objective of the decomposition and the knowledge of the processes affecting the data(Figure 2.3). This gives stl an intrinsic flexibility. The authors of the method suggest diagnostic tests and criteria to make justifiable choices according to the scope of the study. More specifically, the user can determine (after Cleveland (1990); P.Hafen (2010)):

-The *number of iterations* of the inner and outer loop, affecting robustness of the model to outliers

-The *frequency* of the series, which affects the cycle-subseries. For example, in the Bam dataset, there were 48 observations daily. For a smoothing on a daily basis, the method will make a subseries of values which are 48 timestamps away from each other.

-The *parameters* affecting the smoothing and filtering procedures, and defining which part of the variability goes to the trend and which to the seasonal component. This includes the choice of the number of subseries values used for each smoothing (see also STEP 3 described above).

The resulting components are not necessarily smooth between subsequent observations. That is because the smoothing takes places between points corresponding in time (e.g. between Januaries) but not subsequent (e.g. between January and February). Therefore, in case the study requires the extraction of smoother components, there is an option for post-smoothing. The final product of the decomposition, the remainder, is derived by subtracting from the raw time series the daily/seasonal/inter-annual patterns. The remainder encapsulates all the deviations of the pixel from the "expected" behaviour, as this is modeled by the other components; i.e. it consists of all the fluctuations caused by factors which do not have daily, seasonal or inter-annual frequency (Figure 2.3).

As can be deduced by the flexibility in the application of STL, the first step in investigating the potential of the decomposition was to define the settings which more effectively describe the recurrent fluctuations of the data. The trials designed to serve this purpose were again based on the use of synthetic anomalies. In this way, the performance of the model would be evaluated not only in terms of delineating clear daily/seasonal/inter-annual patterns but also in terms of retaining the imposed anomaly intact (and available for being detected) in the remainder of the decomposition.

An examination followed on the possibilities of comparing and combining the reviewed existing detection scheme and the refined STL decomposition into an improved scheme for anomaly retrieval. Finally, the new standardized scheme was applied in the different earthquake datasets.

An overview of the structure of the research is given in Table2.1, along with the experimental settings which were examined in each step.

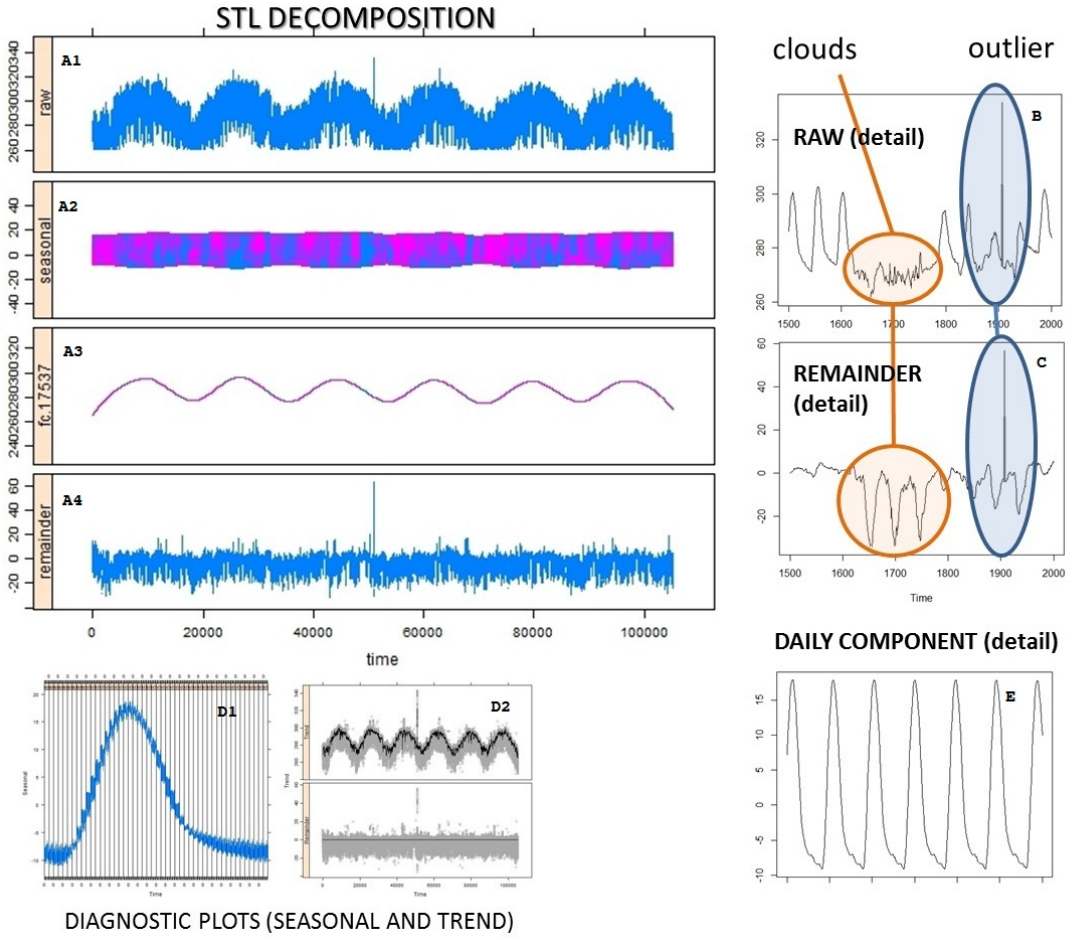


Figure 2.4: STL decomposition: components and diagnostics. A1: raw time series, after cloud removal. A2: Seasonal component. With the present choice of settings, its expresses diurnal fluctuations (see detail E). It becomes wider in the summer, which is consistent with ground measurements of temperatures (see Appendix B). A3: Smooth trend, expressing climatic seasonality. A4: Remainder, containing the result of subtracting the seasonal component and the trend from the raw time series. B: detail of the raw time series, and in C: the corresponding remainder. Clouds and outliers are highlighted. D1: Seasonal diagnostic plot and D2: Trend diagnostic plot, used to evaluate the amount of variability attributed to each of the components of the decomposition.

Table 2.1: Experimental settings

Objective	*Rationale/expected outcome*	*Experimental setting(s)*
Review of the existing algorithm	-functionality with bigger datasets/user interfaces -compatibility of formats for further processing	Do required tests and make adjustments if needed
Examination of the normalization kernel	**check:** -effect of data availability	-compare results of normalizing different sizes of images from the same area
	-choice of kernel size	-sensitivity analysis in different earthquakes
Synthetic anomaly retrieval	Assessment of the performance of the existing scheme	Retrieval of different anomalies using existing scheme
Modeling and tuning of decomposition parameters	Define the sequence and possible iterations of decompositions for the required extraction of daily, seasonal and inter-annual patterns	-yearly and then daily STL -daily and then yearly STL -application of trend post-smoothing in daily STL
Application on different artificial anomalies	Control for the quality of the model, i.e. see if the imposed anomaly is present after decomposition	**Testing anomalies:** -in different seasons -with different data availability
Examination of time series from different locations/terrain	Trace possible differences depending on terrain Get an idea on factors which might influence detection	STL on different pixels and comparisons
Combination of STL with the existing detection scheme	-removal of daily/seasonal patterns -possibility of bypassing σ for anomaly flagging	*Decomposition of normalized pixels* *Decomposition prior to normalization (retrieval of synthetic anomaly with comparison to "mean" reference):* -retrieval in the mean remainder of the neighbours -retrieval in the remainder of the mean neighbour signal
Cloud removal	Noise reduction Removal of false positives Highlight earthquake-related instances	*masking out:* -a portion of the remainder or -a part of the raw TS based on remainder values *examine:* -effects on the model/ -effects on detection

Chapter 3
Results

3.1 IMPROVEMENT OF EXISTING DETECTION SCHEME

3.1.1 Scheme review and technical considerations

The existing algorithms were applied to the available data from Bam and Van to control for its applicability to large datasets (103257 images for a 6-year time series in Bam and 138000 for a 4-year series in Van). The size of the datasets is important for the computational demands of the study, both in terms of time and resources. One single file containing all images has a size exceeding 4GB and each step of processing or subsetting it takes up several hours. Examination of the algorithms and the corresponding products revealed technical problems such as interface compatibility, which were resolved along with defects of the datasets. Observations about the products and corresponding improvements are described below. All new scripts are included in Appendix A.

a. Import According to the default settings of the importing procedure, the existing images were placed sequentially. A new importing program was designed and implemented, to make sure that the resulting stacks also contained blank images in the place of missing observations. In this way the products were in a format compatible for further processing, while at the same time the performance of the filtering procedure in later stages was improved. This happened was because filtering of the complete stacks took into account the real time interval between existing observations, resulting in a decrease of false positives (Figure 3.1).

b. Cloud masking Visual inspection of the products showed that clouds were still present in a number of images even after the application of the RAM mask (Figure 3.1).

c. Normalization The normalization products showed that the procedure did not remove completely all daily and seasonal patterns, as was initially expected (Figure 3.1).

Another issue was the choice of kernel size, which is the (spatial) distance between the pixel and the neighbours used for the normalization. The kernel has to be wide enough so as not to include pixels also affected by the earthquake, because in this case the potential anomalies would be present in both the kernel and the central pixel and the anomaly would not be visible after normalization. On the other hand, the kernel size cannot be increased infinitely because this would result in the pixel being normalized by neighbours with totally different prevailing weather conditions. The kernel is a square-shaped ring; the choice of the size of its sides was based on the literature regarding the expected extent of an earthquake related thermal anomaly and a conventional extent of homogeneity in local climate. Sensitivity analysis was performed for both Bam and Van with kernel sizes corresponding to sides of 60, 120 and 240 km. As seen in Figure 3.4, In both cases, the narrowest kernel size attenuated the observed peaks while the widest kernel choice highlighted smaller peaks and potential “noise”. The selected kernel size was thus set to 120km.

A final concern about the normalization is the variation in data availability. In the edges of the image only part of the kernel ring is available for normalization; the same can happen anywhere in the images because of cloud removal. This raises concerns of statistical consistency,

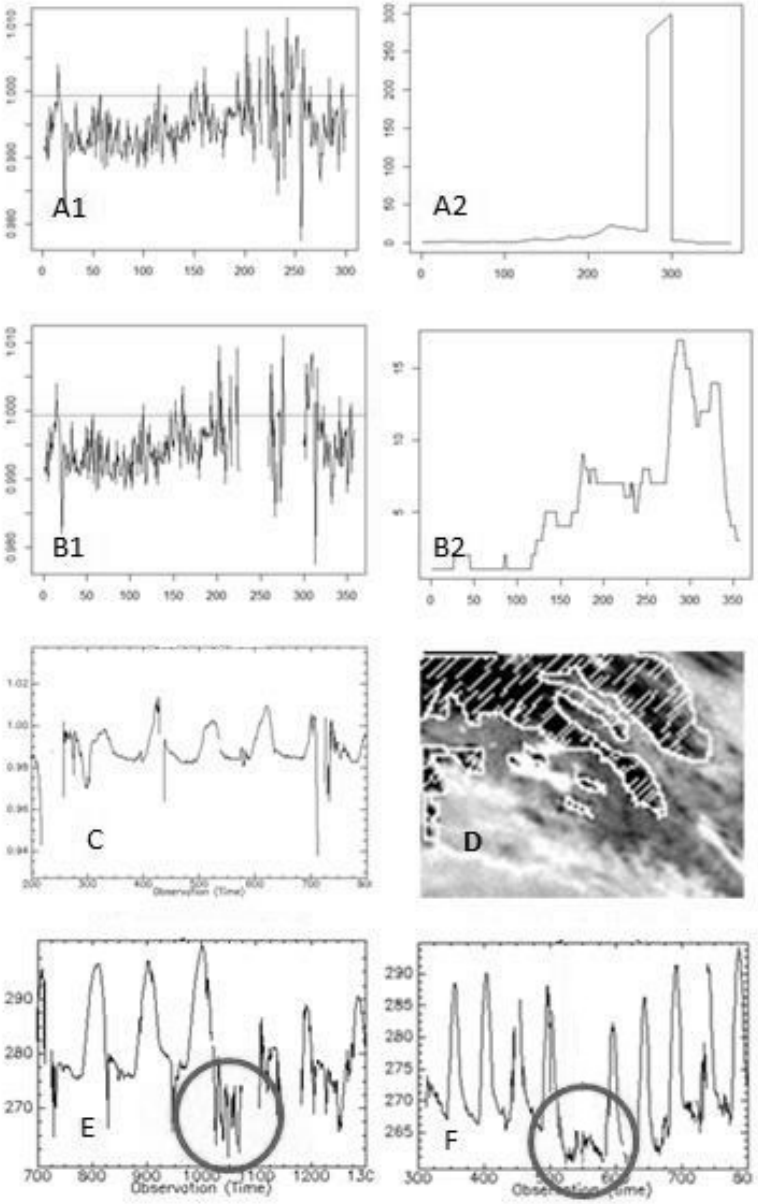
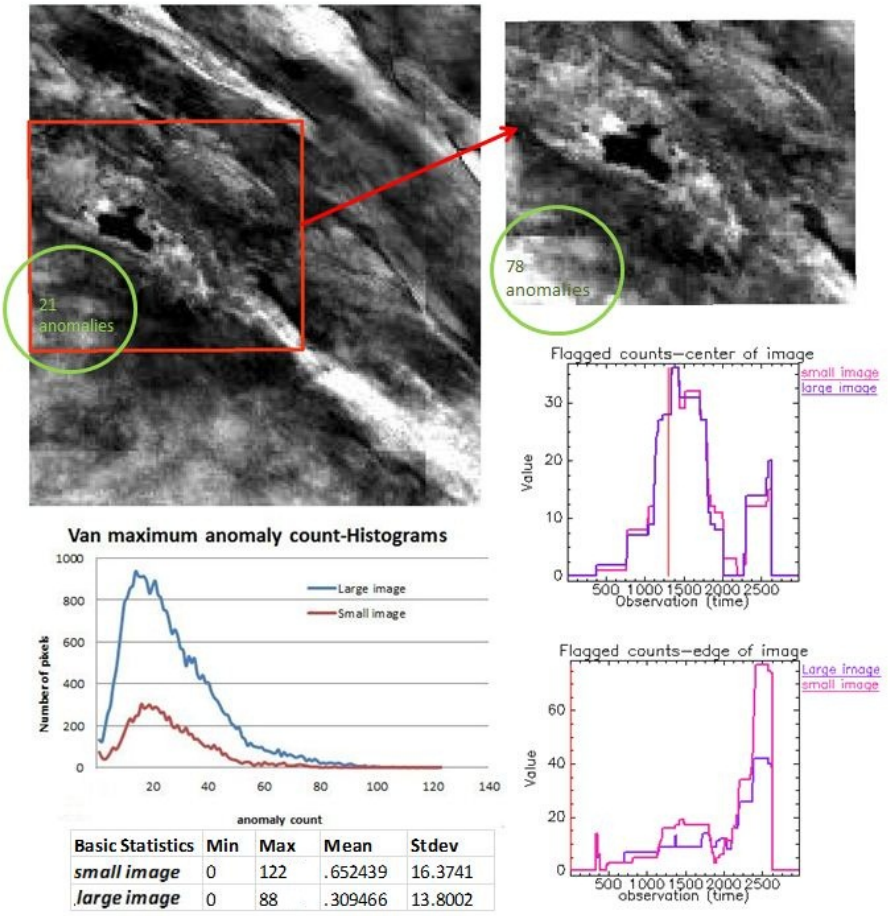


Figure 3.1: Evaluation of existing scheme. A1: detail of normalized time series of a dataset stacked without insertion of blank observations in the place of missing values. A2: the result of anomalous counts based on time series (A1). B1,2: the same for a dataset including blank observations in the place of missing values. A2 and B2 are flagging anomalies in the same available images, but the result is different because B2 is flagging based on the actual time interval between observations. C: detail from normalized time series, showing remnants of diurnal patterns. D: clouds remaining in a TIR image from Van after the application of the RAM mask. E,F: remaining clouds in time series from Van and Bam respectively.

because in the time series of the normalized values, subsequent observations might be the result of normalization with different numbers of neighbours coming from different locations (because a pixel from a specific location could sometimes be available and sometimes masked out as clouded). The effect of data availability was studied by comparing results of the application of the existing scheme in Van, using datasets of different spatial extent. The same area, captured in images of 330x230km, was also studied with images of larger extent (770x550km). It was expected that the edges of the smaller image would appear differently in the larger image after processing, because in the larger image more kernel pixels were available for normalization. Normalization was done with a kernel size of 120km and the temporal window for the filtering was 7 days. The results confirmed the existence of distortions in the edges of the smaller image, where the kernel was by default incomplete. Areas appeared as anomalous in the edges of the smaller image while in the larger image they were not flagged as anomalous. However, the differences in the central part of the images were subtle, and the flagged areas coincided spatially and in terms of intensity (Figure 3.2).



Basic Statistics	Min	Max	Mean	Stdev
small image	0	122	.652439	16.3741
large image	0	88	.309466	13.8002

Figure 3.2: Effect of data availability for normalization. The images show maximum anomaly counts and the corresponding statistics. In the edges of the smaller image, higher anomaly counts can be found (circled green). As seen also by the number of flagged values (bottom right) and the statistics, the smaller image is more variable. In the center of the images the flagged count time series is almost the same (middle right).

d. Anomaly detection Flagging of anomalous instances is based on the mean and standard deviation of the time series. Practically, this implies that a part of the dataset is always declared

anomalous. This can lead to false positives and differences in the detected anomalies depending on the size of the dataset. If the mean and σ of the years 1999-2004 are different than the mean and σ of year 2000 alone, then the flagged anomalies for 2000 will be different in a one-year and the six-year dataset (Figure 3.3). Furthermore, these statistics are affected by outliers, so anomaly detection can be influenced in the case of clouds or artifacts due to sensor malfunction.

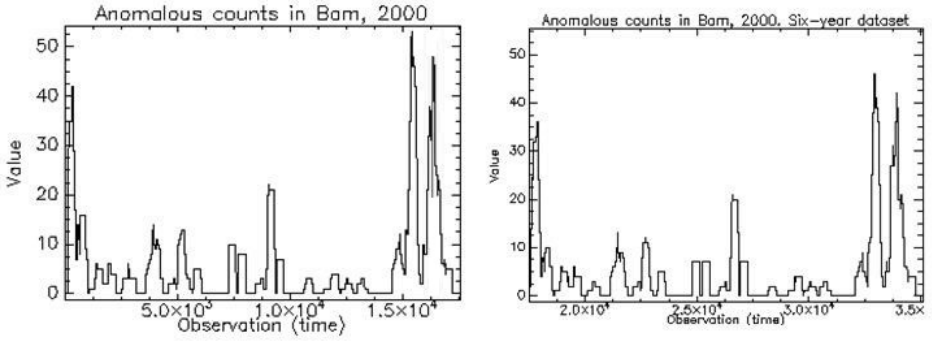


Figure 3.3: Effect of dataset length on anomaly detection.Left, flagging anomalies in Bam (2000) based on a one-year dataset; Right, based on a six-year dataset. Less anomalies are counted with the use of a longer dataset.

3.1.2 Synthetic anomalies

Synthetic anomalies were imposed in both Van and Bam datasets to assess the performance of the existing detection scheme. These were applied in an earthquake-free year (2000 for Bam, 2009 for Van, no earthquakes registered with magnitude >5) and the following points were checked:

(a) the limits of thermal increase which can be detected by the system (3-10^{o}K)

(b) the extent of detectable anomaly (1-20 adjacent pixels)

(c) the season. One setting was the month of occurrence of the future earthquake, to compare the synthetic retrieved peak with the peaks registered in the original dataset in the month of the earthquake. Another synthetic anomaly was placed in a different period of the year, to check for retrievability in different climatic conditions. Preferably, this would contain no peaks in the original processed dataset.

(d) the area (part of the image). This setting was focused in detectability of synthetic anomalies regardless of the differences of signal variability in different locations. One of the filtering products, the maximum anomaly count image (see section 2) shows which locations have more anomalous behaviour; this can be due to many factors, for instance topography. A high anomaly count in an earthquake-free dataset is considered as high natural variability of the recorded signal. In order to assess the role of the existing natural signal variability in the detection procedure, synthetic anomalies were imposed in parts of the original dataset with maximum anomaly counts, and in areas without anomalous instances (respectively, areas of high and low background variability).

The duration of the anomaly was in all cases 5 days.

The results showed that the most important factor affecting the detection of the imposed anomalies was the existence of background variability (Figure 3.5). Regardless of spatial extent (1-20pixels) anomalies ranging from +3-10^{o}K were clearly retrievable when imposed in an area that had shown low levels of natural variability in the original dataset. As can be seen in Figure 3.5, a 5-degree anomaly is clearly detected in July, a month with almost no background variability;

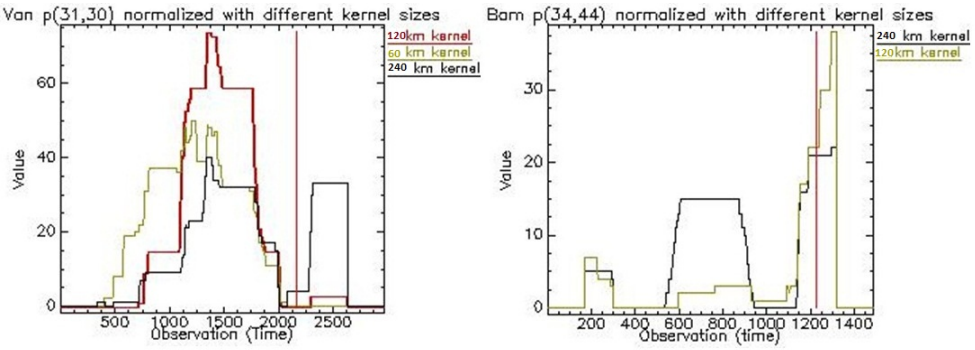


Figure 3.4: Sensitivity analysis: choice of kernel size in Van(left) and Bam(right). The 120km kernel shows the largest anomaly count in both cases, and the resulting peak is more discernible from the smaller ones.

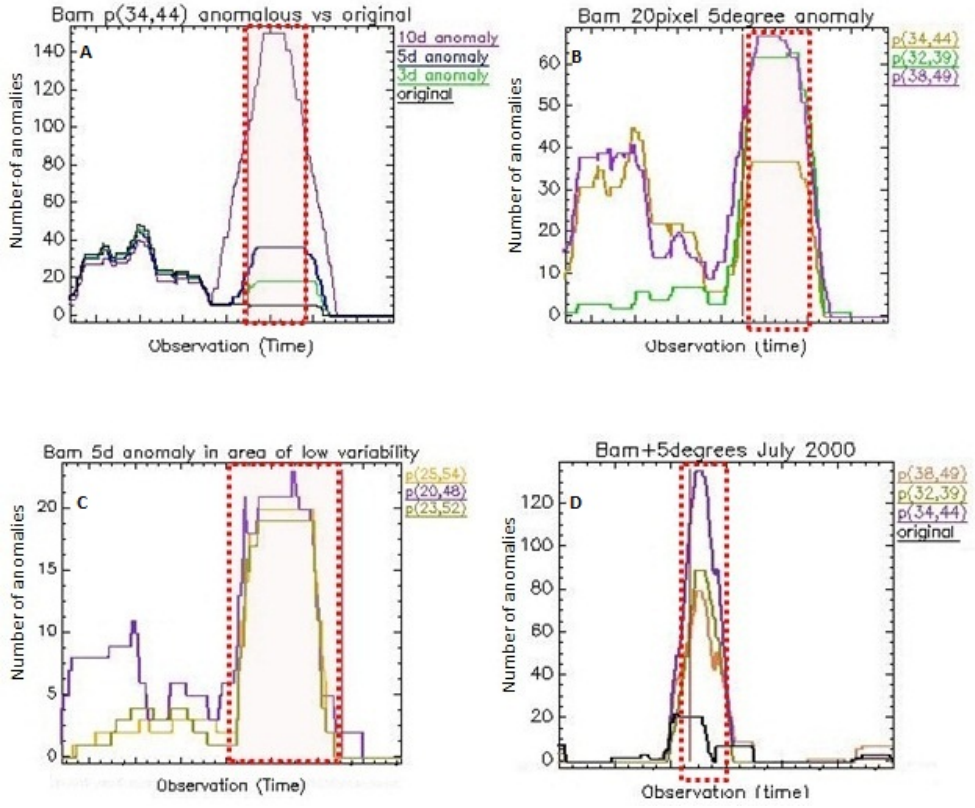


Figure 3.5: Synthetic retrieval based on the existing scheme. In all cases, time of imposed anomaly is designated by a dashed red rectangle. A: retrieval of different degrees of imposed anomaly. In this pixel, only the 10 degree anomaly stands out clearly. B: detectability of the same 5-degree anomaly in different pixels. Depending on the pixel, the anomaly detectable (pixels [32,39], [38,49]) or not. C, D: As opposed to A and B, these graphs show synthetic anomalies imposed in location (C) and period (D) of low background variability. The anomaly is clearly retrievable in both cases regardless of the pixel.

not only was it detected, but it was the highest peak in the whole dataset. The same holds truth for any of the locations (pixels) tested, as well as for the case of December (earthquake month) in an area of low maximum anomaly count. However, the same 5-degree anomaly when placed in a location of high background variability, is detectable in some pixels but not detectable in others: the peak corresponding to the imposed anomaly is still there, but it is lower than the existing peaks of the background signal. The 10^o anomaly was clearly retrieved regardless of the natural variability of the pixel.

3.2 APPLICATION OF STL DECOMPOSITION

The goal of STL application was to define daily, seasonal and inter-annual patterns in the data. Each decomposition defines two kinds of components, one with a low- and one with a high-frequency of fluctuation, so potentially more decompositions would be needed for extracting all of the required components. The resulting products are referred to as "seasonal component" (the one with the highest frequency of fluctuation), "trend" (the one with lower frequency of fluctuation) and "remainder" (the result of removing the other components from the raw data). The decision on which part of the variation goes to which component largely depends on the choice of settings for the decomposition, as defined by the user (section 2.2.2).

Thus, the first task on STL application was to describe optimal settings for decomposition which allowed for extraction of diurnal, seasonal and inter-annual components in a way that keeps anomalous instances intact in the remainder. The trials were carried out in the dataset from Bam. Structural outliers, i.e. observations resulting from an apparent malfunction of the sensor (Figure 2.4) were not removed from the raw dataset, to act as controls for the robustness of the various decompositions to outliers.

The primary choice of STL settings is determining the length of the cycle subseries, which will be the basis for the first LOESS smoothing (step 2, 2.2.2, figure 2.3). Two settings were tested. On one hand, the term "daily decomposition" from here onwards signifies a decomposition where the length of the cycle subseries is equal to the number of observations within 24 hours. In this way, the diurnal fluctuations end up in the so-called seasonal component of STL, and lower-frequency variability is contained in the so-called trend. On the other hand, the subcycle length of a "yearly decomposition" will correspond to the number of observations within a year. The resulting seasonal component will depict the fluctuations between the months of each year (climatic seasonality), the trend will describe inter-annual changes, and the diurnal fluctuations will be shared somehow between these two components.

Because the patterns required for the study were three, and each decomposition describes only two levels of fluctuations, at first the possibility of two subsequent decompositions was examined. The first would remove the daily and seasonal part of the variability and the second would deal with the extraction of the remaining low-frequency fluctuations. This follows the suggestion of the authors of the method (Cleveland, 1990; P.Hafen, 2010) to first remove high-frequency fluctuations, and was implemented as a daily decomposition followed by a yearly. Following this scenario, the yearly STL was applied: (a) on the remainder of the daily STL, (b) on the initial raw time series after removing the daily fluctuations defined by the daily STL. The inverse scenario was also tested, i.e. first a yearly decomposition followed by a daily. The results showed (Figure 3.6),

(a) a high amount of variability ending up in the trend of both daily and yearly STL, (b) an unclear distinction of low and high-frequency fluctuations in the case of yearly STL, as the diurnal variability was shared between components, (c) a lack of robustness in yearly decompositions despite the high number of outer-loop iterations, and (d) minuscule levels of interannual variation.

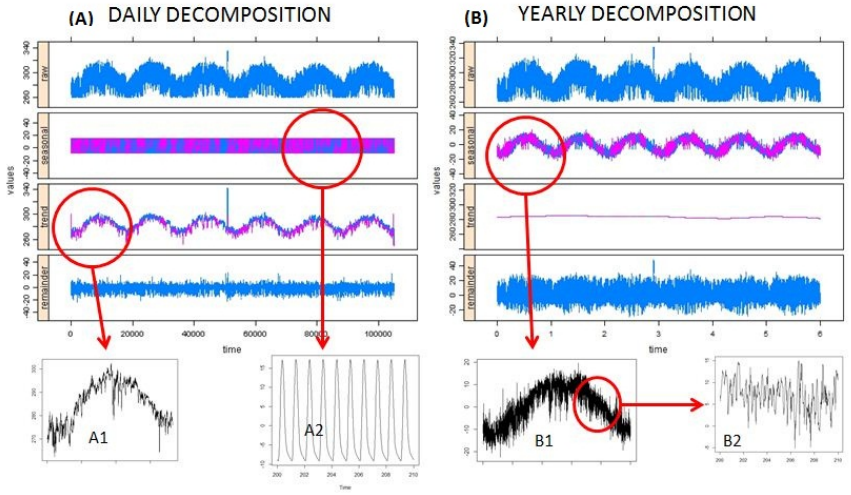


Figure 3.6: Choice of STL settings: Daily (A) vs Yearly (B) decomposition. A1: detail of trend, capturing climatic seasonality. A2: detail of seasonal component, containing diurnal fluctuations. B1: detail of seasonal component, describing climatic seasonality for the same period as A1. B2: zoom in the seasonal component. As can be seen, A1 and B1 are not the same although they should be describing seasonality of the same period. They contain a lot of variability (B2) and it is not clear where this variability comes from. The yellow circle shows that part of the outlier is still modeled as part of the seasonality, despite the high number of robustness iterations. In C (trend of yearly decomposition): the minuscule inter-annual fluctuations throughout the dataset.

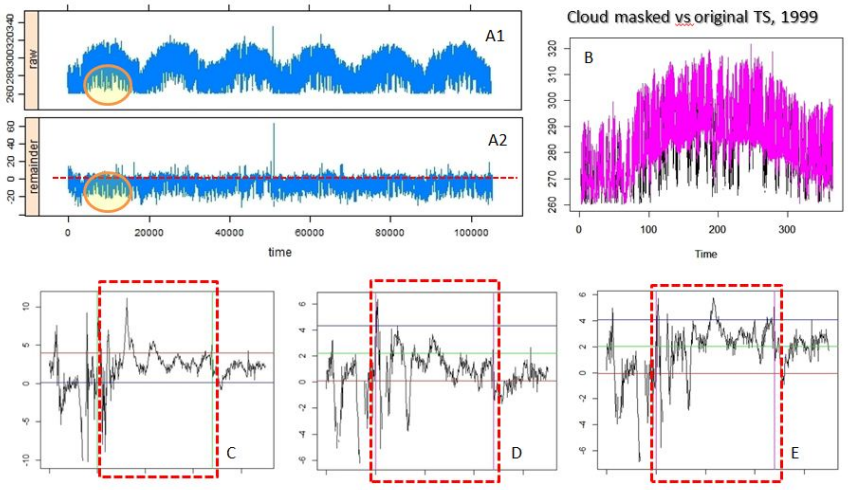


Figure 3.7: Coincidence of clouds and low remainder values. A1: Raw time series, A2: remainder of STL. The circles indicate a characteristic case in the first year, analogous cases can be seen in the rest of the years especially in the summer. B: Result of cloud removal on the basis of -(mean+2*σ). The original time series is superimposed by the cloud-removed one. The remaining black edges represent the parts of the original signal which have been removed as clouds. C, D, E: effect of different cloud removal approaches on synthetic anomaly retrieval. In C, anomalous remainder values without cloud removal. In B, remainder cleared of values below the set threshold. In D, the result of removing Raw TS values where the anomalous remainder was below the threshold, and then performing another decomposition. Setting D shows better performance, as the anomalous area is highlighted compared to C and B.

The above, along with the computational demands of two decompositions, led to the alternative scenario of testing a daily decomposition followed by trend post-smoothing. This would extract a well-defined diurnal pattern and a smoothed seasonal one, leaving as much variability as possible in the remainder to avoid losing anomalous signal. The settings were tried and evaluated using the available diagnostic plots (Figure 2.4), the number of iterations was determined taking into consideration the convergence criterion mentioned by the authors of the method (Cleveland, 1990) and the final settings and R scripts can be found in Appendix A.

The results were cross-examined in the case of artificially imposed anomalies: the original dataset was decomposed before and after imposing synthetic thermal increases of various characteristics. The imposed anomalies had different duration (3-6 days), shape (uniform increase/ gaussian increase), magnitude (4-6 degrees), and position in the time series (in different seasons and in areas with and without missing values). In order to compare original and anomalous remainders, two approaches were tested: division (as would be consistent to the normalization process of the existing detection scheme) and subtraction. The remainder of the decomposition of the original dataset, before imposing synthetic anomalies (from here on, "original remainder") was either divided by or subtracted from the remainder of the decomposition of the same dataset after imposing an artificial anomaly (from here on, "anomalous remainder"). Given the fact that STL is an additive decomposition, it was expected that the subtraction would outperform division. Indeed, subtracting the remainders resulted in a clear delineation of the anomalous part of the dataset, both temporally and in magnitude. While division also gave an indication of the existence of an anomaly, the anomalous part was not clearly defined and did not give a straightforward relation to the magnitude of the imposed anomaly (see figure 3.8). So the choice for all further comparisons and retrieval trials was set to subtraction. Overall, the remainder of the decomposition proved to retain the anomalies fully, regardless of their shape, duration, magnitude, position in the time series and missing values (see figure 3.8).

3.3 ANOMALY DETECTION AND STANDARDIZATION OF DETECTION SCHEME

In the present study, the initial assumption was that the majority of the variability of the data was due to seasonal and diurnal effects. STL was employed to extract this part of the variability. Thus, it was expected that the remaining signal could already be considered anomalous in the sense of not having been generated by the same process as the removed components of the time series (reference to definition of anomaly). This would mean that anomaly would be defined intrinsically by the decomposition and the threshold of $mean + 2\sigma$ for definition of anomaly would no longer be needed. Such a definition of anomaly would be qualitative, instead of statistical, and more straight-forward in its relation to the physical causes of thermal signal variations.

However, the results showed that the remainder of the decomposition contained itself a high level of fluctuations which, at least in the length of the time series examined, was not inflicted by the defined seasonal and diurnal patterns (Figure 3.10). These fluctuations could be due to coincidental factors, as the presence of non-removed clouds or unobserved deficiencies of the dataset (sensor failure, corrupt data). Also, they might be caused by recurrent effects, which would have a frequency different than the defined seasonal component and trend; these fluctuations could potentially be further modeled, but were not extracted by STL decomposition. One decomposition cannot serve as a universal solution to modeling all underlying factors. A detailed modeling of the thermal variability of the signal of a specific dataset was beyond the limitations of the present study. So, when the need emerged to distinguish which part of the deviating values were caused by factors non-related to the impeding earthquake, given the lack of available modeling of these other factors, standard deviation was again the last resort. The $mean + 2\sigma$ threshold was used to

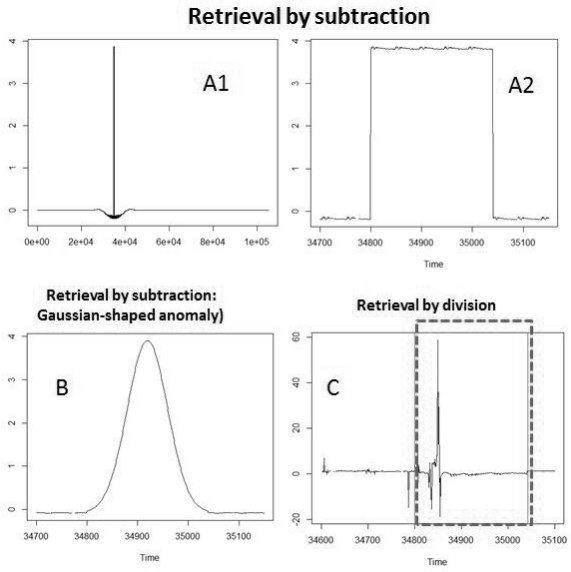


Figure 3.8: Anomaly retrieval by comparing the anomalous and the original time series. Added anomaly of 4 degrees retrieved by subtraction: A1, whole dataset, A2 detailed view; B: added gaussian shape anomaly of 4 degrees, retrieved by subtraction; C, anomaly of 4 degrees retrieved by division of STL remainders. Anomalous part of the dataset is enclosed in dashed lines. With subtraction of the original from the anomalous remainder, the imposed anomaly is fully retrieved both in shape and magnitude, and it is always the only anomaly detected. Division results are rather unclear. The anomaly is still included but not clearly delineated.

isolate the most extreme of the deviating values captured in the remainder (this corresponds to a statistically chosen 5% of the remainder). A further effort was made to qualitatively describe potential factors underlying the observed anomalies, based on topography and geological substrates of the study areas.

Incorporation of STL in the existing scheme was tried in two settings: (a) decomposition of

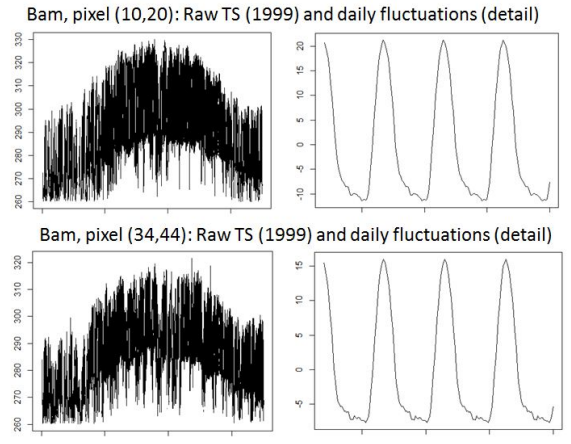


Figure 3.9: Comparison of Time Series of pixels from different substrate. In A, raw time series and detail of diurnal variation, extracted from a location with sandy substrate (thermal inertia=.024, Sabins (1997). In B, time series of a basaltic pixel (thermal inertia=.053, Sabins (1997) and the daily fluctuations of its values. Pixel A has a wider range of temperature values. Sampling locations of the two pixels are shown in figure 3.14

the already normalized time series, (b) decomposition prior to normalization.

For the first setting, the basis of detection was the (spatial) comparison between the pixel and its neighbours. The role of STL was to remove remnants of recurrent seasonal/diurnal patterns from the normalized time series and thus reduce false positives. The normalized time series shows how the difference between the central pixel and its surroundings changes through time; STL removes daily/seasonal fluctuations in these changes, and the remaining changes have to be interpreted in relation to other factors (e.g. a strong earthquake). The application of STL did refine the results of the existing scheme. Less peaks appeared in the flagged count graphs, and the remaining ones were better delineated (Figure3.11, Figure3.12, Figure3.10). The improvement in performance is qualitative: it is based on discarding anomalous instances which were falsely considered anomalous because they contained remnants of diurnal variability.

The second setting used STL to detect deviations from the expected, normal pixel behaviour *before* comparing the pixel to its surroundings. "Normal pixel behaviour" refers to pixel variability determined by daily and seasonal patterns. The remainder of the decomposition held all fluctuations in the data which cannot be attributed to diurnal and seasonal effects. In the second setting, these deviations captured by the remainder of STL of the central pixel were compared to the remainder of the decomposition of neighbouring pixels. This follows the principle of normalization, as is applied in the existing methodology; however, while the comparison in the existing scheme is division-based, the comparisons of STL remainders were performed by subtraction (for reasons based on the results of the previous section). Two approaches were tested for the second setting:

1. Decomposition of all surrounding pixels, averaging of the remainders, and subtracting the average neighbour remainder from the central pixel remainder

2. Averaging the values of the surrounding pixels, decomposing the average, and subtracting the remainder of the mean signal decomposition from the central pixel remainder

In all cases, the central pixel was compared to the average of eight neighbours, perimetrically placed in a kernel-like ring around the central pixel (Figure 3.14). The choice of pixels was made in a way that would cover areas from the existing different types of terrain. Unfortunately, time and computational limitations did not allow the use of a complete kernel of neighbouring pixels. The two approaches gave similar results. The imposed anomalies could only be detected by the $mean + 2\sigma$ threshold if they were imposed on a spot where there was a sufficiently high pre-existing background signal. The existing signal fluctuations were already higher than the imposed alteration. So, if the synthetic anomaly was placed in an area with very low natural variability, it was overshadowed by peaks in other parts of the dataset. The imposed anomaly did clearly increase the signal, as was proved by comparing the remainders of the original and the anomalous time series; however, not enough to overcome existing peaks. Some of the deviations contained in the remainder were already larger than the imposed increases (Figure 3.10)

Both detection combinations (i.e. STL either before or after normalization), when tested with synthetic anomalies, appeared to be prone to the effects of existing background signal. Imposed temperature increases were detectable by the threshold only when the pre-existing signal was high enough or alternatively when the rest of the signal fluctuations were lower than the imposed anomaly. Therefore, taking into consideration the computational requirements both in time and resources, the first setting (normalization prior to decomposition) was preferred for the subsequent analyses.

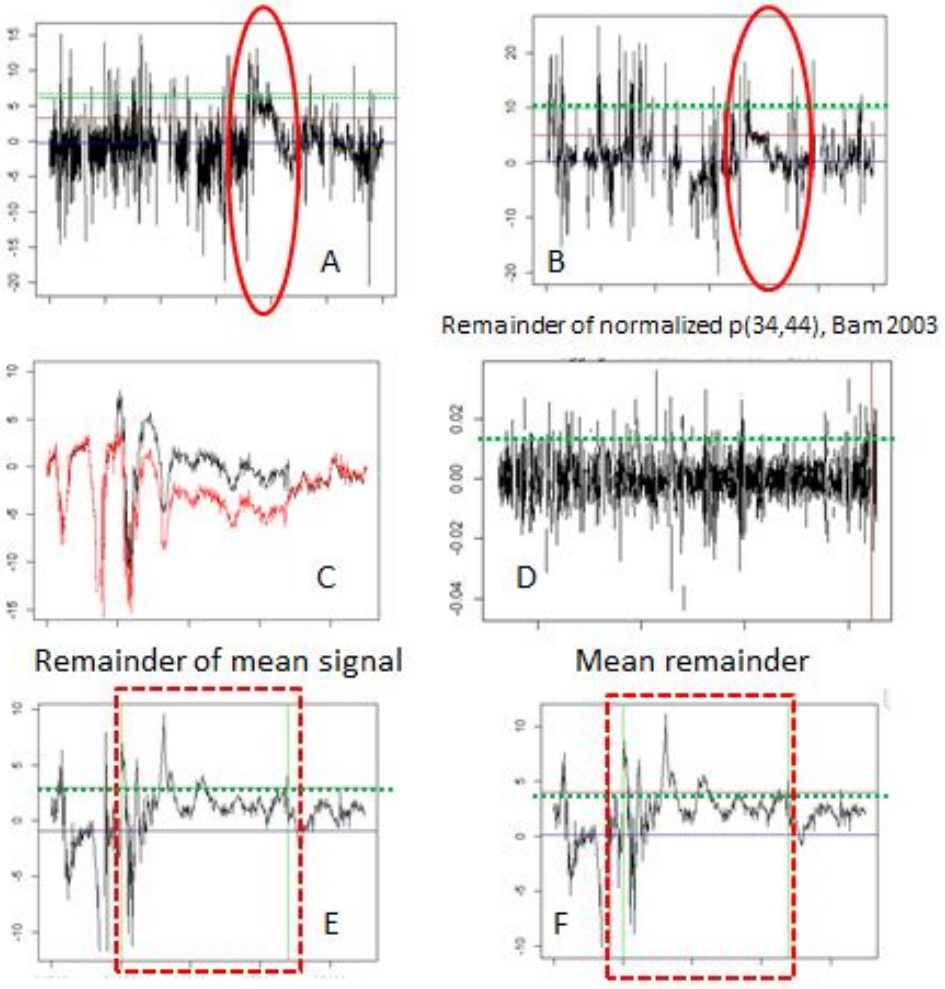


Figure 3.10: A, B: imposed anomalies (highlighted by red circles) in different samples of remainders. The anomalous areas can be clearly distinguished but are not detectable by the threshold (green line) because of the already wider existing signal fluctuations. C: Anomalous remainder (black) superimposed by original (red). The difference is clear in the region of imposed anomaly. Retrieval of anomaly is done with subtraction of the two, see Figure 3.8. D: Remainder of STL decomposition of normalized time series. The values above the threshold (green line) will be flagged as anomalous. E: Anomaly detection by comparing the remainder of the central pixel with the mean remainder of its surroundings. F:Anomaly detection by comparing the remainder of the central pixel with the remainder of the decomposition of the mean signal of the neighbours. The results of E, F are very similar: anomalous areas do not exceed the threshold and cannot be flagged.

3.4 CLOUD REMOVAL

Clouds in the image are introducing artifacts in the processing of the dataset and interfere with anomaly detection. The issue of cloud removal was more important for the case of Bam, because because there was no cloudmask product available for MFG data. There exist cloud-masking products for MSG datasets but unfortunately they were not available for the present study, therefore the cloud-removal attempts could not be validated.

The presence of clouds was obvious even after the application of the RAM mask, and the clouded pixel values were captured by the remainder of the decomposition. In the summer, for instance, it is true that the daily temperature range, especially in the area of Bam, is much wider in the summer (very warm days and very cold nights, as seen in the climatic records of the area- see Appendix B). But this is effectively captured in the decomposition: the seasonal component (Figure2.3) has a wider range every summer. So the remaining extremes can be found in the remainder, as shown by a visual inspection of the data (Figure3.7). This led to an attempt to remove clouds on the basis of the remainder. The idea was to remove values which were colder than the majority of the remainder. The selection was performed on a statistical basis: mean and standard deviation of the remainder were calculated and a threshold was set to $mean - 2\sigma$, so that all values falling below this threshold were considered extremely cold and were removed. Two settings were tested. The first removed *remainder values* lower than the threshold. The second included removal of *raw time series values* where the remainder was lower than the threshold, and followingly a new decomposition of the cloud-removed raw time series. Both settings induced improvements in the sense of reduction of noise (extreme peaks obscuring the imposed anomaly). The second setting showed better performance (Figure3.7).

3.5 APPLICATION OF REVIEWED SCHEME TO BAM AND VAN EARTHQUAKES

The complete 6-year and 4-year datasets from Bam and Van respectively were analyzed with the combined scheme. After normalization, STL decomposition was applied to remove seasonality leftovers, and flagging of the anomalies was carried out on the remainder of the decomposition. The daily components resulting from STL were quite small, but reduced the magnitude of most of the observed peaks, especially the largest ones. In the case of Van there was a noticeable change: all of the observed peaks were reduced after decomposition, with the exception of the peak registered before the earthquake. Although the same diurnal pattern was extracted from the whole dataset, the pre-earthquake peak was not influenced. Apparently the magnitude of the other peaks was partly due to the diurnal variation, and many instances which had been flagged as anomalous were reconsidered as normal when the daily patterns were taken into account. The magnitude of the pre-earthquake peak was obviously unaffected by daily fluctuations (see figure 3.12). Similar results were recorded in Bam, with more clear reductions in all of the observed peaks (Figure3.11). In both Bam and Van, large peaks were recorded prior to the earthquakes (see detailed views in Figure 3.11 and Figure 3.12). Although other large peaks were also present, the larger anomaly counts were the ones occurring before the earthquakes.

Finally, the terrain in both Bam and Van was studied in order to explain anomalous peaks in the study areas. Figures 3.13 and 3.15 focus on geological substrate, and figures 3.16 and 3.14 overlay detection products on DEMs in order to examine the effect of altitude. As can be seen in the figures, in both cases the anomalous pixels are located in areas with higher altitude relative to their surroundings, and in the proximity of fault lines. Also, they both come from volcanic/basaltic substrate, which has higher thermal inertia properties (Sabins, 1997) and is thus faster to heat up and cool down. The role of substrate in the thermal behaviour of the pixels can be seen already in the raw time series: Figure 3.9 shows samples of one pixel from volcanic substrate of higher

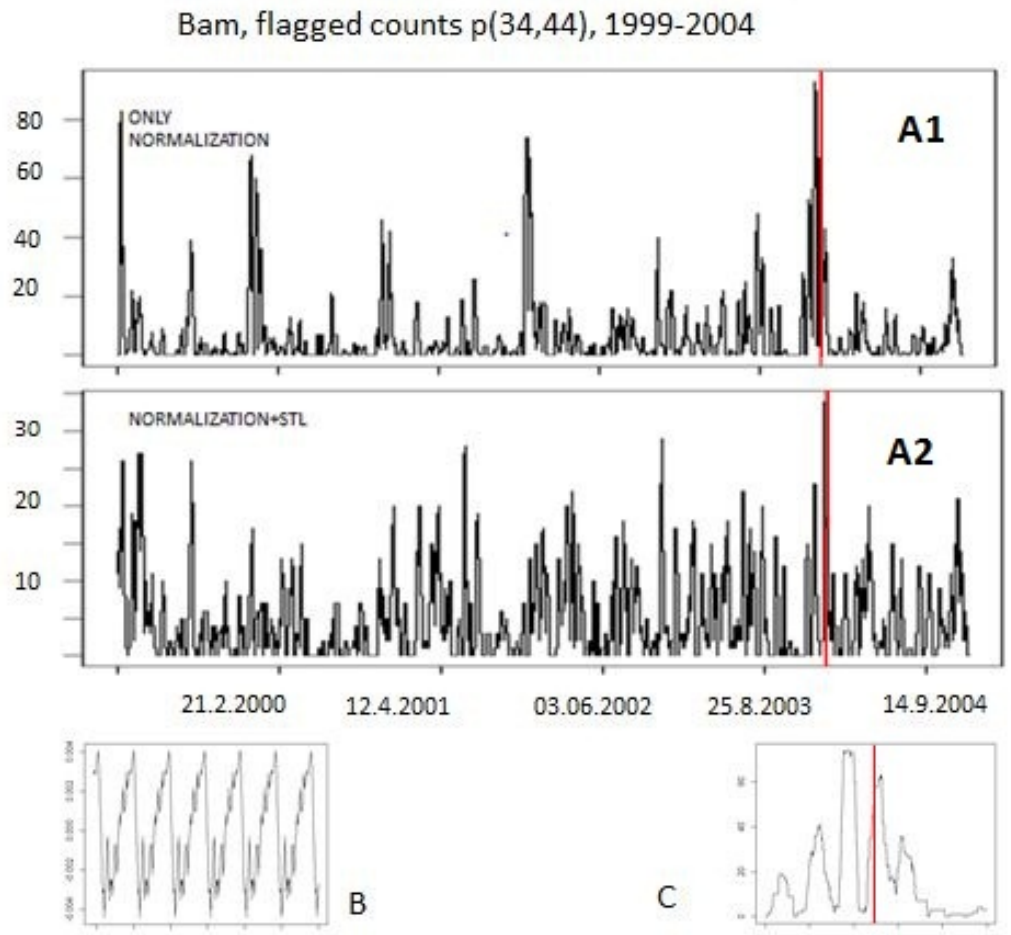


Figure 3.11: Application of reviewed scheme to Bam. Red vertical lines designate occurrence of the earthquake. A1: anomaly detection after normalization only. Numbers of flagged instances in Bam, whole dataset. C: detailed view of the same product. A2: Numbers of flagged values after applying decomposition on the normalized time series. All peaks are attenuated. B: detail of the diurnal fluctuations extracted by the decomposition

altitude and one pixel from an area with sandstones, located in a lower area surrounded by mountains. It is clear that the range of values is wider in the first case. Temperature range is lower for materials with high thermal inertia, however other factors like shading and altitude should also be taken into consideration (Sabins, 1997).

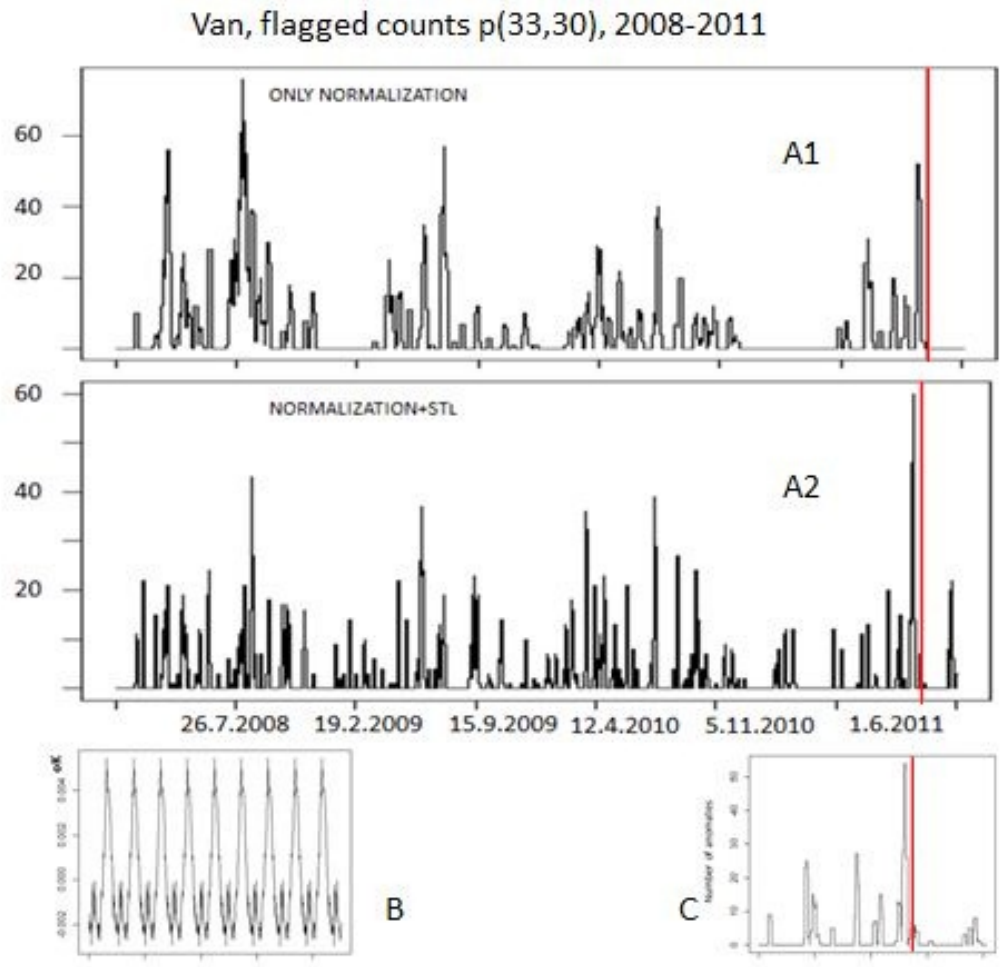


Figure 3.12: Application of reviewed scheme to Van. Red vertical lines designate occurrence of the earthquake. A1: anomaly detection after normalization only. Numbers of flagged instances in Van, whole dataset. C: detailed view of the same product. A2: Numbers of flagged values after applying decomposition on the normalized time series. B: detail of the diurnal fluctuations extracted by the decomposition. All peaks are attenuated after decomposition, but the pre-earthquake peak persists and this changes their relative magnitude.

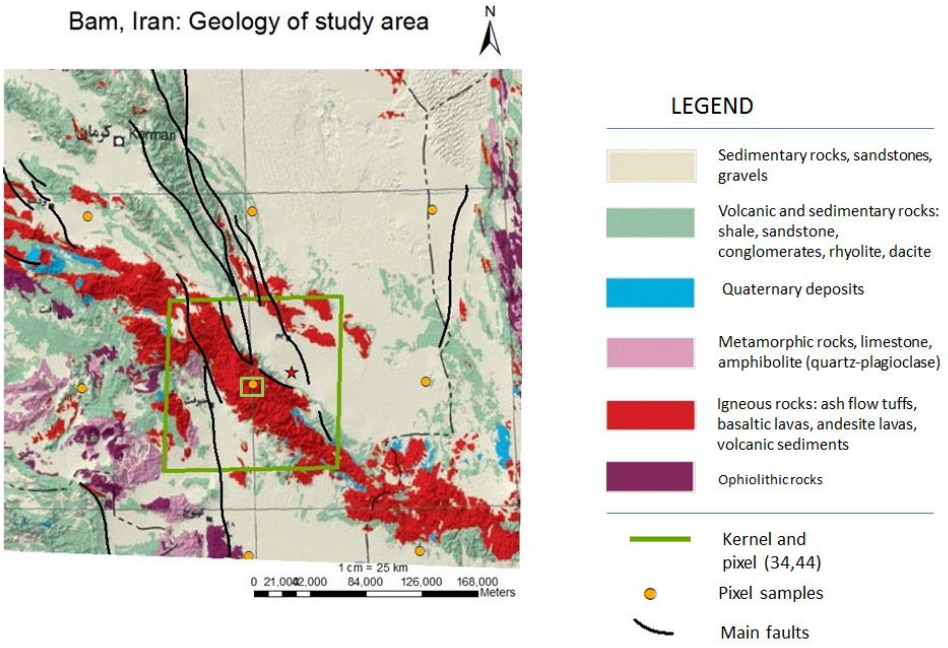


Figure 3.13: Geology of study area in Bam, Iran. Adopted by the Geological Survey of Iran and USGS. The green lines show the kernel used for normalization and the central pixel, which was the pixel which showed the pre-earthquake peak. Yellow points show sample pixels used for retrieval with STL and for tracing differences in thermal behaviour depending on the substrate.

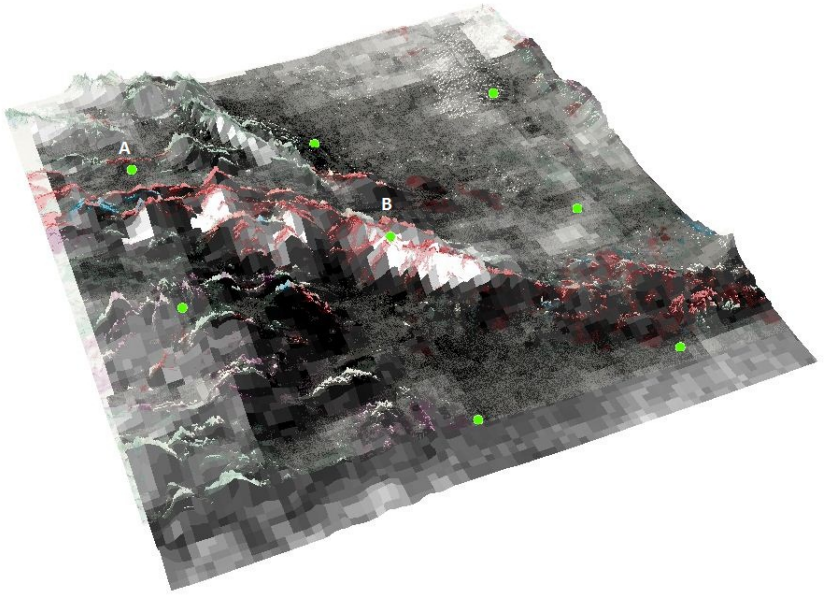


Figure 3.14: Image showing areas of maximum anomaly counts, superimposed on an ASTER GDEM —product of METI and NASA. White/lighter areas have higher anomalous counts, and the colour shades correspond to the geological substrates shown in the previous Figure 3.13. Time series from pixels A and B are further compared in Figure 3.9. Areas with most anomalous values are located in areas of higher altitude and volcanic substrate.

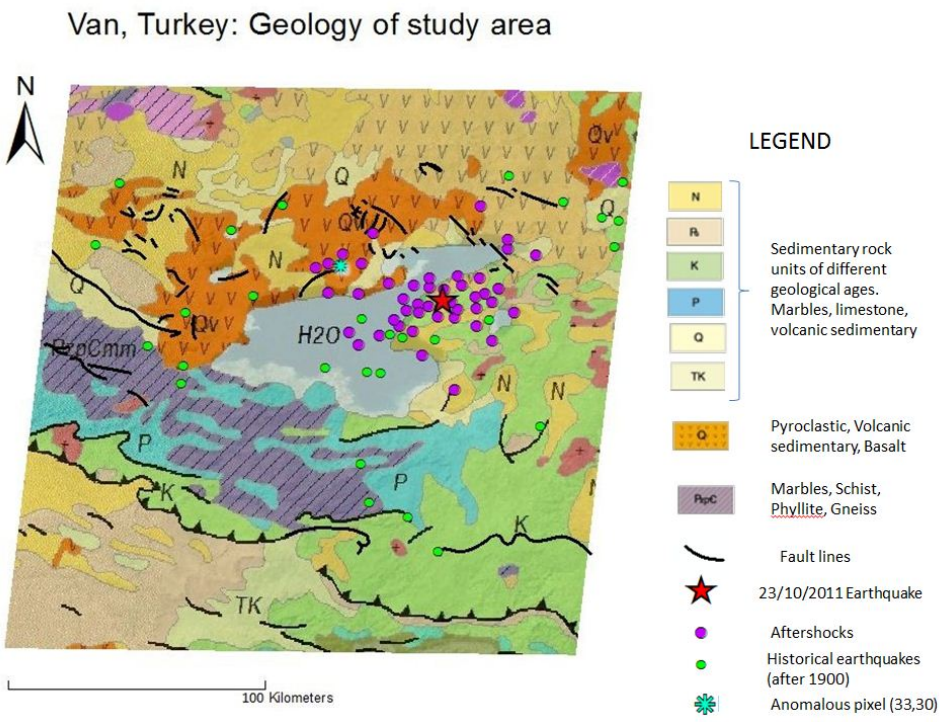


Figure 3.15: Geology of study area in Van, Turkey. Adopted by the Geological Survey of Turkey and USGS. The anomalous pixel is indicated in blue in the north of the lake, on basaltic-volcanic substrate and on the location of a fault line.

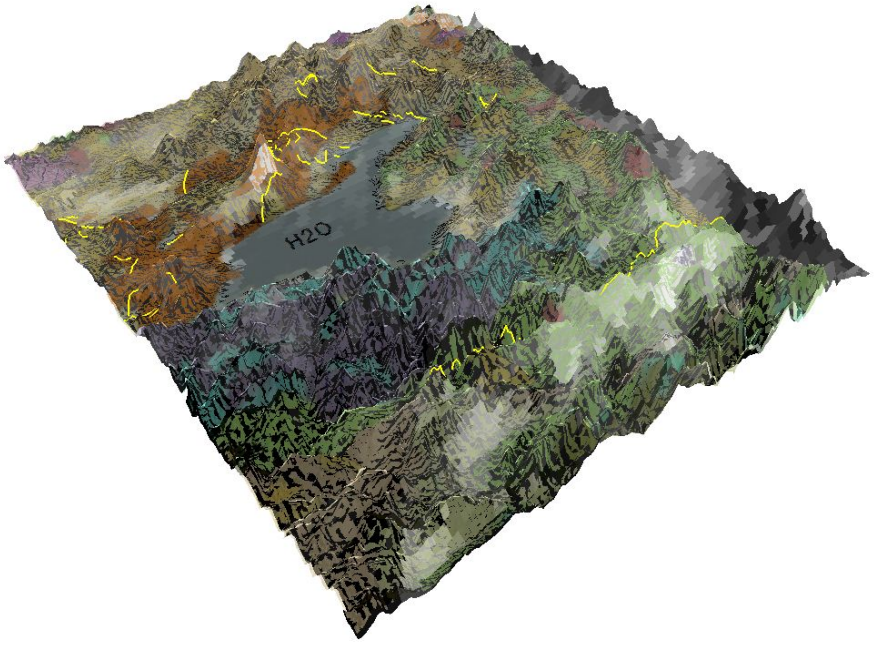


Figure 3.16: Image showing areas of maximum anomaly counts, superimposed on an ASTER GDEM —product of METI and NASA. White/lighter areas have higher anomalous counts, and the colour shades correspond to the geological substrates shown in the previous Figure 3.15. Areas with most anomalous values are located in areas of higher altitude and volcanic substrate.

Chapter 4
Discussion

The key components of the present study were, (a) the review of the current detection scheme, and evaluation of its performance for the first time on the basis of synthetic anomaly retrieval, and (b) the investigation into the potential of the application of STL decomposition for the removal of recurrent patterns in the data and the anomaly detection procedure.

The review of the existing scheme underlined some important issues concerning the detection. First, the statistical inconsistency resulting from the variability in missing data for the calculation of the normalization kernel. The effect of data availability has not been quantified in the present study but it was seen in the results. The potential distortions could be minuscule, but theoretically, a time series of values calculated with different available data points could be containing artifacts and introducing noise in the detection procedure. This problem affects more severely the edges of an image and can be alleviated by choosing sufficient image size according to the area of interest. However, it still exists in the central parts after the removal of clouds.

Secondly, the basis of anomaly detection affects the result in different ways. The temporal side of the detection, which is based on the historical record of of pixel values, has the advantage of being pixel specific; and, given the favourable characteristics of the geostationary sensors, this povides a good basis for the definition of normal, expected pixel behaviour. However, it is faced with difficulties in the case of an unusually cold or warm period: a warm month appears as anomalous for the whole image, while an anomaly contained in an unusually cold month would be invisible, being still colder than usual. This is the biggest advantage of the spatial side of the detection, i.e. comparison of the pixel with its surroundings: it can remove effects of factors common in the central and the neighbouring pixels. An unusually warmer/colder month is warmer/colder for all pixels and the normalization is expected to remove the background temperatures and highlight the spatial differences of the pixels. Still, as shown in the results, some seasonal and diurnal patterns persisted. Terrain differences possibly played a role in that, as the differences in the thermal behaviour of various substrates and geomorphological features can also influence the response of different parts of an image to the same “warmer” or “colder” environmental conditions. Also the choice of kernel size is crucial for normalizing pixels sufficiently different, to permit the distinction of anomalies, but at the same time not too far, so as to avoid normalizing pixels with totally different local climate. This was shown in the sensitivity analysis of the present study. It can be argued that the suggested kernel size is compatible with potential physical processes underlying earthquake effects on the thermal signal, while larger kernels seemed to introduce extra peaks which could be indicative of the differences in prevailing local conditions (Figure 3.4).

Overall, the temporal dimension of the detection is affected by climatic extremes while the spatial by the diversity of the terrain and the morphology. A combination of the two dimensions would be ideal and this is what was attempted in the present study. The challenging part is to identify which marked unusual occurence can be attributed, or linked, to which of the factors potentially involved.

The use of standard deviation, although statistically justifiable, imposes some concerns:
(a)It is in itself prone to the effect of outliers, and these outliers may be artifacts caused by the

presence of clouds but distorting the whole detection process.
(b)The length of the time series is important, because the standard deviation of two-years of data might differ from the one of a six-year dataset and the flagging will consequently change. This was an issue that prevented any quantification of peaks in the case of synthetic anomalies. Synthetic anomalies were tested on a one-year dataset, in order to ensure a complete earthquake-free period. Longer time series with this characteristic were not available. On the same grounds, no tests were done on different durations of anomaly. So any quantifiable connection between the characteristics of the synthetic anomaly and the number of observed anomalies could not be justified; after all, if the detection scheme eventually reaches an operational monitoring level, the lentgh of available time series will be constantly changing (data will be continuously aded). So this remains an issue to be dealt with in the future.
(c) Anomaly detection based on mean and standard deviation implies that part of the dataset is always considered anomalous, which is definitely not always the case. This is an issue that holds truth also for the cloud mask tested in the present study-it is a positive thing that it is affected by outliers (clouds) but removes always a part of the dataset. The criterion of too much deviation from the expected value is justifiable and the mask appeared to reduce cloud effects, but it does not always lead to removal of all the clouds. Already Khan mentioned about the RAM mask that it implies the existance of a cloud free part of the image all the time. Cloud removal in itself is a difficult issue and also a matter of definition: there are many types of clouds, low clouds, ice clouds, dust clouds-they cannot be distinguished and removed by the application of a single set of criteria. And their effect in changing the temperature differs accordingly.

Returning to the issue of anomaly detection, the decomposition remainder revealed the existence of quite some deviations from the expected pixel values, which could be attributed to a diversity of factors. This prevented the much desired bypass of the σ-based anomaly definition. Imposed anomalies were detectable only depending on the existing background signal. STL proved to be a powerful tool in defining and removing diurnal and seasonal variations, and its flexibility may imply there is room for further improvement if the user decides to surpass the strict definition of seasonality and include more of the variability in the seasonal component. However, this should mean a good knowledge of which parts of the remainder can be considered "normal" as in "recurrent, although with lower frequency". The present study was exploratory in this sense and for that reason the attempts were focused on removing only clear diurnal/seasonal patterns and leaving as much of the variability in the remainder, in fear of discarding potential anomalies. Maybe a subsequent attempt of modeling the remainder could reveal patterns related to specific factors, like topography.

Actually the result of processing the whole datasets of Bam and Van, it can be noticed that peaks are recorded almost recurrently in the period before the earthquakes. In Bam, the time series has more irregular variability. The anomalous pixel is located in a high-altitude, shaded position, of volcanic/basaltic substrate. It seems to be usually colder than its surroundings. In Van, again the pixel comes from a high altitude volcanic area, and appears warmer in late summer/early autumn; the peaks there are more recurrent. This might also be related to the vegetation in the area: the land use in the area is dominated by agriculture, and this period coincides with crop yield and routine works on the soil of the farms. The common characteristics of the location of the peaks in the study areas support the possibility of potential further modeling of the remainder, and it is important that the modeling is based less on statistical description and more on the physical factors affecting the temperature changes. The anomalous pixels are in both cases very close to faults and the magnitude of the peaks observed before the earthquakes is higher than the rest. Spatial and temporal coincidence with an earthquake is an indication of a relation between earthquakes and anomalies, but more research is needed to reach positive conclusions. Given

the role of existing background signal, it could be argued that thermal anomalies caused by the earthquakes are contained in the excess anomaly counts and highlighted because the background variability is already high. And, on the other hand, an earthquake-related anomaly which happens in a period of low variability, would remain undetected. If these patterns of recurrent anomalous peaks could be modeled on a physical basis, then the difference in magnitude of the peaks could be a more direct and quantifiable indication of a link to the earthquakes. In combination with a more effective cloud masking product, this would reduce the noice and excessive deviations considerably. Another option would be the application of different approaches to anomaly detection in the remainder. For example, ARIMA (as suggested by P.Hafen (2010)) or the setting of a trained, normal time series of earthquake-free years, to used as a reference for detection in a test-series (as proposed in Cheboli (2010)). Window-based or spectral techniques for reducing the dimensionality of the data might also be applicable. Cross-checking with ground-based observations could further rule-out some of the observed peaks.

More technical remarks include the number of pixels used for the remainder comparisons in the previous section. It can be argued that it was a very limited number of pixels and therefore the results included a lot of noise. The present study was exploratory and the computational and time limitations did not allow for the use of a complete "kernel" in the settings, so indeed there is room for improvement in this sense. Comparisons between the different settings of the detection scheme might also be more accurate with the use of more pixels.

Also, the number of synthetic anomalies tested was not exhaustive. The attempt was made to include a variety of settings and the most important parameters, but more tests can be done to explore the limits of the existing methodology. As to the settings of the temporal window used for the detection, it should be noted that the length of the window has to be larger than the anomaly in order to detect it (Cheboli, 2010). This is not a major issue so long as it is taken into account. Also, it would be more accurate if the temporal window was not centered on the pixel timestamp but included only past values until the pixel: this would be more functional in operational mode and would also avoid temporal misplacements of the anomalous counts towards future timeslots.

Finally, visualization of the results is a very important issue; given the variability of the image because of different factors, the maximum anomaly count proved insufficient to trace the anomalous areas. Indeed, when testing synthetic anomalies, thermal increases imposed in areas of low variability were not distinguished by their maximum anomaly count because this remained still lower compared to areas of high natural variability. Investigation for anomalous instances was done manually, knowing the area of synthetic thermal increase. This means that in the case of a real earthquake, when there is no indication of the location of a potential anomaly, there needs to be a sufficient visualization technique to trace the anomaly in the whole dataset. An attempt was made before by Buzzo (2012); it gave a three-dimensional image of the occurence of anomalies and the earthquake epicentres, but could only depict a small part of the dataset and valuable information was lost due to computational limitations.

In conclusion, the present study led to improvements in the performance of the existing scheme, explored its limitations and supported the application of STL decomposition for handling the temporal aspects of thermal anomaly detection. The results showed that thermal peaks have been recorded prior to the tested earthquakes and in fault areas, but it is still not possible to determine whether they contain earthquake-induced anomalies or quantify them. There are apparently many factors influencing the TIR signal, and so far it is not possible to quantify the contribution of each one of them in the deviations from the expected pixel behaviour. The current approach can be further elaborated and developed. Furthermore, it can see applications in many other fields related to thermal anomaly detection and processing meteorological and other physical effects.

Bibliography

Abraham, B., Chuang, A., 1989. Outlier detection and time series modeling. Technometrics 31, 241–248.

Aliano, C., Corrado, R., Filizzola, C., Genzano, N., Pergola, N., Tramutoli, V., 2008. Robust tir satellite techniques for monitoring earthquake active regions: limits, main achievements and perspectives. Annals of Geophysics 51. URL: http://www.annalsofgeophysics.eu/index.php/annals/article/view/3050.

Blackett, M.e.a., 2011. Exploring land surface temperature earthquake precursors: A focus on the gujarat (india) earthquake of 2001. Geophysical Research Letters 38.

Buzzo, G., 2012. Earthquake precursors detected via hypertemporal satellite thermal data. Technical Report. ITC.

Chandola, V., Banerjee, A., Kumar, V., 2009. Anomaly detection: A survey. ACM Comput. Surv. 41, 1–58.

Chawla, C., Sun, P., 2006. Outlier detection: Principles, techniques and applications.

Cheboli, D., 2010. Anomaly detection of time series. Master's thesis. University of Minessota. URL: http://purl.umn.edu/92985.

Cleveland, R.e.a., 1990. Stl: A seasonal-trend decomposition procedure based on loess. Journal of Official Statistics 6, 3–73.

Cowpertwait, P., Metcalfe, A., 2009. Introductory time Series with R. Use R!, Springer, New York.

EUMETSAT, 2012. Official website of the european organisation for the exploitation of meteorological satellites. URL: http://www.eumetsat.int/Home/index.htm/.

Fox, A., 1972. Outliers in time series. J. Royal Statis. Soc. Series B 34, 350–363.

Freund, F., 2002. Charge generation and propagation in igneous rocks. Journal of Geodynamics 33, 543–570.

Freund, F., 2003. Rocks that crackle and sparkle and glow: Strange pre-earthquake phenomena. Journal of Scientific Exploration 17, 61–64.

Freund, F., 2011. Pre-earthquake signals: Underlying physical processes. Journal of Asian Earth Sciences 41, 383–400.

Freund, F.T., Takeuchi, A., Lau, B.W.S., Al-Manaseer, A., Fu, C.C., Bryant, N.A., Ouzounov, D., 2007. Stimulated infrared emission from rocks: assessing a stress indicator. eEarth 2, 7–16. URL: http://www.electronic-earth.net/2/7/2007/, doi:10.5194/ee-2-7-2007.

Genzano, N., Aliano, C., Filizzola, C., Pergola, N., Tramutoli, V., 2007. A robust satellite technique for monitoring seismically active areas: The case of bhujï£¡gujarat earthquake. Tectonophysics 431, 197–210.

Hawkins, D., 1980. Identification of Outliers. Chapman and Hall.

Huguet, C.e.a., 2012. A seasonal cycle of terrestrial inputs in lake van, turkey. Environm.Sci.Pollut.Res. .

ISRO, 2012. Official website of the indian space research organization. URL: http://www.isro.org/.

Khan, N., 2010. Thermal infrared time series analysis for earthquake precursory detection. Master's thesis. University of Twente Faculty of Geo-Information and Earth Observation ITC.

Nicola Pergola, Francesco Marchese, V.T., 2004. Automated detection of thermal features of active volcanoes by means of infrared avhrr records. Remote Sensing of Environment 93, 311 – 327.

NOAA, 2012. Official website of the national oceanic and atmospheric administration. URL: http://www.noaa.gov/.

Ouzounov, D., Freund, F., 2004. Mid-infrared emission prior to strong earthquakes analyzed by remote sensing data. Advances in Space Research 33, 268–273.

P.Hafen, R., 2010. Local regression models: Advancements, applications, and new methods. Ph.D. thesis. Purdue University.

Prime Ministry, D., Emergency Management Presidency, E.D.o.T., 2011. Report on Van earthquake (Eastern Turkey). Technical Report. AFAD.

Pulinets, S., Ouzounov, D., Karelin, A., Boyarchuk, K., Pokhmelnykh, L., 2006. The physical nature of thermal anomalies observed before strong earthquakes. Physics and Chemistry of the Earth 31, 143 – 153.

R Core Team, 2012. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. URL: http://www.R-project.org.

Rossiter, D., 2012. Tutorial: Time series in r.

Rousseeuw, P., Leroy, A., 1987. Robust Regression and Outlier Detection. John Wiley & Sons, Inc.

Sabins, F., 1997. Remote sensing : principles and interpretation. Third edition ed., W.H. Freeman, New York.

Saraf, A., Rawat, V., Banerjee, P., Choudhury, S., Panda, S., Dasgupta, S., Das, J., 2008. Satellite detection of earthquake thermal infrared precursors in iran. Natural Hazards 47, 119–135.

Saraf, A., Rawat, V., Das, J., Zia, M., Sharma, K., 2012. Satellite detection of thermal precursors of yamnotri, ravar and dalbandin earthquakes. Natural Hazards 61, 861–872.

Tiampo, K.F., Shcherbakov, R., 2012. Seismicity-based earthquake forecasting techniques: Ten years of progress. Tectonophysics 522-523, 89–121.

Tjemkes, S., 2005. On the Conversion from Radiances to Equivalent Brightness Temperatures. URL: http://www.eumetsat.int/groups/ops/documents/document/pdf_msg_seviri_rad2bright.pdf.

Tramutoli, V., Cuomo, V., Filizzola, C., Pergola, N., Pietrapertosa, C., 2005. Assessing the potential of thermal infrared satellite surveys for monitoring seismically active areas: The case of kocaeli (izmit) earthquake, august 17, 1999. Remote Sensing of Environment 96, 409 – 426.

Tramutoli, V., Di Bello, G., Pergola, N., Piscitelli, S., 2001. Robust satellite techniques for remote sensing of seismically active areas.

Tronin, A., 2009a. Satellite remote sensing in seismology. a review. Remote Sensing 2, 124–150.

Tronin, A.A., 2006. Remote sensing and earthquakes: A review. Physics and Chemistry of the Earth, Parts A/B/C 31, 138–142.

Tronin, A.A., 2009b. Satellite remote sensing in seismology. a review. Remote Sensing 2, 124–150. URL: http://www.mdpi.com/2072-4292/2/1/124, doi:10.3390/rs2010124.

Tsay, R.e.a., 2000. Outliers in multi-variate time series. Biometrika 87, 789–804.

USGS, 2012. Official website of the us geological survey. URL: http://www.usgs.gov/.

Wan, Z., 2008. New refinements and validation of the modis land-surface temperature/emissivity products. Remote Sensing of Environment 112, 59–74.

Wan, Z., Dozier, J., 1996. A generalized split-window algorithm for retrieving land-surface temperature from space. Geoscience and Remote Sensing, IEEE Transactions on 34, 892 –905. doi:10.1109/36.508406.

Wick, L.e.a., 2003. Evidence of lateglacial and holocene climatic change and human impact in eastern anatolia: high-resolution pollen, charcoal, isotopic and geochemical records from the laminated sediments of lake van, turkey. The Holocene 13, 665–675.

Yao, Q.L., Qiang, Z.J., 2012. Thermal infrared anomalies as a precursor of strong earthquakes in the distant future. Natural Hazards , 1–13.

Zahradnik, J., Sokos, E., 2011. Multiple-point source solution of the Mw7.2 Van earthquake, October 23, 2011, Eastern Turkey. Technical Report. EMSC.

Zhengming, W., Zhao-Liang, L., 1997. A physics-based algorithm for retrieving land-surface emissivity and temperature from eos/modis data. Geoscience and Remote Sensing, IEEE Transactions on 35, 980–996.

Ziqi, G., ShuQing, Q., Chao, W., Zhi, L., Xing, G., Weiguo, Z., Yong, Y., Hong, Z., Jishuang, Q., 2002. The mechanism of earthquake's thermal infrared radiation precursory on remote sensing images, in: Geoscience and Remote Sensing Symposium, 2002. IGARSS '02. 2002 IEEE International, pp. 2036 – 2038.

Appendix A
IDL and R Scripts

Importing program, including blank images in the place of missing values. Written on the basis of the previous importing program by Harald van der Werf.

```
pro complete_import, ev
  compile_opt idl2

  title = 'complete_import'

  ; select meteosat tiff files
  file_dir=dialog_pickfile(/read,/directory,/must_exist)
  file_names=file_search(file_dir,'*.tif')

  ;check error
  if (file_names[0]) eq '' then begin
  print, 'error filename'
  return
  endif

  ; query first input file
  ok = query_tiff(file_names[0],tif_info,geotiff=geotiff)
  ;help, geotiff
  if (ok) eq 0 then begin
  print, 'error tiff'
  return
  endif

    ; define data dimensions
  ns = tif_info.dimensions[0]
  nl = tif_info.dimensions[1]
  nel = n_elements(file_names)


  ; create array for bandnames (date strings)
  b_names = strarr(nel)

  ;create arrays for all years and their day numbers, months and their day
        ;numbers, days, hours, minutes
  all_years = lonarr(nel)
  yr_days=lonarr(nel)
```

```
  mon = lonarr(nel)
  m_days=lonarr(nel)
  dd = lonarr(nel)
  hh = lonarr(nel)
  minutes = lonarr(nel)

  ;(A)create a b_names array with elements=names of files i, only as date,
        ;imported either from mfg or from msg
  for i = 0, nel-1 do begin
    file_name = file_names[i]
    ; add filename datestring to bandnames array
      file_name = file_basename(file_name,'.tif')
        ;if IR108 exists in whichever file, then make bandname array
                                  following mfg procedure, else msg
        image_source='IR108'
  if STRCMP(image_source, file_name , 5 ) then begin
 b_names[i] = strmid(file_name,strpos(file_name,'-',/reverse_search)+1)
 hourly_images=2
 endif else begin
 b_names[i] = (strsplit(file_name,'_',/extract))[0]+'00'
 hourly_images=4
 endelse
 endfor
        ;sort the resulting b_names array
        b_names=b_names[sort(b_names)]


;(B)split the name further to extract year, month, day, hour, minutes sorted arra
;create sorted array of unique years
          for j=0, nel-1 do begin
          b_name=b_names[j]
          ;extract year string
          year_string=strmid(b_name, 0, 4)
          ;convert year string to long and store it in all_years array
          all_years[j]= long(year_string)
          ;create all months array
          mon[j]=long(strmid(b_name, 4, 2))
          ;create all days array
          dd[j]=long(strmid(b_name, 6, 2))
          ;create all hours array
          hh[j]=long(strmid(b_name, 8, 2))
          ;create all_months array
          minutes[j]=long(strmid(b_name, 10, 2))
        endfor

   ;create array holding sorted, unique values of years
   years= all_years[uniq(all_years)]
```

```
;(C) calculate nr of days in year and total expected images in the batch
;WARNING: the expected nr of images is calculated to make
;COMPLETE YEAR BATCHES.
;this program won't just fill the gaps from the first to the last image
      nr_images=0
      a= where (years ne 0, count)
       for k=0, count-1 do begin
        days_in_year=365

 if ((years[k] mod 4 eq 0) && (years[k] mod 100 ne 0))
then days_in_year=366
if ((years[k] mod 4 eq 0) && (years[k] mod 100 eq 0)
&& (years[k] mod 400 eq 0)
then days_in_year=366

;calculate nr of images expected, depending on image source
nr_images=nr_images+days_in_year*24*hourly_images
endfor

print, 'nr_images=', nr_images

 ;(D) create input array, all values NaNs, seperate array for names
  data_array = make_array(ns,nl,nr_images, /float, value= !values.f_nan)

;(E) fill in the input array with the existing images where they exist

  ;calculate the name components of the first expected image
     expected_min=00
     expected_hr=00
     expected_dd=01
     expected_mm=01
     expected_yr=years[0]


;creat bname NaN string array to hold the names of the existing bands
;in the place where they're written
bnames= make_array(nr_images, /string, value= string(0))

;position counter
w=0

;image loop
for y=0, nel-1 do begin

jump: print, 'next position=', w

if (w le nr_images-1) then begin
```

```
 print, expected_yr, expected_mm, expected_dd, expected_hr, expected_min

;calculate differences
dmin= minutes[y]-expected_min
dhrs= hh[y]-expected_hr
ddays= dd[y]-expected_dd
dmon= mon[y]-expected_mm
dyr= all_years[y]-expected_yr

print, dyr,dmon,ddays, dhrs, dmin

;calculate d=maximum days of expected month for subsequent calculations
     case expected_mm of
      01:d=31
      02:d=28
      03:d=31
      04:d=30
      05:d=31
      06:d=30
      07:d=31
      08:d=31
      09:d=30
      10:d=31
      11:d=30
      12:d=31
    endcase

days_of_exp_year=365
if ((expected_yr mod 4 eq 0) && (expected_yr mod 100 ne 0))
then days_of_exp_year=366
if ((expected_yr mod 4 eq 0) && (expected_yr mod 100 eq 0) &&
(expected_yr mod 400 eq 0))
then days_of_exp_year=366

if ((days_of_exp_year eq 366) && (expected_mm eq 02)) then d=29



;calculate subsequent expected image name components

if ((hourly_images eq 2) && (expected_min eq 00))
then expected_min=30 else begin
if ((hourly_images eq 2) && (expected_min ne 00)) then begin
expected_min= 00
 if (expected_hr ne 23) then (expected_hr=expected_hr+1) else begin
    expected_hr= 00
       if (expected_dd ne d) then (expected_dd=expected_dd+1) else begin
```

```
        expected_dd=1
        if (expected_mm ne 12) then expected_mm=expected_mm+1 else begin
        expected_mm=1
         expected_yr= expected_yr+1
          endelse
         endelse
       endelse
     endif
  endelse

if (hourly_images eq 4) && (expected_min ne 45) then
 expected_min=expected_min+15 else begin
   if ((hourly_images eq 4) && (expected_min eq 45)) then begin
     expected_min= 00
       if (expected_hr ne 23) then expected_hr=expected_hr+1 else begin
         expected_hr= 00
           if (expected_dd ne d) then expected_dd=expected_dd+1 else begin
            expected_dd=1
            if (expected_mm ne 12) then expected_mm=expected_mm+1 else begin
            expected_mm=1
            expected_yr= expected_yr+1
            endelse
         endelse
       endelse
     endif
   endelse


 print, expected_yr, expected_mm, expected_dd, expected_hr, expected_min


;if image y matches the expected then write it in position w of stack
;and go to check another image (exit position w for loop)
;else repeat loop (let w=w+1) calculating subsequent expected
;till you find the position w
;where image y components match the expected ones

if ((dyr eq 0) && (dmon eq 0) &&(ddays eq 0) && (dhrs eq 0) &&
(dmin eq 0)) eq 0 then begin
w=w+1
goto, jump
endif

if ((dyr eq 0) && (dmon eq 0) &&(ddays eq 0) && (dhrs eq 0) &&
(dmin eq 0)) eq 1 then begin
data_array[*,*,w]=read_tiff(file_names[y])
bnames[w]=b_names[y]
w=w+1
```

```
endif

endif

endfor  ; back to y


print, 'loading done'


; ;(F) if mfg perform the calibration

if hourly_images eq 2 then begin

  ;first make array of only existing images
  data=fltarr(ns, nl, nel)


  ;write the array of existing images
f=0
for e=0, nr_images-1 do begin
    if f le (nel-1) then begin
      if (bnames[e] ne 0) then begin
      data[*,*,f]=data_array[*,*,e]
      f=f+1
      endif
    endif
endfor; back to e, i.e. change stack position


  ;calibrate the array of existing images
    ; get meteosat 5 calibration data
    org_calibration = hw_shaky_caldata_mfg()

    ; create empty array for calibration data
    ; to be resampled to selected dates
    res_calibration = fltarr(2,nel)

    imgdates = float(strmid(b_names,0,12))
    caldates = org_calibration[0,*]

    starting = where(caldates ge imgdates[0],starting_exist)
    if starting_exist ge 0 && starting[0] ne 0 then
                starting =
                starting[0]-1 else starting = 0
    ending   = where(caldates le imgdates[nel-1],ending_exist)
    if ending_exist ge 0
                then ending = (reverse(ending))[0]+1
```

```
                else ending = n_elements(caldates)-1

    ; loop through calibration dates
    for i = starting, ending-1 do begin
      ; check where image data is greater
                        or equal to a calibration date
      index = where(imgdates ge caldates[i],count)

      ; in case a match is found...
      if (count gt 0) then begin
        ; ...add the calibration date to the
                                resampled calibration array
        res_calibration[*,index] =
                                rebin(org_calibration[[1,2],i],2,count)
      endif
    endfor

    print, 'resampling done'

; convert to radiance
data = data -
transpose(rebin(reform(res_calibration[1,*]),nel,ns,nl),[1,2,0])
data = data *
transpose(rebin(reform(res_calibration[0,*]),nel,ns,nl),[1,2,0])

    print, 'calibration done'

      ; convert to brightness temperature
      A =  6.7348
      B = -1272.2
      ;BT = B / (alog(radiance) - A)

      data = alog(data)
      data = B / (data - A)

      print, 'conversion done'

  ;return the calibrated images in their correct place in the stack

v=0 ;image counter

; z position in data_array loop
  for z=0, nr_images-1 do begin
   if v le (nel-1) then begin
    if b_names[v] eq bnames[z] then begin
    data_array[*,*,z]=data[*,*,v]
    v=v+1 ;check next image of small stack
    endif
```

```
    endif
  endfor ;back to z, i.e. check next big stack position


endif


;write the output file, whether mfg or msg
;remember to change the names :-P
  openw,unit, 'bam_total', /get_lun
  writeu, unit, data_array
  free_lun, unit

  map_info = 'None'
  envi_setup_head,
    fname = 'bam_total',
    descrip = descrip,
    ns = tif_info.dimensions[0],
    nl = tif_info.dimensions[1],
    nb = nr_images,
    bbl = ok,
    data_type = 4,
    interleave = 0,
    bnames = bnames,
    wavelength_units = 5,
    map_info = map_info,
    /write

  print, 'import complete'

  end
```

Program for imposing artificial anomalies

```
pro fake_shake_new, ev

compile_opt idl2

title= 'fake_shake_new'

;Definitions
filter = '*.tif'
bsq_filename = dialog_pickfile(/read,/must_exist)
mask_filename = dialog_pickfile(/read,/must_exist)
anomaly_start='20111013000000'
anomaly_finish='20111015000000'


; OPEN BSQ file
envi_open_file, bsq_filename, /no_realize, r_fid = bsq_fid
```

```
envi_file_query, bsq_fid, dims = dims1, data_type=data_type,
file_type = file_type, interleave = interleave,
NB = NB, NL=NL, NS=NS, spec_names=spec_names,WL=WL,
FNAME=FNAME, WAVELENGTH_UNITS=wavelength_units,
REFLECTANCE_SCALE_FACTOR=reflectance_scale_factor,
FWHM=fwhm,BNAMES=bnames, descrip=descrip, bbl=bbl
;print, BNAMES[5]
;read the stack in
bsq = fltarr(ns,nl,nb)
for j = 0, nb-1 do begin
; apply to all bands and concatenate
image_band = ENVI_get_data(fid = bsq_fid, dims  = dims1, POS = j)
bsq[*,*,j]=image_band
endfor

;print, bsq [1,1,3]

; Open mask
envi_open_file, mask_filename, r_fid = mask_fid
envi_file_query, mask_fid, dims = dims2
mask = ENVI_get_data(fid = mask_fid, dims  = dims2, POS = 0)


;find anomaly range in band names
anomaly_brange = lonarr(2)
for i = 0, NB-1 do begin
 if (strmatch(bnames[i],anomaly_start))then anomaly_brange[0] = i
 if (strmatch(bnames[i],anomaly_finish))then anomaly_brange[1] = i
endfor;
print, anomaly_brange[0], anomaly_brange[1]

;find pixel serial numbers of the mask, only one dimension
result = where(mask eq 1)
print, result
;increase the values of all masked pixels,
;within the bands of the anomaly range
for i=(anomaly_brange[0]), anomaly_brange[1] do begin
serial_nr=result+((ns*nl)*i)
;calculate the serial numbers of all masked pixels in all bands
;print, serial_nr
;print, anomaly_slice[serial_nr]
bsq[serial_nr]=bsq[serial_nr]+4
;increase masked image values by 4 degrees
print, bsq[serial_nr]
endfor

;output
 openw, unit,'VanEQ4an.bsq', /get_lun
```

```
  writeu, unit, bsq
  free_lun, unit

  descrip  = 'imposed_anomaly_of_4degrees'
    map_info = 'None'
    envi_setup_head,
    fname = 'VanEQ4an.bsq',
    descrip = descrip,
    ns = ns,
    nl = nl,
    nb = nb,
    data_type = 4,
    interleave = 0,
    bnames = bnames,
    wavelength_units = 5,
    map_info = map_info,
    /write


close,/all

ENVI_FILE_MNG, ID=mask_fid,     /REMOVE
ENVI_FILE_MNG, ID=bsq_fid,     /REMOVE
end
```

Program for normalization, adapted by existing script of Harald van der Werf

```
pro kernel_calc, ev

  compile_opt idl2

  title = 'Kernel_calc'

  ;; INPUT

  envi_open_file,/no_realize,r_fid=fid
  if fid eq -1 then return

  ; query input image
  envi_file_query,fid,dims=dims,ns=ns,nl=nl,nb=nb,
    bnames=bnames,data_type = data_type, fname=fname
  pos = lindgen(nb)
  ns = dims[2]-dims[1]+1
  nl = dims[4]-dims[3]+1

  ; create input array
  maskedfile = fltarr(nl,nb,ns)

  ; loop through image lines
for l= 0,nl-1 do begin
```

```
    ; get a line out of the image
    ; and add to input array
maskedfile[l,*,*] = envi_get_slice(fid=fid,line=l,pos=pos,/bip)
endfor

print, 'loaded data'

   ; close input file
   envi_file_mng, id=fid, /remove

   ; transpose input data to bsq [ns,nl,nb]
  maskedfile = transpose(maskedfile,[2,0,1])

  ; create input array
  testarray = fltarr(ns,nl)

  ; loop to load images, and fill data arrays
  for b = 0, nb-1 do begin


      testarray[*,*] = (maskedfile[*,*,b]

;set kernelsize
ks= 10


;calculate expected max nr of pixels in the kernel
;kernel_pixels=4*(2*kernelsize+1)-4


kernelvalue=0
NaN_count=0
kp_count=0 ;number of existing kernel pixels

;create output file to store normalized pixel values
output=fltarr(ns, nl)

for cl=0, nl-1 do begin
for cs=0, ns-1 do begin

;general kernel scheme
if ((cs ge ks) && (cs le (ns-1-ks)) && (cl ge ks)
&& (cl le (nl-1-ks))) then begin

;left side
for j= cl-ks, cl+ks do begin

    if testarray[cs-ks, j] eq testarray[cs-ks, j] then begin
```

```
      kp_count=kp_count+1
      kernelvalue=kernelvalue+testarray[cs-ks, j]
      endif else begin
         NaN_count=NaN_count+1
         testarray[cs-ks, j]=0
         kernelvalue=kernelvalue+testarray[cs-ks, j]
         testarray[cs-ks, j]=!values.f_nan
         endelse

;right side

     if testarray[cs+ks, j] eq testarray[cs+ks, j] then begin
      kp_count=kp_count+1
      kernelvalue=kernelvalue+testarray[cs+ks, j]
      endif else begin
         NaN_count=NaN_count+1
         testarray[cs+ks, j]=0
         kernelvalue=kernelvalue+testarray[cs+ks, j]
         testarray[cs+ks, j]=!values.f_nan
         endelse

endfor

;upper part
for i= cs-(ks-1), cs+(ks-1) do begin

     if testarray[i, cl-ks] eq testarray[i, cl-ks] then begin
      kp_count=kp_count+1
      kernelvalue=kernelvalue+testarray[i, cl-ks]
      endif else begin
         NaN_count=NaN_count+1
         testarray[i, cl-ks]=0
         kernelvalue=kernelvalue+testarray[i, cl-ks]
         testarray[i, cl-ks]=!values.f_nan
         endelse


;lower part

     if testarray[i, cl+ks] eq testarray[i, cl+ks] then begin
      kp_count=kp_count+1
      kernelvalue=kernelvalue+testarray[i, cl+ks]
      endif else begin
         NaN_count=NaN_count+1
         testarray[i, cl+ks]=0
         kernelvalue=kernelvalue+testarray[i, cl+ks]
         testarray[i, cl+ks]=!values.f_nan
         endelse
```

```
endfor

if kp_count eq 0 then kernel= !values.f_nan
else kernel= kernelvalue/kp_count
if kernel eq 0 then output[cs,cl]= !values.f_nan
else output[cs,cl]= testarray[cs,cl]/kernel

kernelvalue=0
NaN_count=0
kp_count=0


endif

;left middle part of margin
if ((cs lt ks) && (cl ge ks) && (cl le (nl-1-ks))) then begin

;right side
for j= cl-ks, cl+ks do begin

    if testarray[cs+ks, j] eq testarray[cs+ks, j] then begin
     kp_count=kp_count+1
     kernelvalue=kernelvalue+testarray[cs+ks, j]
     endif else begin
        NaN_count=NaN_count+1
        testarray[cs+ks, j]=0
        kernelvalue=kernelvalue+testarray[cs+ks, j]
        testarray[cs+ks, j]=!values.f_nan
        endelse

endfor

;upper part
for i= 0, cs+(ks-1) do begin

    if testarray[i, cl-ks] eq testarray[i, cl-ks] then begin
     kp_count=kp_count+1
     kernelvalue=kernelvalue+testarray[i, cl-ks]
     endif else begin
        NaN_count=NaN_count+1
        testarray[i, cl-ks]=0
        kernelvalue=kernelvalue+testarray[i, cl-ks]
        testarray[i, cl-ks]=!values.f_nan
        endelse


;lower part
```

```
    if testarray[i, cl+ks] eq testarray[i, cl+ks] then begin
     kp_count=kp_count+1
     kernelvalue=kernelvalue+testarray[i, cl+ks]
     endif else begin
        NaN_count=NaN_count+1
        testarray[i, cl+ks]=0
        kernelvalue=kernelvalue+testarray[i, cl+ks]
        testarray[i, cl+ks]=!values.f_nan
        endelse
endfor

if kp_count eq 0 then kernel= !values.f_nan
else kernel= kernelvalue/kp_count
if kernel eq 0 then output[cs,cl]= !values.f_nan
else output[cs,cl]= testarray[cs,cl]/kernel

kernelvalue=0
NaN_count=0
kp_count=0

endif

; right middle margin
if ((cs gt (ns-1-ks)) && (cl ge ks) && (cl le (nl-1-ks))) then begin

;left side
for j= cl-ks, cl+ks do begin

    if testarray[cs-ks, j] eq testarray[cs-ks, j] then begin
     kp_count=kp_count+1
     kernelvalue=kernelvalue+testarray[cs-ks, j]
     endif else begin
        NaN_count=NaN_count+1
        testarray[cs-ks, j]=0
        kernelvalue=kernelvalue+testarray[cs-ks, j]
        testarray[cs-ks, j]=!values.f_nan
        endelse

endfor

;upper part
for i= cs-(ks-1), ns-1 do begin

    if testarray[i, cl-ks] eq testarray[i, cl-ks] then begin
     kp_count=kp_count+1
     kernelvalue=kernelvalue+testarray[i, cl-ks]
     endif else begin
        NaN_count=NaN_count+1
```

```
        testarray[i, cl-ks]=0
        kernelvalue=kernelvalue+testarray[i, cl-ks]
        testarray[i, cl-ks]=!values.f_nan
        endelse


;lower part

    if testarray[i, cl+ks] eq testarray[i, cl+ks] then begin
     kp_count=kp_count+1
     kernelvalue=kernelvalue+testarray[i, cl+ks]
     endif else begin
        NaN_count=NaN_count+1
        testarray[i, cl+ks]=0
        kernelvalue=kernelvalue+testarray[i, cl+ks]
        testarray[i, cl+ks]=!values.f_nan
        endelse
endfor

if kp_count eq 0 then kernel= !values.f_nan
else kernel= kernelvalue/kp_count
if kernel eq 0 then output[cs,cl]= !values.f_nan
else output[cs,cl]= testarray[cs,cl]/kernel

kernelvalue=0
NaN_count=0
kp_count=0

endif

;upper middle margin
if ((cs ge ks) && (cs le (ns-1-ks)) && (cl lt ks)) then begin

;left side
for j= 0, cl+ks do begin

    if testarray[cs-ks, j] eq testarray[cs-ks, j] then begin
     kp_count=kp_count+1
     kernelvalue=kernelvalue+testarray[cs-ks, j]
     endif else begin
        NaN_count=NaN_count+1
        testarray[cs-ks, j]=0
        kernelvalue=kernelvalue+testarray[cs-ks, j]
        testarray[cs-ks, j]=!values.f_nan
        endelse

;right side
```

```
    if testarray[cs+ks, j] eq testarray[cs+ks, j] then begin
     kp_count=kp_count+1
     kernelvalue=kernelvalue+testarray[cs+ks, j]
     endif else begin
        NaN_count=NaN_count+1
        testarray[cs+ks, j]=0
        kernelvalue=kernelvalue+testarray[cs+ks, j]
        testarray[cs+ks, j]=!values.f_nan
        endelse

endfor

;lower part
for i= cs-(ks-1), cs+(ks-1) do begin

    if testarray[i, cl+ks] eq testarray[i, cl+ks] then begin
     kp_count=kp_count+1
     kernelvalue=kernelvalue+testarray[i, cl+ks]
     endif else begin
        NaN_count=NaN_count+1
        testarray[i, cl+ks]=0
        kernelvalue=kernelvalue+testarray[i, cl+ks]
        testarray[i, cl+ks]=!values.f_nan
        endelse
endfor

if kp_count eq 0 then kernel= !values.f_nan
else kernel= kernelvalue/kp_count
if kernel eq 0 then output[cs,cl]= !values.f_nan
else output[cs,cl]= testarray[cs,cl]/kernel


kernelvalue=0
NaN_count=0
kp_count=0

endif

;lower middle margin
if ((cs ge ks) && (cs le (ns-1-ks)) && (cl gt (nl-1-ks))) then begin

;left side
for j= cl-(ks-1), nl-1 do begin

    if testarray[cs-ks, j] eq testarray[cs-ks, j] then begin
     kp_count=kp_count+1
     kernelvalue=kernelvalue+testarray[cs-ks, j]
     endif else begin
```

```
        NaN_count=NaN_count+1
        testarray[cs-ks, j]=0
        kernelvalue=kernelvalue+testarray[cs-ks, j]
        testarray[cs-ks, j]=!values.f_nan
        endelse

;right side

    if testarray[cs+ks, j] eq testarray[cs+ks, j] then begin
     kp_count=kp_count+1
     kernelvalue=kernelvalue+testarray[cs+ks, j]
     endif else begin
        NaN_count=NaN_count+1
        testarray[cs+ks, j]=0
        kernelvalue=kernelvalue+testarray[cs+ks, j]
        testarray[cs+ks, j]=!values.f_nan
        endelse

endfor

;upper part
for i= cs-ks, cs+ks do begin

    if testarray[i, cl-ks] eq testarray[i, cl-ks] then begin
     kp_count=kp_count+1
     kernelvalue=kernelvalue+testarray[i, cl-ks]
     endif else begin
        NaN_count=NaN_count+1
        testarray[i, cl-ks]=0
        kernelvalue=kernelvalue+testarray[i, cl-ks]
        testarray[i, cl-ks]=!values.f_nan
        endelse

endfor

if kp_count eq 0 then kernel= !values.f_nan
else kernel= kernelvalue/kp_count
if kernel eq 0 then output[cs,cl]= !values.f_nan
else output[cs,cl]= testarray[cs,cl]/kernel

kernelvalue=0
NaN_count=0
kp_count=0

endif

; up left corner
if ((cs lt ks) && (cl lt ks)) then begin
```

```
;right side
for j= 0, cl+ks do begin

    if testarray[cs+ks, j] eq testarray[cs+ks, j] then begin
     kp_count=kp_count+1
     kernelvalue=kernelvalue+testarray[cs+ks, j]
     endif else begin
        NaN_count=NaN_count+1
        testarray[cs+ks, j]=0
        kernelvalue=kernelvalue+testarray[cs+ks, j]
        testarray[cs+ks, j]=!values.f_nan
        endelse

endfor

;lower part
for i= 0, cs+(ks-1) do begin

    if testarray[i, cl+ks] eq testarray[i, cl+ks] then begin
     kp_count=kp_count+1
     kernelvalue=kernelvalue+testarray[i, cl+ks]
     endif else begin
        NaN_count=NaN_count+1
        testarray[i, cl+ks]=0
        kernelvalue=kernelvalue+testarray[i, cl+ks]
        testarray[i, cl+ks]=!values.f_nan
        endelse
endfor

if kp_count eq 0 then kernel= !values.f_nan
else kernel= kernelvalue/kp_count
if kernel eq 0 then output[cs,cl]= !values.f_nan
else output[cs,cl]= testarray[cs,cl]/kernel

kernelvalue=0
NaN_count=0
kp_count=0

endif

;low left corner
if ((cs lt ks) && (cl gt (nl-1-ks))) then begin

;right side
for j= cl-ks, nl-1 do begin

    if testarray[cs+ks, j] eq testarray[cs+ks, j] then begin
```

```
      kp_count=kp_count+1
      kernelvalue=kernelvalue+testarray[cs+ks, j]
      endif else begin
         NaN_count=NaN_count+1
         testarray[cs+ks, j]=0
         kernelvalue=kernelvalue+testarray[cs+ks, j]
         testarray[cs+ks, j]=!values.f_nan
         endelse

endfor

;upper part
for i= 0, cs+(ks-1) do begin

     if testarray[i, cl-ks] eq testarray[i, cl-ks] then begin
      kp_count=kp_count+1
      kernelvalue=kernelvalue+testarray[i, cl-ks]
      endif else begin
         NaN_count=NaN_count+1
         testarray[i, cl-ks]=0
         kernelvalue=kernelvalue+testarray[i, cl-ks]
         testarray[i, cl-ks]=!values.f_nan
         endelse

endfor

if kp_count eq 0 then kernel= !values.f_nan
else kernel= kernelvalue/kp_count
if kernel eq 0 then output[cs,cl]= !values.f_nan
else output[cs,cl]= testarray[cs,cl]/kernel

kernelvalue=0
NaN_count=0
kp_count=0

endif

;upper right corner
if ((cs gt (ns-1-ks)) && (cl lt ks)) then begin

;left side
for j= 0, cl+ks do begin

     if testarray[cs-ks, j] eq testarray[cs-ks, j] then begin
      kp_count=kp_count+1
      kernelvalue=kernelvalue+testarray[cs-ks, j]
      endif else begin
         NaN_count=NaN_count+1
```

```
        testarray[cs-ks, j]=0
        kernelvalue=kernelvalue+testarray[cs-ks, j]
        testarray[cs-ks, j]=!values.f_nan
        endelse
endfor

;lower part
for i= cs-(ks-1),ns-1 do begin

     if testarray[i, cl+ks] eq testarray[i, cl+ks] then begin
     kp_count=kp_count+1
     kernelvalue=kernelvalue+testarray[i, cl+ks]
     endif else begin
        NaN_count=NaN_count+1
        testarray[i, cl+ks]=0
        kernelvalue=kernelvalue+testarray[i, cl+ks]
        testarray[i, cl+ks]=!values.f_nan
        endelse

endfor

if kp_count eq 0 then kernel= !values.f_nan
else kernel= kernelvalue/kp_count
if kernel eq 0 then output[cs,cl]= !values.f_nan
else output[cs,cl]= testarray[cs,cl]/kernel

kernelvalue=0
NaN_count=0
kp_count=0

endif

;low right corner
if ((cs gt (ns-1-ks)) && (cl gt (nl-1-ks))) then begin

;left side
for j= cl-ks, nl-1 do begin

    if testarray[cs-ks, j] eq testarray[cs-ks, j] then begin
     kp_count=kp_count+1
     kernelvalue=kernelvalue+testarray[cs-ks, j]
     endif else begin
        NaN_count=NaN_count+1
        testarray[cs-ks, j]=0
        kernelvalue=kernelvalue+testarray[cs-ks, j]
        testarray[cs-ks, j]=!values.f_nan
        endelse
```

```
endfor

;upper part
for i= cs-(ks-1), ns-1 do begin

    if testarray[i, cl-ks] eq testarray[i, cl-ks] then begin
     kp_count=kp_count+1
     kernelvalue=kernelvalue+testarray[i, cl-ks]
     endif else begin
        NaN_count=NaN_count+1
        testarray[i, cl-ks]=0
        kernelvalue=kernelvalue+testarray[i, cl-ks]
        testarray[i, cl-ks]=!values.f_nan
        endelse

endfor

if kp_count eq 0 then kernel= !values.f_nan
else kernel= kernelvalue/kp_count
if kernel eq 0 then output[cs,cl]= !values.f_nan
else output[cs,cl]= testarray[cs,cl]/kernel

kernelvalue=0
NaN_count=0
kp_count=0


endif

endfor
endfor

maskedfile[*,*,b]=output[*,*]

endfor

;; OUTPUT

  ; output to regular file
  openw,unit, 'Van10else.bsq' , /get_lun
  writeu, unit, maskedfile
  free_lun, unit

  descrip = 'test stack normalized data'
  map_info = 'None'
  envi_setup_head,
    fname = 'Van10else.bsq',
    descrip = descrip,
```

```
    ns = ns,
    nl = nl,
    nb = nb,
    data_type = 4,
    interleave = 0,
    bnames = bnames,
    wavelength_units = 5,
    map_info = map_info,
    /write
print, 'normalization done'
end
```

STL decomposition with the settings that resulted from the analysis

```
>bam32.stl<-stl2(Bam32.ts, t=NULL, 48, s.window=143, s.degree=0,
t.window=73, t.degree=0, fc.window=17536, fc.degree=2, l.window=49,
l.degree=0, inner=1, outer=5)
```

R script to capture values exceeding the anomaly threshold

```
> for (i in 1:105216){
+ if (is.na(rem_b3749norm_robust.ts[i]) |
(rem_b3749norm_robust.ts[i]<threshold_robust)){
+ flags_rob3749[i]<-NA
+ }
+ else{
+ flags_rob3749[i]<-1
+ }
+ }
```

R script to calculate sum of anomalous values in a 7-day window centered in the pixel, when 48 observations/daily

```
> for (j in 169:105048){
+ flagcount_rob3749[j]<-sum(flags_rob3749[(j-168):(j+168)], na.rm=TRUE)
+ }
> for (j in 105048:105216){
+ flagcount_rob3749[j]<-sum(flags_rob3749[(j-168):105216], na.rm=TRUE)
+ }
> for (j in 1:169){
+ flagcount_rob3749[j]<-sum(flags_rob3749[1:(j+168)], na.rm=TRUE)
+ }
```

R script to produce the mean time series of neighbouring pixels, and calculate their mean:

```
> for (i in 1:105216){
+ if (is.na(clear_rem1020[i])){
+ cl.na_count[i]<-cl.na_count[i]+1
+ }
+ else{
+ cl.p_count[i]<-cl.p_count[i]+1
+ clear_rem[i]<-clear_rem[i]+clear_rem1020[i]
+ }
```

```
+ }

> for (i in 1:105216){
+ if (cl.p_count[i]==0){
+ clear_rem[i]<-NA
+ }
+ else{
+ clear_rem[i]<- clear_rem[i]/cl.p_count[i]
+ }
+ }
```

R script to remove from the raw time series the values which correspond to a remainder lower than the threshold

```
> for (i in 1:105216){
+ if(is.na(rem_b3444.ts[i])){
+ masked.ts[i]<-NA
+ }
+ else{
+ if(rem_b3444.ts[i]<(-2*stdev)){
+ masked.ts[i]<-NA
+ }
+ else{
+ masked.ts[i]<-b3444.ts[i]
+ }
+ }
+ }
```

Appendix B
Additional information about the study areas

Table B.1: Bam, Iran: average temperatures (^{o}C) by month, after the Meteorological Organization of Iran, http://www.irimet.net/

Month	*Average max*	*Average min*	*Absolute max*	*Absolute min*
January	6.7	1.1	20.4	-14
February	11.6	5.4	26	-1.5
March	14.7	6.7	32	6.4
April	19.5	7.4	37	9.6
May	23.8	13.6	42.6	13
June	27.6	17.6	45	19
July	27.8	15.4	46	-12
August	27.8	15.1	44	13
September	23.4	13.8	39.8	15
October	25.4	18.9	37	11
November	17.8	12.1	31	5
December	13.7	8.6	27	-5

Table B.2: Van, Turkey: average temperatures (^{o}C) by month, after the Turkish State Meteorological Service, http://www.dmi.gov.tr/index.aspx

Month	*Average max*	*Average min*	*Absolute max*	*Absolute min*
January	2.1	-8	17.7	-24
February	2.7	-7.2	16.4	-23.4
March	7.3	-2.5	21.8	-18.4
April	13.7	3	26.2	-10
May	18.3	7.2	29	-1.4
June	23.6	11.2	36	0.9
July	27.5	15	38	3.6
August	27.8	15	37	9.2
September	23.6	10.5	33.9	0.8
October	18.3	5.6	30.3	-5
November	10.7	-0.1	24.8	-12.4
December	4.7	-5	22	-22.2

Geographic Grid Search Earthquakes= 39
Latitude: 32.000N - 26.000N
Longitude: 61.000E - 54.000E
Catalog Used: PDE
Date Range: 1999 to 2004
Magnitude Range: 4.0 - 10.0
Depth Range: 0 - 20
Data Selection: Historical & Preliminary Data

CAT	YEAR	MO	DA	ORIG TIME	LAT	LONG	DEP	MAGNITUDE
PDE	2002	04	04	154432.18	27.06	55.24	16	4.8 mbGS
PDE	2003	07	10	170637.66	28.35	54.17	10	5.8 MwHRV
PDE	2003	07	10	174015.90	28.30	54.13	10	5.7 MwHRV
PDE	2003	07	10	192608.24	28.38	54.05	10	4.2 mbGS
PDE	2003	07	10	193706.85	28.29	54.03	10	4.3 mbGS
PDE	2003	07	10	203700.45	28.18	54.02	10	4.0 mbGS
PDE	2003	07	10	220015.64	28.03	54.18	10	4.0 mbGS
PDE	2003	07	10	230243.88	28.28	54.20	10	4.1 mbGS
PDE	2003	07	10	234001.37	28.12	54.28	10	4.1 mbGS
PDE	2003	07	11	031752.09	28.46	54.05	10	4.0 mbGS
PDE	2003	07	11	102332.80	28.22	54.13	10	4.6 mbGS
PDE	2003	07	11	235544.42	28.47	54.04	10	4.9 mbGS
PDE	2003	07	13	011551.60	28.45	54.10	10	4.2 mbGS
PDE	2003	07	13	030011.23	27.72	54.13	10	4.2 mbGS
PDE	2003	07	13	093327.14	28.23	54.08	10	4.3 mbGS
PDE	2003	07	14	042855.63	28.60	54.07	10	4.2 mbGS
PDE	2003	07	20	014030.09	28.27	54.01	10	4.5 mbGS
PDE	2003	08	21	040209.17	29.05	59.77	20	5.9 MwGS
PDE	2003	12	15	225722.67	28.31	54.15	10	5.1 MwHRV
PDE	2003	12	26	015652.44	29.00	58.31	10	6.6 MwHRV
PDE	2003	12	26	030613.64	28.93	58.33	10	5.1 mbGS
PDE	2003	12	26	035325.18	28.79	58.12	10	4.5 mbGS
PDE	2003	12	26	091635.34	29.01	58.28	10	4.0 mbGS
PDE	2003	12	26	095958.98	29.01	58.32	10	4.0 mbGS
PDE	2003	12	26	140816.02	29.08	58.20	10	4.6 mbGS
PDE	2003	12	28	092422.92	29.21	58.55	10	4.1 mbGS
PDE	2003	12	28	150212.01	29.01	58.39	10	4.2 mbGS
PDE	2004	01	11	050604.75	29.19	58.55	10	4.1 mbGS
PDE	2004	01	21	132336.33	29.17	58.24	10	4.3 mbGS
PDE	2004	01	28	172931.57	29.01	58.33	10	4.1 mbGS
PDE	2004	04	08	142304.03	28.97	58.20	10	4.0 mbGS
PDE	2004	05	29	023348.27	29.11	58.35	10	4.2 mbGS
PDE	2004	07	22	045136.01	29.11	58.31	10	4.8 MwHRV
PDE	2004	08	05	162521.27	28.86	57.22	10	4.1 mbGS
PDE	2004	08	28	043210.82	31.14	57.40	10	4.2 mbGS
PDE	2004	10	10	102810.27	26.90	57.54	10	4.6 mbGS
PDE	2004	10	14	012118.49	30.77	56.76	4	4.2 mbGS
PDE	2004	11	13	043538.06	27.85	55.28	10	4.1 mbGS
PDE	2004	12	11	033555.44	27.77	57.57	20	4.6 mbGS

Figure B.1: Search results for Earthquakes M$>$4, Bam 1999-2004, according to NEIC-USGS (part)

Geographic Grid Search Earthquakes= 181
Latitude: 40.000N - 37.000N
Longitude: 45.000E - 42.000E
Catalog Used: PDE
Date Range: 2008 to 2011
Magnitude Range: 4.0 - 10.0
Depth Range: 0 - 20
Data Selection: Historical & Preliminary Data

CAT	YEAR	MO	DA	ORIG TIME	LAT	LONG	DEP	MAGNITUDE
PDE	2008	03	14	050932.90	39.02	44.71	14	4.0 mbGS
PDE	2008	04	03	191950.40	38.53	44.86	18	4.0 mbGS
PDE	2008	04	15	012153.20	37.28	44.94	15	4.0 LgTEH
PDE	2008	05	11	231959.16	37.21	43.35	10	4.0 mbGS
PDE	2008	07	02	194235.20	37.57	43.97	12	4.0 mbGS
PDE	2008	07	26	221650	38.44	43.39	10	4.4 mbGS
PDE	2008	08	01	151622.55	37.62	43.97	5	4.0 MLISK
PDE	2008	10	03	124152.80	38.16	44.67	2	4.3 mbGS
PDE	2010	11	06	010517	38.90	44.28	5	4.1 mbGS
PDE	2011	02	22	090814.59	38.98	42.31	5	4.4 mbGS
PDE	2011	02	22	091144	38.98	42.30	5	4.6 mbGS
PDE	2011	02	22	091606	38.99	42.32	5	4.3 MLISK
PDE	2011	02	22	181552.90	38.93	42.33	4	4.5 mbGS
PDE-W	2011	03	14	185710.91	38.72	44.11	4	4.7 mbGS
PDE-W	2011	05	29	110205.54	37.41	42.58	11	4.5 mbGS
PDE-W	2011	07	10	042259.29	38.56	44.34	5	4.0 mbGS
PDE-W	2011	10	23	104122.93	38.72	43.51	16	7.1 MwWCMT
PDE-W	2011	10	23	104817.42	38.75	43.60	9	5.6 mbGS
PDE-W	2011	10	23	105205	38.76	43.25	7	4.8 mbGS
PDE-W	2011	10	23	105341	38.75	43.20	7	4.4 mbGS
PDE-W	2011	10	23	105649	38.81	43.45	5	5.6 mbGS
PDE-W	2011	10	23	110014.31	38.59	43.52	10	4.6 mbGS
PDE-W	2011	10	23	110029.14	38.60	43.10	10	5.2 mbGS
PDE-W	2011	10	23	110733.72	38.56	43.28	10	4.5 mbGS
PDE-W	2011	10	23	111047.05	38.33	42.61	10	5.0 mbGS
PDE-W	2011	10	23	111626.87	38.74	43.34	10	4.7 mbGS
PDE-W	2011	10	23	111904.51	38.69	43.57	11	4.5 mbGS
PDE-W	2011	10	23	113241.22	38.81	43.30	5	5.7 mbGS
PDE-W	2011	10	23	114423.10	38.48	43.40	10	4.3 mbGS
PDE-W	2011	10	23	120310	38.70	43.14	7	4.7 mbGS
PDE-W	2011	10	23	121307.83	38.79	43.16	5	4.5 mbGS
PDE-W	2011	10	23	122330.06	38.72	43.69	10	4.3 mbGS
PDE-W	2011	10	23	123014	38.75	43.18	5	4.0 mbGS
PDE-W	2011	10	23	124209.69	38.73	43.19	5	4.8 mbGS
PDE-W	2011	10	23	125649.02	38.76	43.34	7	4.5 mbGS
PDE-W	2011	10	23	130651.48	38.69	43.15	5	4.2 mbGS
PDE-W	2011	10	23	131704.47	38.72	43.62	15	4.4 mbGS
PDE-W	2011	10	23	133349.22	38.81	43.23	5	4.1 MLISK
PDE-W	2011	10	23	134609	38.78	43.26	7	4.1 mbGS
PDE-W	2011	10	23	134749	38.78	43.40	8	4.0 mbGS
PDE-W	2011	10	23	141338.77	38.60	43.41	10	4.1 mbGS
PDE-W	2011	10	23	145249	38.71	43.18	5	4.0 mbGS

Figure B.2: Search results for Earthquakes M$>$4, Van 2008-2011, according to NEIC-USGS (part)