

TIME-SERIES ANALYSIS OF REMOTELY-SENSED  
THERMAL INFRA-RED EMISSIONS:  
LINKING ANOMALIES TO PHYSICAL PROCESSES

Efthymia Pavlidou

Graduation committee:

**Chairman/Secretary**

Prof.dr.ir. A. Veldkamp      University of Twente

**Supervisor**

Prof.dr. M. van der Meijde      University of Twente

**Co-supervisors**

Dr. C.A. Hecker      University of Twente  
Dr. H.M.A. van der Werff      University of Twente

**Members**

Prof.dr. F.D. van der Meer      University of Twente  
Prof.dr. V.G. Jetten      University of Twente  
Prof.dr. S.M. de Jong      Utrecht University  
Prof.dr. L. Evers      KNMI, Delft University of Technology  
Dr. M. Gerstenberger      GNS Science, New Zealand

This research was conducted under the auspices of the Graduate School for  
Socio-Economic and Natural Sciences of the Environment (SENSE)

ITC dissertation number 329  
ITC, P.O. Box 217, 7500 AE Enschede, The Netherlands

ISBN 978-90-365-4624-9  
DOI 10.3990/1.9789036546249

Cover designed by Job Duim  
Printed by ITC Printing Department  
Copyright © 2018 by Efthymia Pavlidou



TIME-SERIES ANALYSIS OF REMOTELY-SENSED  
THERMAL INFRA-RED EMISSIONS:  
LINKING ANOMALIES TO PHYSICAL PROCESSES

DISSERTATION

to obtain  
the degree of doctor at the University of Twente,  
on the authority of the rector magnificus,  
prof.dr. T.T.M. Palstra,  
on account of the decision of the graduation committee,  
to be publicly defended  
on Thursday 27<sup>th</sup> September 2018 at 12:45

by

Efthymia Pavlidou

born on 27<sup>th</sup> February 1980

in Alexandroupolis, Greece

This thesis has been approved by  
**Prof.dr. M. van der Meijde**, supervisor  
**Dr. C.A. Hecker**, co-supervisor  
**Dr. H.M.A. van der Werff**, co-supervisor



*Valeu a pena? Tudo vale a pena  
Se a alma não é pequena.  
Quem quer passar além do Bojador  
Tem que passar além da dor.  
Deus ao mar o perigo e o abismo deu,  
Mas nele é que espelhou o céu.*

-Fernando Pessoa



## Acknowledgements

It is very difficult to fit the last few years in a few paragraphs.

I am grateful to the University of Twente, the Twente Graduate School, ITC and the ESA Department for hosting me and this project. I would like to very much thank my promotor, Prof. Dr. Mark van der Meijde, and also my supervisors Dr. Christoff A. Hecker and Dr. Harald M.A. van der Werff, along with their colleagues who unofficially advised and supported me during this work.

I consider myself very lucky to have been welcomed to, and inspired by, meetings with different scientific communities which were relevant to this project. I have especially benefited from the workshops of the International LST and Emissivity Working Group (ILSTE) and from the International Workshops on Statistical Seismology.

The SENSE network, and the PhD communities of ITC, the University of Twente (PNUT) and The Netherlands (PNN) offered support, good times and valuable training. I have also been honoured to participate in the University Council; this was for me a challenging and mind-opening experience.

Finally I would like to thank all the people who, intentionally or unintentionally, helped me or hurt me in any way during my PhD.

I have learned a lot, and I am grateful.



# Table of Contents

Acknowledgements.....	i
List of figures .....	iv
List of tables.....	v
1.1 Introduction .....	1
1.2 Problem statement .....	2
1.3 Research objectives .....	3
1.4 Structure of the thesis .....	4
1.5 Data sources, satellite sensors, abbreviations.....	5
Chapter 2. Anomalous pre-earthquake thermal emissions reviewed .....	9
2.1 Introduction .....	9
2.2 Compilation of thermal anomalies linked to the Wenchuan and Gujarat earthquakes .....	10
2.3 Comparisons and discussion .....	19
2.4 Conclusions and implications for future research .....	24
Chapter 3. Finding a needle by removing the haystack: A spatio-temporal normalization method for geophysical data .....	27
3.0 Abstract.....	27
3.1 Introduction .....	27
3.2 Methodology.....	30
3.3 Application and evaluation .....	34
3.4 Discussion.....	42
3.5 Conclusions .....	44
Chapter 4. Study of Volcanic Activity at Different Time Scales Using Hypertemporal Land Surface Temperature Data.....	45
4.0 Abstract .....	45
4.1 Introduction .....	45
4.2 Materials and Methods .....	49
4.3 Results .....	54
4.4 Discussion.....	58
4.5 Conclusions .....	62
Chapter 5. Time series analysis of Land Surface Temperatures in 20 earthquake cases .....	63
5.0 Abstract.....	63
5.1 Introduction .....	63
5.2 Datasets .....	64
5.3 Methodology.....	66
5.4 Results .....	73
5.5 Discussion.....	81
5.6 Conclusions .....	86
Chapter 6. Uncertainty propagation in the normalization procedure .....	87
6.1 Introduction .....	87
6.2 Data.....	88

6.3	Methodology.....	88
6.4	Results and discussion .....	91
6.5	Conclusion .....	94
Chapter 7.	Synthesis and outlook .....	97
7.1	Normalization and anomaly detection .....	97
7.2	Volcanic application .....	99
7.3	Earthquake application.....	100
7.4	Outlook for future studies .....	102
Summary.....		105
Samenvatting .....		107
References .....		109
Appendix 1: Volcanic case study .....		127
Appendix 2: Anomaly density.....		134
Appendix 3: Statistical analyses .....		139

## List of figures

Figure 3.1. Methodology.....	31
Figure 3.2. Study areas.....	36
Figure 3.3. Effect of algorithm settings and missing data.....	38
Figure 3.4 Detection of synthetic anomalies of different magnitude.....	40
Figure 3.5 Comparisons between detection methods. ....	41
Figure 4.1. Study areas over Mount Etna (panel a) and Virunga National Park, D.R.Congo (panel b) .....	50
Figure 4.2. Methodology.....	52
Figure 4.3. Hotspot detection at Mount Etna.....	55
Figure 4.4. Analysis of data from different wavebands .....	56
Figure 4.5. Hotspot detection in Virunga National park.....	57
Figure 4.6 Detection results from the lava lake of Nyiragongo.....	59
Figure 5.1. An example of anomaly detection. ....	68
Figure 5.2. Definition of distance zones .....	71
Figure 5.3. Anomalies detected in the study area of Italy. ....	76
Figure 5.4. Per pixel numbers of anomalies, averaged every three months, in the study area of Baja California.....	77
Figure 5.5 Anomaly density for all studied earthquakes in earthquake and no-earthquake years .....	78
Figure 5.6. Relation between atmospheric parameters and anomalies in four study areas. ....	85
Figure 6.1. Factors influencing uncertainty propagation.....	93
Figure 6.2. Uncertainty propagation. ....	95
Figure 7.1. Comparison of midwave and longwave IR images. ....	99
Figure A1.2. Detection with different thresholds.....	129
Figure A1.4. Application of data availability thresholds on the normalization frame. ....	132
Figure A1.5. Application of data availability thresholds on the temporal window.....	133
Figure A3.1. Anomaly density using a $\mu+2\sigma$ threshold .....	140
Figure A3.2. Anomaly density using a $\mu+2\sigma$ threshold with co-seismic period adjustment .....	141
Figure A3.3. Anomaly density using a $\mu+3\sigma$ threshold .....	142
Figure A3.4. Anomaly density using a $\mu+3\sigma$ threshold after a co-seismic period adjustment.....	143

## List of tables

Table 1.1. Satellites and sensors mentioned in this thesis .....	6
Table 1.2 Reanalyses mentioned in this thesis, and their temporal and spatial resolution. ....	6
Table 2.1 Research on Wenchuan earthquake.....	11
Table 2.2 Research on Gujarat earthquake.....	16
Table 3.1 Experiments and related choices.....	37
Table 5.1. Earthquake events.....	66
Table 5.2 Definitions of spatial zones, temporal periods and adjusted co-seismic periods applied for anomaly density calculations. ....	70
Table 5.3. Summary of the results of ANOVA statistical tests.....	80
Table 6.1. Results of uncertainty propagation.....	91
Table A1. Algorithm settings.....	127





# Chapter 1. Introduction

## 1.1 *Introduction*

### 1.1.1 Earthquake precursors

Earthquakes form a considerable hazard to exposed populations and structures. As a result, effort is being invested internationally to investigate the “utilization of possible forerunners of large earthquakes to drive civil protection actions” [Jordan et al., 2011]. The current standard in earthquake predictability research consists of time-dependent earthquake hazard assessment, along with associated probabilities and errors [Tiampo and Shcherbakov, 2012]. At the same time, a variety of parameters appear in literature as potential earthquake precursors [Cicerone et al., 2009], often sparking debate [1999]. The term “diagnostic precursor” refers to “observations and/or interactions of physical parameters which can be linked, above the levels of chance, to subsequent earthquake occurrence” [Jordan et al., 2011]. Diagnostic precursors, although extensively studied (see chapter 2), are still poorly understood and their utility to produce short-term prediction schemes is very limited [Jordan et al., 2011]. However, more research on earthquake precursors could provide better understanding of earthquake-related processes and potentially support physical-based seismicity models.

Potential precursors include, among others, hydro-geochemical changes in groundwater aquifers [Ingebritsen and Manga, 2014, Du et al., 2010], ionospheric perturbations [Pulinets, 2004, De Santis et al., 2015], electromagnetic variations [Cicerone et al., 2009] and temperature increases [Tronin, 1996].

### 1.1.2 Thermal earthquake precursors

This thesis is triggered by the idea that some of the stress building up during the interaction of tectonic plates, could be expressed shortly before an earthquake as radiation of thermal energy (as explained in detail in [Freund, 2003]).

Reports have appeared in literature in the last forty years on anomalies in surface temperature retrievals, ground-based measurements, top-of-atmosphere satellite observations and numerical simulations [Shen et al., 2013, Tronin, 2010, Tramutoli et al., 2015, Pulinets and Dunajacka, 2007, Wang and Zhou, 1984]. These studies describe emissions prior to earthquakes in occurring worldwide, including Japan, Kamchatka, Central Asia, India, China, Greece, California (USA) and Iran. If such observations were proven to be related to earthquake occurrence and could be traced by satellite sensors, the

synoptic coverage of satellites will provide a tool to monitor earthquake-prone areas worldwide.

However, in literature related to earthquakes, the definition of the term 'thermal anomaly' is broad and not systematic. Researchers have adopted different views of what is normal and what is anomalous. As phrased by Tramutoli et al. [2015], "a clear definition of anomaly as well as a clear description of the processing phases which could isolate anomalies connected with seismic activities from any other cause, is very hard to find". The methodologies range from visual inspection of a few satellite images [Saraf et al., 2008] to complicated image processing procedures [Ouzounov et al., 2006]. The same methodology is often applied using different parameters and settings without sufficient explanation for the choices made [Saradjian and Akhoondzadeh, 2011]. As a result, reported findings on the same earthquake often contradict. Given the different settings, one cannot determine the cause of these differences. Furthermore, reported anomalies may be located hundreds of kilometers away from the earthquake and not even cover the epicenter [Zhang et al., 2010], challenging the existence of links between earthquakes and observed anomalies. Anomalies are often identified based on short time periods, covering two-three months around the earthquake with only a few images [Dey and Singh, 2003]. This does not allow an examination of possible anomalies without earthquake occurrence. As a result, other potential causes of observed anomalies (diurnal or seasonal variations, geomorphological features) are not considered and a statistical evaluation of the findings is often missing [Eneva et al., 2008].

Extracting the part of the data that relates to a specific process requires good knowledge of the process. However, processes that may cause thermal emissions due to pre-earthquake stress built-up are not well understood [Bhardwaj et al., 2017, Ouzounov and Freund, 2004]. This implies that any anomaly detection procedure would need to be designed without *a priori* known physical constraints on the characteristics of a typical earthquake-related anomaly. The challenge is extended by the fact that earthquake-related temperature fluctuations may not stand out than temperature fluctuations caused by other mechanisms. An earthquake is the result of natural processes and any earthquake contribution to the recorded emissions can overlay other signatures, potentially not increasing the signal enough to be considered abnormal.

## **1.2 Problem statement**

To address the limitations above, there is a need for:

1. a consistent *a priori* definition of anomaly, keeping in mind that there is no proven theory to constrain its physical characteristics;

2. an analysis of spatially and temporally extended datasets to account for periods/areas affected and not affected by earthquakes;
3. an analysis of different earthquake cases in different areas, to address potential links between earthquake characteristics and anomalous signals;
4. a detailed description of the spatial characteristics of detected anomalies, including their extent and distance from earthquakes;
5. a consideration of other potential effects on the data, like seasonal and climatic influences that may be changing with time; and
6. a comprehensive statistical evaluation of analysis results.

This thesis is aimed at addressing the above limitations in order to answer the following research questions:

1. Are there anomalous signals in thermal IR satellite sensor data, which uniquely coincide, spatially and temporally, with the occurrence of earthquakes?
2. When do such anomalies appear, where do they appear, and how do they change when earthquakes have different magnitude, focal depth or focal mechanism?

### **1.3 Research objectives**

#### **General objective:**

Evaluate if there are thermal IR anomalies, detectable from a satellite, which can be spatially and temporally linked to earthquake occurrences.

#### **Specific objectives:**

1. Formulate a suitable methodological approach in order to isolate spatially and temporally limited (localized) signal fluctuations.
2. Analyse multiple earthquakes at different locations to determine potential relations between observed anomalies, earthquakes and local conditions.
3. Statistically evaluate the occurrence (space, time) of potential anomalies in relation to magnitude, mechanism and location of earthquakes.

**Hypothesis 1:** Pre-earthquake processes, and/or the earthquake itself, have a traceable contribution to the thermal emissions from the earth surface which are recorded by satellite sensors.

This hypothesis is supported by literature concluding on the existence of earthquake-induced thermal emissions which can be detected prior to earthquakes, see for example Tronin [2000] and the literature review in Chapter 2.

**Hypothesis 2:** More anomalies are expected to appear before/during the date of the earthquake; at closer distance to the location of the epicentre; and only in the year when an earthquake takes place.

This hypothesis is supported by published research which applies criteria of spatial proximity and temporal coincidence to isolate earthquake-related anomalies (for example, Qin et al.,[2013]) and reports less pronounced presence of anomalies in years without earthquakes (for example, Tramutoli et al. [2005]).

**Hypothesis 3:** More anomalies are expected to appear before shallower earthquakes or earthquakes of higher magnitude.

This hypothesis is supported by research concluding that increasing magnitude and decreasing depth result in more easily observed anomalies [Xiong and Shen, 2017].

## **1.4 Structure of the thesis**

**Chapter 2** is a literature review on pre-earthquake thermal anomalies. Comparisons are made among findings published for the same earthquake, and differences are traced back to the applied methodologies. The study of published research leads to a detailed identification of the challenges related to the research objective, and informs methodological choices in the chapters to follow.

**Chapter 3** contains an introduction and test of a methodology, building on kernel-based image processing approaches. Because there is no physically derived description of an earthquake-induced thermal anomaly, the methodology instead suppresses commonalities between observations in order to highlight local differences. As a result, seasonal and spatially extended patterns are not misinterpreted as potential earthquake-induced effects.

**Chapter 4** presents a first real-life application of the methodology to monitor known volcanic targets using Land Surface Temperature (LST) data. This serves as a proof of concept for the performance of the methodology. Moreover, it has its own merit in the field of satellite volcanology: it facilitates the utilization of imagery which was previously not widely applicable for hotspot detection. Application of the methodology of chapter 3 highlights volcanic signatures in LST data, it allows utilization of 30-year-long longwave infrared (LWIR) archives to study volcanic dynamics in longer time series. It can complement analysis based on existing hotspot detection.

**Chapter 5** describes an earthquake-anomaly related study of LST datasets around the world. Areas are examined at times with and without an earthquake occurrence. The methodology provides a constraint on the spatial extent of

detected anomalies. A statistical evaluation investigates if there are significant differences between the anomaly density calculated before, after or during earthquakes, at different distances from the earthquake epicenter. Meteorological information and the Weather Forecasting and Research model (WRF) are applied to examine atmospheric influences on the results of detection.

In **Chapter 6**, uncertainty information of LST products is utilized to quantify error propagation through the processing chain. The results are used to evaluate the effect of LST retrieval uncertainty on the anomalies detected in the previous chapter. Finally, the results of all chapters are synthesized and concluded in **Chapter 7**, which also includes suggestions on further research.

## ***1.5 Data sources, satellite sensors, abbreviations***

Throughout this thesis different types of input, data sources and satellite sensors are mentioned. These are shortly presented here for reference.

### **Data sources**

**Satellite sensors** record the relative intensity of emissions from the earth surface and the atmosphere as digital numbers, which are converted to radiance units during radiometric preprocessing. The sensors mentioned in this thesis are summarized, along with their spatial and temporal resolution, in Table 1.1.

**Reanalyses.** A climate reanalysis gives a numerical description of the recent climate, produced by combining models with observations. Reanalyses provide an estimate of atmospheric and surface parameters such as air temperature, pressure and wind at different altitudes [Dee et al., 2011, Dee et al., 2014, Smith et al., 2001]. The reanalyses, and the temporal and spatial resolution of the data used in the publications and the tests mentioned in this thesis, are shown in Table 1.2.

Table 1.1. Satellites and sensors mentioned in this thesis, and their temporal and spatial resolution.

Satellite/sensor	Temporal Resolution	Spatial resolution
FY-2/VISSR	every 3hrs	0.1°
NOAA/AVHRR	daily (2.5°)/	2.5° (before 2006)/
	twice per day (1°)	1° (after 2006)
AQUA/AIRS	twice per day	1°
AQUA-TERRA/ MODIS	four times per day	1km
Meteosat First Generation (MFG)/ MVIRI	every half hour	5km
Meteosat Second Generation (MSG)/ SEVIRI	every 15min	3km
GOES	hourly	4km
GMS-5/VISSR	Every half hour	5km
Landsat TM	Every 16 days	60m
AMS-E	Twice daily	45km, 5.4km (resampled)
SSM/I	Twice daily	25km, 12.5km

Table 1.2 Reanalyses mentioned in this thesis, and their temporal and spatial resolution.

Reanalysis	Maximum Temporal Resolution	Maximum Spatial resolution
<b>NCEP/NCAR</b>	every 3 hours	1.9°x1.9°
<b>NCEP/FNL</b>	every 6 hours	1°x1°
<b>ERA-Interim</b>	every 6 hours	79 km
<b>ERA-5</b>	every hour	31 km

## Data types

Top-of-Atmosphere (**ToA**) or more accurately, at-sensor radiance is converted to Brightness Temperature (**BT**), the equivalent blackbody temperature in Kelvin, following Plank's equation. The BT mentioned in this thesis are derived from radiance recorded either in the Infrared bands (**TIR**, else longwave infrared, **LWIR**, with wavelength range 8-12µm; and midwave infrared, **MWIR**, with wavelength range 3-5 µm) or in the microwave bands (**MW**, typical frequency range 0.4-35 GHz) of passive satellite sensors.

Outgoing Longwave Radiation (**OLR**, W/m<sup>2</sup>), represents electromagnetic radiation emitted from earth and its atmosphere out to space at wavelengths between 4 and 100µm. OLR is the result of processes of absorption, scattering, and emissions from atmospheric gases, aerosols, clouds and the surface. OLR is not a measurable observation; it is derived from satellite radiance using

algorithms based on multispectral regression models [Ellingson et al., 1989, Lee, 2014] or from reanalysis.

Surface Latent Heat Flux (**SLHF**,  $\text{W/m}^2$ ) is an atmospheric parameter representing flux of heat from the earth's surface to the atmosphere due to phase transitions of water. SHLF can be calculated from ground-based measurements or satellite observations of wind speed, air humidity and surface temperature [Schulz et al., 1997]. Alternatively, it can be derived from numerical assimilation as a reanalysis product.

Land Surface Temperature (**LST**) is the directional radiometric temperature of the land surface, used as a best approximation to the thermodynamic temperature [Norman and Becker, 1995]. It is estimated from Top-of-Atmosphere brightness temperatures registered in the infrared spectral channels of satellite sensors, after application of radiometric and atmospheric corrections and temperature-emissivity separation. Surface temperatures can also be approximated by reanalysis products as Skin Temperatures (**T<sub>skin</sub>**). However, it should be noted that LST and T<sub>skin</sub> are in principle different variables, calculated in very different ways and representing different quantities.





## **Chapter 2. Anomalous pre-earthquake thermal emissions reviewed**

### **2.1 *Introduction***

Remote sensing has been reported as a tool for the study of earthquakes [Tronin, 1996,2010]. The first remote sensing reports on pre-earthquake thermal anomalies appear as early as 1988 ([Gorny et al., 1988], in [Tramutoli et al., 2015, Tronin, 2000b, Blackett et al., 2011]). Since then researchers reported thermal anomalies studying different parameters, satellite-based, numerically retrieved or measured on-site. These parameters are Top-of Atmosphere observations in the longwave thermal bands (TIR) of satellite sensors; satellite-based Land Surface Temperature (LST) retrievals; satellite-based or numerical retrievals of Surface Latent Heat Flux (SLHF); ground-based, near-surface air temperature measurements (usually at 2m above the surface) or soil temperature measurements; and satellite-based or numerical retrievals of Top-of-Atmosphere Outgoing Longwave Radiation (OLR) [Shen et al., 2013, Tronin, 2010, Tramutoli et al., 2015, Pulinets and Dunajacka, 2007, Wang and Zhou, 1984].

Observations that are referred to as 'anomalies' are described as short-term, sudden increases in signal [Tramutoli et al., 2015]. They may appear a few hours [Akhoondzadeh, 2013b] to a few years [Yao, 2010] prior to an earthquake and they sometimes reappear shortly after an earthquake [Tronin, 2000a]. Such anomalies are reported worldwide, including Japan, Russia, countries in Central Asia, India, China, Greece, California (USA) and Iran. Most literature is focused on shallow depth earthquakes (focal depth < 35km) with magnitudes above  $M_w$  4.

Different theories have been put forward to explain the potential physical link between the appearance of anomalies and processes that take place prior to an earthquake. These include the expulsion of warm gases and/ or liquids from stressed rock; the generation and propagation of electric currents with subsequent electromagnetic emissions; air ionization, water vapor condensation and latent heat release [Freund, 2011, Pulinets and Ouzounov, 2011, Hamza, 2001, Saraf et al., 2009, Liperovsky et al., 2008, Tramutoli et al., 2013]. Whereas some theories have been tested in laboratory conditions [Freund, 2011, Freund, 2003a, Freund, 2003b, Takeuchi et al., 2006, Wu et al., 2006, Zhang and Liu, 2011, Umarkhodgaev et al., 2012], none of them has been proven to relate universally to precursory observations in the field [Bhardwaj et al., 2017, Ouzounov and Freund, 2004, Ouzounov et al., 2006]. Since there is no physically-based description of the characteristics of earthquake-related thermal emissions, anomalies are identified in literature

based on different definitions with the application of different methodologies at different locations. This hinders the comparability of results and a quantitative validation of observed anomalies [Jiao et al., 2017].

Some earthquakes have been the subject of more than one study. In such cases, it is possible to compare the studies based on the same earthquake and examine commonalities and differences in their findings. This would allow a better understanding of applied methodologies and provide direction for future study. This chapter focuses on the two widely studied earthquakes, in terms of thermal anomaly-related literature: the Wenchuan, China (2008) earthquake, appearing in 11 studies and triggering a dedicated review [Ma and Wu, 2012] and the Gujarat, India (2003) earthquake, which is studied in 9 articles. First, an inventory of published reports on each earthquake is compiled (Tables 2.1 and 2.2). The results of the studies are presented and discussed, and an analysis follows on potential similarities and differences examined in relation to the applied methodologies.

## ***2.2 Compilation of thermal anomalies linked to the Wenchuan and Gujarat earthquakes***

### **2.2.1 Wenchuan (China, 2008)**

The Mw 7.9 earthquake of Wenchuan occurred on May 12, 2008 at a focal depth of 19km [Hayes et al., 2017]. It generated a 240km-long, right-lateral oblique faulting surface rupture zone along the Beichuan fault, and a 72km-long, dip-slip reverse faulting surface rupture zone along the Pengguan fault [Xu et al, 2009]. Among the first ones to report on thermal anomalies are Yang and Mi [2009], who use NCEP-NCAR reanalysis data at a spatial resolution of 1.9° latitude by 1.9° longitude over an area of 66°x78°. The authors study daily surface upward longwave radiation flux (ULWRF), soil temperature at 10-20cm depth ( $T_{\text{soil}}$ ) and air temperature at 2 meters above the ground ( $T_{2m}$ ). They calculate the 20-year mean and standard deviation for each day as a reference background.

*Table 2.1 Research on Wenchuan earthquake. Compilation of published thermal anomalies related to the earthquake of Wenchuan, 2008. Hyphens are placed when information is not available.*

Study	Data type	Spatial resolution	Reported duration of study	Methodology	First appearance of anomaly (days related to earthquake)	Last appearance of anomaly (days related to earthquake)	Max. duration (days)	Max. Extent (km <sup>2</sup> )
Yang et al [2009]	Reanalysis ULWRF		130 days		-60, several	10	-	-
Yang et al [2009]	Reanalysis Soil temp. at 10-20cm	1.9x1.9°	131 days	Difference from long-term reference	-120, several	-	15	-
Yang et al [2009]	Reanalysis Air Temp.		132 days		-60, several	10	-	-
Wei et al [2009]	FY-2 TIR BT	-	80 days	Difference from one reference image	-55, several	-	13	160x10 <sup>4</sup>
Singh et al [2010]	Microwave SSM/I BT	25x25km	-		-18	-	-	-
Singh et al [2010]	Air Temp. AIRS	-	90 days	visual inspection	-15	-	-	-
Singh et al [2010]	Humidity AIRS	-	-		-11	-	-	-
Zhang et al [2010]	FY-2 TIR BT	5x5km	2.5 years	wavelet and RPS	-11	-	60	-
Xiao et al [2010]	FY-2 OLR	5x5km	3 years	wavelet and RPS	-8 maximum +5	-	45	1200
Yao et al [2010]	-	-	-	visual inspection	-52	-	-	450x10 <sup>4</sup>
Yao et al [2012]	-	-	-	visual inspection	-years	-	-	450x10 <sup>4</sup>
Wu et al [2012]	FY-2 OLR	0.1x0.1°	-		no	-	-	-
Wu et al [2012]	NOAA-AVHRR OLR	2.5x2.5° 1x1°	-	deviation from long-term reference and criteria of spatial/temporal coincidence	no	-	-	-
Wu et al [2012]	AQUA-AIRS OLR	1x1°	-		-6	-	-	-
Wu et al [2012]	FY-2 TIR BT	5x5km	-		no	-	-	-
Wu et al [2012]	reanalysis skin temp.		-		-6	-	-	-
Wu et al [2012]	reanalysis Air Temp.		-		-6	-	-	-
Wu et al [2012]	reanalysis daily Temp. Range	1.9x1.9°	-		-6	-	-	-
Wu et al [2012]	reanalysis SHLF		-		no	-	-	-

Table 2.1, continued

Study	Data type	Spatial resolution	Reported duration of study	Methodology	First appearance of anomaly (days related to earthquake)	Last appearance of anomaly (days related to earthquake)	Max. duration (days)	Max. Extent (km <sup>2</sup> )
Qin et al [2013]	same as Wu et al [2010]		-	Same as Wu[2012] plus quantify criteria in a reliability index	-	-	-	-
Jing et al [2013]	NOAA-AVHRR OLR	2.5 x 2.5° 1 x 1°	-		-2 months -13 days	-1	2	2x10 <sup>4</sup>
Jing et al [2013]	reanalysis SLHF		-		-1	-	-	-
Jing et al [2013]	reanalysis air Temp		-	deviation from long-term reference	-10	-	-	-
Jing et al [2013]	reanalysis air pressure	1.9 x 1.9°	-		-14	-	-	-
Jing et al [2013]	reanalysis relative humidity		-		-10	-	-	-
Wei et al [2013]	FY-2 TIR BT	5x5km	5 years	wavelet and RPS	-4	-	80	-

Anomalies are reported as deviations from the 20-year maximum, minimum or mean, or as exceedance of twice the 20-year standard deviation. The authors find that, in the two months preceding the earthquake, ULWRF exceeded the 20-year maximum three times and also reached the 20-year minimum another three times. ULWRF values above the 20-year maximum persisted for ten days after the earthquake. ULWRF values for the rest of the year are not discussed.  $T_{soil}$  in the month preceding the earthquake is found to be close to the 20-year average. Anomalies are not present in the month before the earthquake, but the authors report 15 days with  $T_{soil}$  above the 20-year maximum occurring four months before the earthquake and more than 30 days with  $T_{soil}$  lower than the 20-year minimum occurring 3-2 months before the earthquake. Finally, the authors report several anomalies in  $T_{air}$  during the two months preceding the earthquake and the ten days after the earthquake. Again, there is no mention about  $T_{air}$  values in the rest of the year.

Wei et al. [2009] study geostationary satellite-based TIR data (10.3-11.3 $\mu$ m) recorded up to 80 days prior to the earthquake over an area 25°x40°. They

choose one satellite image in February as reference image and report anomalies as deviations from this reference. They find the first anomalies 55 days prior to the earthquake, when the temperature difference from the reference exceeded  $10^{\circ}\text{C}$  in an area  $25 \times 10^4 \text{ km}^2$  located approximately  $10^{\circ}$  east and south of the earthquake. They report similar anomalies throughout the period of study and lasting up to 13 days, with an extent that covers almost the entire study area, where deviations  $>10^{\circ}\text{C}$  extent  $160 \times 10^4 \text{ km}^2$ .

Singh et al. [2010] study brightness temperatures derived from the Special Sensor Microwave Imager (SSM/I, ground resolution 25km), meteorological data from a ground-based station closest to the earthquake epicenter, and air temperature and relative humidity data derived from the Atmospheric InfraRed Sounder (AIRS) at different pressure levels. Results are shown for a period of one month before the earthquake until two months after. No specific anomaly definition is provided; instead, the authors describe increases in absolute values of the variables of interest, which they consider anomalous. Such increases are reported 18 days prior to the earthquake for microwave BT; 11 days prior to the earthquake for relative humidity; and 11-15 days prior to the earthquake for air temperatures at different levels. It should be noted that similar increases are visible on other dates throughout the year of data in the graphs provided.

Zhang et al. [2010] use TIR BT values derived from the FY-2 geostationary satellite series (spatial resolution 5km). They use one night-time image per 24 hours to construct a time series on which a wavelet transform is used to model its frequency components. They state that the basic earth temperature field and the annual variation temperature fields are removed by removing the wavelet seventh order part. Clouds and cold-heat air currents act on a short scale of several hours to several days, and their effects can be removed from the signal by rounding off the wavelet second-order part. The authors then apply Fourier transform in order to perform power spectrum estimation. They study the frequency content of the signal in time windows of 64 observations and observe changes in dominant frequency among consequent windows. In their findings, the Wenchuan earthquake was preceded by anomalies eleven days before the earthquake, and anomalies are persistent until the end of the month following the earthquake.

The same methodology is followed by Xiao et al. [2010] who use OLR data derived from two TIR and one water vapor channel of the FY-2 geostationary satellite series. The authors use three night images per 24 hours for a period of three years over an area of  $50^{\circ} \times 95^{\circ}$ . The authors report that an anomaly lasting for 45 days appeared the month before the earthquake and affected an area of  $1200 \text{ km}^2$ . The power spectrum showed a peak 8 days before the

earthquake and reached its three-year maximum amplitude five days after the earthquake.

Yao and Qiang [2010] describe in the region between 28°–45°N, 70°–102°E a series of what they call thermal ellipses, designating areas of increased temperature that extended up to 3000km in length and 1500km in width. These areas were identified using visual inspection of FY-2 satellite imagery, but the analysis approach is not further described. The ellipses are suggested to connect the epicenters of the Wenchuan and Yutian earthquakes, located more than 2000km apart and occurring within a time interval of 52 days. In subsequent work [2012] and using surface temperatures derived from geostationary based FY-2 TIR observations, the same authors relate the appearance of such ellipses prior to the Wenchuan earthquake with the occurrence of another earthquake, which took place in Yushu, China, 2 years, 1 month and 10 days later.

Wu et al. [2012] use a multitude of data to investigate the presence of anomalies prior to the Wenchuan earthquake. These include geostationary- and polar- satellite-derived OLR, geostationary-based TIR BT observations, and NCEP-NCAR reanalysis-derived surface temperatures, air temperatures, SLHF and Diurnal Temperature range (DTR). Their study covers an 8°x12° area over the epicenter of the earthquake. They study a period three months before and one month after the earthquake, using reference values of 3-28 years depending on data source. They calculate the mean  $\mu$  and standard deviation  $\sigma$  of reference years on the same dates and identify candidate anomalies when observations in the year of the earthquake exceed the  $\mu + \sigma$  of the reference values. Anomalies are declared when (a) candidate anomalies appear over tectonic structures and active seismogenic zones of the area, and (b) candidate anomalies appear “at approximately the same time” when different input data are analyzed. Following this approach, the authors find anomalies close to the epicenter 6 days prior to the earthquake in four parameters: DTR, OLR from AQUA-AIRS, Air Temperature (at 500hPa) and skin temperature. SLHF, TIR BT from FY-2, and OLR from NOAA AVHRR and FY-2 did not show any anomaly.

Qin et al. [2013] evaluate the results of Wu et al. [2012] by quantifying the reliability of an anomaly using three indices. The first index describes the degree of deviation of observations in the year of the earthquake from the long-term average (intensity of the anomaly). The second index describes the degree of spatial adjacency between the location of an anomaly and the earthquake epicenter or a seismogenic fault. The third index describes the spatial coincidence between the anomalies detected using different inputs, within a temporal window. The three indices are numerically combined to produce a reliability index. Using this approach, the authors calculate that the reliability index of the results reported in Wu et al. [2012] is 69.27%.

Jing et al. [2013] also use multiple data sources to detect anomalies prior to the earthquake of Wenchuan. They study NOAA-AVHRR OLR, and SLHF, air temperature, air pressure and relative humidity, derived from NCEP-NCAR reanalysis. The authors identify anomalies as deviations from long-term reference values. Using a year of monthly OLR data at  $2.5^{\circ} \times 2.5^{\circ}$  spatial resolution, they report anomalies close to faults and the epicenter of the earthquake, starting two months before the earthquake, reaching a peak in the month of the earthquake and lasting until the end of the month after the earthquake. Using daily OLR data with  $1^{\circ} \times 1^{\circ}$  spatial resolution, they report frequent 1-2 day long anomalies, appearing for the first time 13 days before the earthquake and extending over an area of  $20,000 \text{ km}^2$ . The last OLR anomaly is reported one day prior to the earthquake. One SLHF anomaly is reported one day before the earthquake and another on the day of the earthquake, close to the epicenter. Using monthly and daily meteorological data of 38 years and 30 days respectively, from two gridcells, the authors find that high air temperature (but not the highest) and the lowest air humidity appear together ten days and one day before the earthquake; and air pressure drops 14 days and 7 days before the earthquake.

Finally, Wei et al. [2013] study the Wenchuan earthquake using five years of TIR BT recorded in FY-2 geostationary satellite sensors at a spatial resolution of 5km. The authors average five night-time images per 24 hours and apply the wavelet transform (WT) and relative power spectrum (RPS) estimation as described in Zhang et al [2010] and Xiao et al [2010]. The authors report an anomaly that reached a peak 4 days before earthquake and lasted 80 days.

### 2.2.2 Gujarat (India, 2003)

The Mw7.6, earthquake of Gujarat took place within the continental crust of the Indian plate on January 26th, 2001 at a focal depth of 17km, and was the result of shallow oblique reverse faulting [Hayes et al, 2017]. The first study on this earthquake is published by Dey and Singh [2003], who use SLHF data from the NCEP/NCAR reanalysis because they claim that SLHF is related to the increases in TIR and LST which appear prior to earthquakes. Their study area has a spatial extent of approximately  $12^{\circ} \times 12^{\circ}$  and a spatial resolution of  $1.9^{\circ} \times 1.9^{\circ}$ . The authors study daily SLHF data of one month before and one month after the earthquake. Monthly means are calculated and subtracted from daily SLHF with the intention to remove seasonality. An anomaly is declared when the result of this subtraction exceeds the  $\mu + 1.5\sigma$  of the same date in ten previous years. Anomalies over the epicenter are shown 25, 4 and 2 days prior to the earthquake, as well as 3 days after. Spatially, the detected anomaly, at its maximum extent, is shown to affect approximately half of the  $12^{\circ} \times 12^{\circ}$  image. It is interesting to note that the resolution of the images shown in the article do not seem to correspond with the spatial resolution of the

product: in some gridcells, the anomaly appears to cover only part of that gridcell.

*Table 2.2 Research on Gujarat earthquake. Compilation of published thermal anomalies related to the earthquake of Gujarat, 2003. Hyphens are placed when information is not available.*

Study	Data type	Spatial resolution	Reported duration of study	Methodology	First appearance of anomaly (days related to earthquake)	Last appearance of anomaly (days related to earthquake)	Max. duration (days)	Max. Extent (km <sup>2</sup> )
Dey and Singh [2003]	Reanalysis SLHF	1.9x1.9°	60 days	Difference de-seasonalized data from long-term reference	-25, several	+3	-	-
Ouzounov and Freund [2004]	MODIS LST	1x1km	90 days	Difference spatially averaged LST from reference year	-6	-	-	10 <sup>4</sup>
Ouzounov et al [2006]	METEOSAT TIR	4x4km	-	Regression slopes for night cooling	-1	-	2	-
	MODIS LST	1x1km	85 days	Difference spatiotemporal averages from reference	-6	+2	-	-
Ouzounov et al [2007]	Reanalysis OLR	2.5x2.5° 1x1°	60days	Difference spatiotemporal averages from reference	-18	-	-	-
Saraf et al [2005a,b]	NOAA-AVHRR LST	1.1x1.1km	7 images 77 days	Visual inspection	-12	-	-	-
Cervone et al [2005]	Reanalysis SHLF	1.9x1.9°	5 years	Wavelet transform and spatial constraints	-3	+5	-	-
Genzano et al [2007]	METEOSAT BT	5x5km	2 months	RETIRA index and spatiotemporal constraints	-15	-	-	-
Blackett et al [2011]	MODIS LST	1x1km	6 years	Differencing as in Ouzounov [2004] and RETIRA as in Genzano [2007]	no	-	-	-

A report on thermal anomalies preceding this earthquake appears in 2004 by Ouzounov and Freund [2004]. The authors use MODIS LST data with spatial resolution of 1km over an area of 100x100km. They study a period of two months before and one month after the earthquake, in the earthquake year and in the following year. They calculate the daily mean LST, spatially averaged over the whole study area, in both years. They define an anomaly as a deviation of the earthquake year daily mean from the non-earthquake daily mean. Following this approach, they identify the highest positive deviation 6 days before the earthquake. In their findings, anomalies appear also after the earthquake. Ouzounov et al. [2006] continue to study the same earthquake by applying two approaches. In the first approach, they use night-time



geostationary-based TIR observations (spatial resolution 4km, temporal resolution 30min) to calculate regression slopes during the night. Anomalies are identified when night-time warming (positive slopes) are observed, instead of the expected night-time cooling. Following this approach, anomalies are found on the night before the earthquake and persist on the night of the earthquake. In the second approach, the authors use the MODIS LST product (1km spatial resolution) and define a 100x100km area around the earthquake epicenter. In this area they calculate the square root of the spatial average of  $LST^2$  of all pixels. They further average this quantity (a) daily and (b) for the whole duration of the dataset. The authors define an anomaly as the difference between the daily and full-length-dataset average  $\Delta LST$ . Following this approach, the authors show anomalies 6 days before the earthquake and 2 days after. The temporal length of the dataset is 85 days. Finally, Ouzounov et al. [2007] study the same earthquake using OLR data. Anomaly detection is performed using monthly means of  $2.5^\circ$  spatial resolution in an area of  $5^\circ \times 5^\circ$  over the earthquake epicenter, and a gridcell with  $1^\circ$  spatial resolution centered on the earthquake location. In monthly data, the authors subtract from the epicentral gridcell, the average value of four adjacent gridcells. These differences are used for anomaly definition. Using daily data, the spatial average of the gridcells in a  $10^\circ \times 10^\circ$  area over the epicenter is subtracted from the epicentral gridcell value. This is calculated for each day and as a five-year average. An anomaly is defined as a deviation of the difference between the daily value and its five-year average, when the deviation exceeds  $+1\sigma$  of the differences of the year of the earthquake. Anomalies are shown 5, 10, 16, 26, 32, 38, 48 days before, and 2 and 4 days after the earthquake. Results are shown only for the month of the earthquake and one month before.

Saraf and Choudhury [2005a, 2005b] recognize anomalies prior to the Gujarat earthquake by visual inspection of LST images. The authors derive LST data from NOAA Top of Atmosphere TIR observations at a nominal spatial resolution of 1.1km. For their analysis, they use a total of 7 images recorded over a period of 77 days around the earthquake, and they identify anomalies applying user-defined thresholds which are not further described. The authors report that the first anomalies appeared on the image recorded 12 days before the earthquake and their amplitude was  $+5-7^\circ\text{C}$ . When the same approach is applied on the same period for other years, similar anomalies are not found.

Cervone et al. [2005] apply a 1D Wavelet transformation on the reanalysis-based Surface Latent Heat Flux (SLHF) dataset of NCEP/NCAR, at a spatial resolution of  $1.9^\circ \times 1.9^\circ$ . The transformation is intended to isolate wavelet maxima which propagate from finer to coarser scales. These maxima are identified as anomalies when they occur at the same time (within 2 days) and when their spatial distribution coincides with local geological features, e.g. continental boundaries or faults. The authors mention that, due to the complex

geology of the area, they constrain the spatial continuity of earthquake-related anomalies using previously published reports rather than by local geological features. They report that within the year of the earthquake, four anomalous signals had the same geometric path, and they estimate the significance of their findings using statistics based on 5 years of data. Of these anomalous signals, one appeared 3 days prior to the earthquake of Gujarat and one appeared 5 days after. The authors do not describe the characteristics of anomalies they found within or outside of the geographical areas pre-defined as related to the earthquake. They mention that the highest anomalies are found over the ocean close to the epicenter, because SHLF values are low over land in this time of year.

Genzano et al. [2007] study the same earthquake by applying an anomaly detection methodology based on the RETIRA-index introduced by Tramutoli et al. [2001]. The index is calculated by subtracting the spatial average of an undefined number of observations in the neighborhood of a pixel ( $\Delta T$ ), from the current observation on that pixel; then subtracting the temporal average of  $\Delta T$  of previous years, on the same location and date; and finally dividing that difference by the standard deviation of the  $\Delta T$  of previous years, on the same location and date. The authors use METEOSAT TIR data, with a spatial resolution of 5x5km and a sampling frequency of one night image per 24 hours. They use five years of data to construct reference fields and perform anomaly detection in a period of two months: the month of the earthquake and the month that follows. The analysis is repeated in the same two months in a year without Mw>5.5 earthquake. In the year of the earthquake of Gujarat, the authors report anomalous pixels appearing 15 days before the earthquake. The number of anomalous pixels reaches its peak four days before the earthquake and then slowly decreases, but anomalies are present also after the earthquake. The authors use a criterion of space-time persistence to distinguish which of these anomalies are significant. The criterion is based on expert opinion and is broadly defined as a requirement for anomalies to be 'spatially extended and persistent in time, together with high intensity'. Using visual inspection, the authors also exclude sequences of anomalies that may be contaminated by clouds. On this basis, the authors identify three meaningful sequences of anomalies. One appearing at sea at 11 days before the earthquake, and reappearing in the same area 5 days before the earthquake; one that appears in the Himalayas the day before the earthquake and lasts until four days after; and one in central India, with anomalies detected within 15 to 5 days before the earthquake, which according to the authors may also be related to cloud cover. None of the anomalous sequences described in this study are found over the epicentral area of the earthquake. Their spatial extent is variable and not clearly defined.

Finally, the Gujarat earthquake is studied by Blackett et al [2011]. Their study is based on the LST product of MODIS, at a km spatial resolution and for a continuous record of 6 years. The authors apply the same differencing method as Ouzounov and Freund [2004] but on longer time series. They compare not only observations of the earthquake year with observations of one non-earthquake reference year, but they apply the same differencing technique by pairing observations between all available years, to better understand the normal thermal variability in the area. They repeat the analysis in two areas, 100x100km and 1500x1500km, and they additionally use the same RETIRA-based methodology as Genzano et al [2007]. Their findings show that observations, which are identified as anomalous in the analysis of Ouzounov and Freund [2004], appear to fall within the range of natural variability when longer time series are considered using the same methodology. The authors do identify anomalies using the RETIRA-based analysis. However, they also find that these anomalies are not related to the earthquake, but are caused by data gaps due to cloud masking and image mosaicking. The authors conclude that there is no evidence to support the presence of thermal anomalies prior to the earthquake and recommend caution in anomaly detection procedures, including the use of long time series and the examination of the effect of missing values, cloud masking and data mosaicking.

## **2.3 Comparisons and discussion**

This section presents an analysis of the observations listed above in both earthquake cases, with special focus on issues related to methodology.

Different authors have used a variety of inputs with different spatial and temporal resolution, and with the application of different methods. The extent, duration and time of appearance of reported anomalies vary greatly among methodologies and data. Anomalies reported for the Wenchuan earthquake, for example, cover areas that range among different studies between 1200km<sup>2</sup> and 450x10<sup>4</sup> km<sup>2</sup>. Their first appearance is reported from years to only one day before the earthquake, and it also happens [Xiao et al, 2010] that the largest anomaly appears after the earthquake. Wu et al. [2012] find anomalies in OLR data from AIRS but not from NOAA-AVHRR (which are of the same spatial resolution), even though these data are studied with the same methodology over the same area and for the same time period. When the same methodology is followed, the use of (slightly) longer datasets leads to different characteristics of the detected anomalies. For example, comparing the work of Zhang et al [2010], Xiao et al [2010] and Wei et al. [2013], who used the same wavelet transform-based approach to study the Wenchuan earthquake on datasets of slightly different duration, find anomalies 11, 8 and 4 days respectively prior to the event. The reported duration of anomalies is different (from 45 to 80 days) as well as the amplitude. The influence of the length of a

data time series is clearly shown in the study of Blackett et al [2011]. With the use of the same methodology on the same data as Ouzounov and Freund [2004], but including more years in the analysis, it is concluded that (1) anomalies that were identified before appear to fall within the natural variation and (2) observations which seemed anomalous actually appear quite often without the presence of an earthquake. The above indicate that the results of anomaly detection are not robust and they are sensitive to the duration of the analyzed dataset.

Below, the reasons behind the disagreement of these findings are traced back to methodological choices, which are earthquake independent. The studies are compared in terms of their data, the spatial extent of study areas and detected anomalies, the temporal extent of study areas and detected anomalies, the methodological approach and the statistical evaluation. The data from all reviewed studies are summarized in Tables 2.1 and 2.2. Limitations in current research are identified to identify choices in future research.

### 2.3.1 Data

The choice of data is a factor that affects detection results. The spatial and temporal resolution of the data affects the detail in which the anomalies can be characterized, especially in their relation to earthquakes. The parameter used for anomaly detection, the related product uncertainties, and missing values affect the uncertainty on detected anomalies.

Earthquake-related processes develop at a scale of kilometers (a study of statistical physics of earthquake-related processes can be found, for example, in Kawamura [2012]), and therefore need to be studied using data of a spatial resolution corresponding to that scale. The use of input data with coarse spatial resolution (of  $1^{\circ}$ - $2.5^{\circ}$ ) in literature does not provide sufficient information on anomaly location, anomaly extent and the spatial relation between anomalies and earthquakes. A strong anomaly of limited spatial extent, averaged over a gridcell of  $2^{\circ}$ , could show similar to a weak anomaly extending to the whole gridcell. An anomaly in a gridcell of  $2^{\circ}$  over the earthquake epicenter may be located at the epicenter or more than 200km away. Coarse input with spatial resolution in the scale of degrees latitude/longitude is used in nine of the twenty reviewed studies, and there are also cases in which finer resolution data are spatially averaged to coarser resolution [like in Ouzounov et al, 2004; 2006].

In earthquake-related literature, reported anomalies can have a duration as short as only a few hours up to one day [Akhoondzadeh, 2013, Cervone et al., 2006]. Three of the reviewed studies are based on monthly or yearly averages. With these data it is not possible to know if anomalies appear two hours or twenty days prior to the earthquake. Two studies are using one image every

~10 days [Saraf et al 2005a,b], with which it is challenging to support that a detected anomaly persists between two subsequent images. Of the remaining studies, three are using daily averages, six are using one image per 24 hours and only three are using 2-5 images per 24 hours. The use of data with a temporal resolution higher than in the reviewed studies would be necessary to capture transient anomalies, to sufficiently describe the duration of an anomaly, and to specify the time of appearance of an anomaly relative to the occurrence of an earthquake.

All types of data used in the reviewed studies suffer from the effect of clouds, cloud cover and cloud remnants. The opacity of clouds for TIR sensors can obscure transient anomalies, and introduces gaps in the dataset thereby reducing the temporal resolution. This problem is inherent to the wavelength of TIR sensors, and can only be mitigated with the application of cloud masking. There are studies which do not apply cloud removal on the data, assuming that cloud presence can be sufficiently modelled by wavelet components [Zhang et al, 2010; Xiao et al., 2010; Wei et al, 2013]. However, in most other reviewed studies, it is recognized that cloud removal is an intricate procedure and cloud presence affects the results of detection [e.g. Blackett et al, 2011; Genzano et al., 2007]. Passive microwave sensors (used in Singh et al., 2010) can register radiance regardless of the presence of clouds, however even in that case clouds have a cooling effect and reduce observed emissions.

Artifacts in the data may be introduced due to image mosaicking, viewing angle and geolocation errors. These affect primarily observations of polar-orbiting sensors [Blackett et al., 2011; Tramutoli et al., 2001], which are used in half of the reviewed studies. Geostationary sensors are less affected by geolocation errors [Aliano et al, 2008]. There are still geolocations errors due to satellite movement, but the viewing angle is stable in this case, as opposed to orbiting satellites. Data from geostationary sensors are used in eight of the reviewed studies. Another six studies use NCEP reanalysis data as input. The concern in this case, as brought up by Zhang et al [2013], is that the NCEP reanalysis system, although consistent, has evolved through time and the data accuracy is time dependent. In particular surface fluxes are heavily dependent on the model and may contain regional biases. None of the reviewed studies quantified the uncertainty that is related to the input.

In some of the studies, basic information on the type of data used, the duration or the total extent of the dataset or the characteristics of detected anomalies is completely missing [e.g. Yao et al, 2010].

### 2.3.2 Spatial extent

Eleven of the twenty reviewed studies do not provide information on the extent of study area. For the other nine studies, the area ranges from 10x100km to

55x95°. Similarly, the spatial extent of detected anomalies is provided in only five cases, ranging from 1200km- 450x10<sup>4</sup> km among different studies. The extent of the study area is, first of all, relevant for methods which include spatial averages. For example, in the calculation of the RETIRA index, the results are influenced by scene statistics [Bhardwaj et al., 2017b, Bhardwaj et al., 2017a]. Secondly, sufficient area should be included to allow for an examination of the spatial extent of observed anomalies. In some of the reviewed studies, the analysis is confined to a few pixels over the earthquake epicenter [Ouzounov and Freund, 2004] or even along a tectonic feature [Cervone 2005]. In such cases, there is not enough information to describe the full extent of a detected anomaly. It is also not possible to examine the potential presence of the anomaly over areas unaffected by the earthquake. In terms of the distance between anomalies and earthquakes, it can sometimes be noted that anomalies are described in locations thousands of kilometers far from the earthquake which do not cover the epicenter [Zhang et al, 2010; Wei et al, 2009]. This makes it challenging to physically explain the link between the observed anomalies and earthquakes. Piroddi and Ranieri [2012] argue that observable phenomena further than 60km from the earthquake, even if they were associable to the seismic event, would not be practically useful as precursors because the potential alarm areas would be too big.

### 2.3.3 Temporal extent

The majority of the studied articles analyze only a few days of data before and after the earthquake. Ten of the reviewed studies have a duration of a 1-3 months, and in two of the studies [Saraf et al., 2005a,b] a tested period of 77 days is represented by only seven images. Anomalies are also commonly reported after the earthquakes [Saraf et al, 2005a,b; Xiao et al, 2010], and it would be interesting to see if they actually disappear or that they are a recurrent, unrelated phenomenon. Testing only short time windows in the year of the earthquake leaves the question whether the reported anomalies would be present at times without earthquake occurrence. Only two of the studies repeat the analysis in years without earthquake occurrence [Blackett et al, 2011; Genzano et al, 2007], and in both studies anomalies are found also in years without earthquakes.

### 2.3.4 Methodological approach

The methods applied in the reviewed studies fall into five broad categories: differencing from a reference background [Yang et al., 2009; Wei et al., 2009], Visual inspection [Yao et al., 2010; Singh et al., 2010; Saraf et al., 2005a,b], Time-frequency analysis with wavelet transform [Zhang et al., 2010; Cervone et al., 2005], anomaly indices [Genzano et al., 2007; Qin et al., 2013] and modeling the rate of night cooling [Ouzounov et al., 2006]. Anomaly definitions vary among these methodologies. Due to these differences, quantitative

comparisons of the findings are difficult, not only among results of analysis in different earthquakes but also among results on the same case study [Wu et al, 2012]. For example, the amplitude of anomaly that is calculated by differencing observations from long-term averages of reference years cannot be compared to the amplitude of a peak in the power spectrum or to the value of an anomaly index. There are cases where an explicit anomaly definition is missing [Singh et al, 2010] or is based on subjective decisions without further justification [Yao et al, 2010; Saraf et al, 2005a,b].

The most common approach is differencing from, or comparing to, a long-term reference of historical observations. A crucial issue, both for anomaly detection and for comparability of results, is the length of the chosen reference dataset. Two studies use 20-year averages, another two use 6-year averages, one uses 10-year averages and one uses 3-year averages as a reference. In two of the studies, the reference was only one image [Ouzounov et al., 2006; Wei et al., 2009], and in another, the reference consisted of images of a few days [Ouzounov et al., 2004]. Furthermore, reference fields are calculated in monthly/daily/yearly averages which follow the formal calendar. Seasonal patterns, however, do not necessarily follow the civil calendar. In fact, as found by Eneva et al [2008], the change to a moving average-based calculation affects detection results. Additionally, comparisons with previous years do not account for interannual variations and longer term trends. NASA's Goddard Institute for Space Studies (GISS) in 2015 declared the first six months of 2015 as the warmest since 1880; this statement was reviewed when 2016 became the new warmest of the record. Comparing an earthquake month of 2016 with a reference month of preceding years might therefore result in an anomalous warm period unrelated to any potential earthquake activity. Comparison based studies should therefore account for climatic and meteorological variation before considering potential earthquake warming effects.

The reviewed methods often depend on choices that are not necessarily based on, or backed by, physical evidence. The lack of quantitative physical understanding of the phenomena underlying the appearance of thermal phenomena prior to earthquakes could justify exploratory studies testing different detection thresholds [Jiao et al., 2017]. However, some of the choices made in the reviewed studies are neither exploratory nor backed by evidence. For example, in the case of time-frequency analysis, the 'rounding-off' of the second-order wavelet part is claimed to remove all weather effects, including clouds, rain and air currents [Zhang et al, 2010; Xiao et al., 2010; Wei et al, 2013]. It is not clear why the presence of clouds, air currents and precipitation can be described by the same wavelet order since they can have very different time scales and are not necessarily correlated to each other (only a small amount of all clouds leads to rain). Jiao et al. [2017] comment that the

wavelet-based approach has a certain arbitrariness and requires gap-filling of missing values, which introduces further uncertainty. Other approaches involve complex spatio-temporal averaging procedures [Ouzounov et al 2006, 2007], which is not based on physical characteristics of an expected anomaly, nor on a statistical distribution of the data. Moreover, in some of the reviewed studies [Cervone et al, 2006; Genzano et al, 2007; Wu et al, 2012; Qin et al, 2013], detected 'candidate' anomalies are declared as actual anomalies only if they fulfill specific criteria which are often subjective, vary among researchers and not physically backed. For example, Genzano et al [2007] use expert judgement to discard anomalies related to so-called 'spurious effects' and keep only anomalies with spatial and temporal persistence not further defined. Expert interpretations may be subjective, especially when they are not well defined, and as a result this work cannot be reproduced.

Weather and meteorological influences could affect detection but are rarely considered in the reviewed studies. Genzano et al [2007] and Blackett et al [2011] specifically do mention the effect of clouds in the results of detection. Such effects should be considered, as in earthquake related literature [Qu et al., 2006, Jie and Guangmeng, 2014] pre-earthquake anomalies are reported to be linked with meteorological effects (temperature inversions, presence of clouds) rather than earthquakes.

### 2.3.5 Statistical evaluation

Years without earthquakes are, when they are considered, mostly used as a reference but not for checking if anomalies also occur in these years. Statistical evaluation of results in general, and over multiple years in particular, is not commonly applied [Eneva et al., 2008]. Statistical analyses in the reviewed articles are mostly used to describe the significance of detected anomalies [Cervone et al, 2005; Wu et al, 2012] or to evaluate relations between earthquake characteristics and detected anomalies [Dey and Singh, 2003]. None of these studies discusses the actual distribution of SLHF or Brightness Temperature values and if it can support the use of the given Confidence Intervals or the chosen significance tests. Genzano et al [2007] examine the case of years without earthquakes for confutation reasons. They do find anomalies in periods without earthquakes but they describe them as sporadic and not persistent. The criteria of spatial and temporal persistence of a reliable anomaly, however, are based on expert opinion and are not described in the study.

## **2.4 Conclusions and implications for future research**

The literature review on thermal anomalies preceding two widely studied earthquakes showed large discrepancies in the characteristics of reported



anomalies. Findings are often conflicting and that questions the robustness of results. Anomalies are not well defined and sufficient description of the applied data processing is sometimes lacking. Large scale, recurrent patterns in the data are either not addressed or are tackled with procedures that depend on the duration and spatial extent of the chosen datasets. Data uncertainty propagation is not considered in anomaly detection, and the spatial and temporal resolution of data is in certain studies coarse and does not allow for detailed examination of the spatiotemporal relation between earthquakes and detected anomalies. In most reviewed studies, other potential causes of the detected anomalies are not taken into account, periods and areas without earthquakes are not examined, and the detection results are not statistically evaluated.

Future research should be based on spatially and temporally extended datasets to allow for examination of non-earthquake affected areas and periods. Full years should be examined, including all seasons and including years without earthquakes. Data of high temporal resolution should mitigate data gaps due to cloudiness, address seasonal and other recurrent components of the signal which change with time, capture transient anomalies, and resolve the temporal relation between earthquakes and detected anomalies. Input of spatial resolution in the order of kilometers, rather than degrees latitude/longitude leading to tens of kilometers, would allow to characterize the extent of anomalies and their distance from earthquakes. The choice of data should involve a consideration on the effect of clouds, observational and processing errors, and the propagation of data uncertainties should be accounted for.

Anomaly definition should be based on well described criteria, consistently applied in the whole study area. Earthquakes do not occur at known periodicities and their effects are spatially finite. Thus, recurrent temporal patterns (for example, seasonality) and patterns extending in the whole study area (for example, extended weather fronts) should be addressed in the methodology. Ideally, anomaly detection would consider the context of each data instance and not only its historical values. This would reduce the sensitivity of the detection procedure to changes in the length of the datasets. Finally, research should be extended to multiple earthquake case studies to study the effect of different magnitude, source mechanisms and epicentral depth, and the findings should be statistically evaluated.



# Chapter 3. Finding a needle by removing the haystack: A spatio-temporal normalization method for geophysical data<sup>1</sup>

## 3.0 Abstract

We introduce a normalization algorithm which highlights short-term, localized, non-periodic fluctuations in hyper-temporal satellite data by dividing each pixel by the mean value of its spatial neighbourhood set. In this way we suppress signal patterns that are common in the central and surrounding pixels, utilizing both spatial and temporal information at different scales. We test the method on two subsets of a hyper-temporal thermal infra-red (TIR) dataset. Both subsets are acquired from the SEVIRI instrument on board the Meteosat-9 geostationary satellite; they cover areas with different spatiotemporal TIR variability. We impose artificial fluctuations on the original data and apply a window-based technique to retrieve them from the normalized time series. We show that localized short-term fluctuations as low as 2 K, which were obscured by large-scale variable patterns, can be retrieved in the normalized time series. Sensitivity of retrieval is determined by the intrinsic variability of the normalized TIR signal and by the amount of missing values in the dataset. Finally, we compare our approach with widely used techniques of statistical and spectral analysis and we discuss the improvements introduced by our method.

## 3.1 Introduction

Short-term, localized, non-periodic fluctuations in hyper-temporal measurements are often obscured by background patterns in the data. The terms 'short-term' and 'localized' respectively refer to duration and spatial extent considerably smaller than the rest of the dataset. Such fluctuations are often of interest for geoscience applications based on detection of extremes and/or environmental monitoring. Potential examples include fires, volcanic and geothermal activity, fluctuations of climatic variables, urban heating incidents, leakage of pollutants, abrupt changes in vegetation, irrigation leakages, and weather extremes. All these phenomena would be recorded as fluctuations in a satellite signal. They may be expressed in different parts of the spectrum, evolve in different spatiotemporal scales, and they can influence the original signal without exceeding the range of normal values. They may occur regularly or unexpectedly, in known or unknown locations.

---

<sup>1</sup> Pavlidou, E., van der Meijde, M., van der Werff, H. M. A., & Hecker, C. A. [2016]. Finding a needle by removing the haystack: a spatio-temporal normalization method for geophysical data. 90(A), 78-86. DOI: 10.1016/j.cageo.2016.02.016

Usually, there is not enough information available on a fluctuation of interest to facilitate its isolation. In such cases, local fluctuations can be made more visible by suppressing patterns that are common to the majority of the dataset. Patterns may be suppressed by explicitly modelling and removing signal components, if characteristics of these components are a priori available. An alternative option would be the application of normalization techniques, which rescale the data, provide adjustment for overall patterns and allow the data to become internally comparable.

Popular choices to identify and remove general trends and periodic signal components, include autoregressive and ordinary regression models, filtering and decomposition techniques (for example, [Cleveland et al., 1990, West, 1997, Jonsson and Eklundh, 2002, Grieser et al., 2002, Alegana et al., 2013, Wang et al., 2014]). Wavelets and Fourier transforms are widely used to define signal patterns of different periodicity in a variety of geophysical applications, ranging from climatic studies to hazards and environmental research (e.g. [Meyers et al., 1993, Kumar and Foufoula-Georgiou, 1997, Ghil et al., 2001, Sajda et al., 2002, de Jong and van der Meer, 2004, Labat, 2005, Scharlemann et al., 2008, Humlum et al., 2011, Pyayt et al., 2013, Tary et al., 2014, Qader et al., 2015]). Randolph [2005] describes typical methods to normalize signals and/or images; the adjustments presented in his work include corrections for constant and non-constant shifts, scaling, and combinations thereof. Adaptations of the Standard Normal Variate method he mentions are often applied in geosciences to identify extremes (for example, [Tramutoli, 1998, Jiménez-Muñoz et al., 2013, Jiménez-Muñoz et al., 2015]).

These methods face three possible limitations. First of all, they cannot easily define signal components with a period longer than the temporal length of the available dataset. Secondly, large-scale patterns which are changing through time may still obscure fluctuations of very small intensity. A third limitation arises because existing methods rely only on the temporal dimension of the measurement. This is an issue especially relevant for research based on multidimensional data, as is the case of hyper-temporal satellite measurements. Satellite sensors provide synoptic, time-synchronous and consistent sampling of geophysical parameters over large areas and over long periods of time; both the temporal and spatial characteristics of the data are needed to extract fluctuations of these parameters from a complicated mix of different influences and noise. As noted by Tary et al. [2014] multidimensional geophysical data are traditionally analysed individually in a one-dimensional manner, but it would be very important to consider more dimensions in the analysis.

Spatial information is commonly used in satellite image processing. Such image processing techniques have been applied, among others, for feature extraction

e.g. van der Werff et al. [2006], Soto-Pinto et al. [2013]; change detection, e.g. Coppin et al. [2004], Canty and Nielsen [2012]; normalization of specific influences in the satellite signal, for example in Yang and Lo [2000], Canty and Nielsen [2008], Ulusoy et al. [2012]; active fire monitoring [Giglio et al., 1999, Giglio et al., 2003, Ichoku et al., 2003, Stolle et al., 2004, Kuenzer et al., 2007, Kuenzer et al., 2008, Calle et al., 2008, Xu et al., 2010, Wooster et al., 2012], and studies of volcanic and geothermal activity [Coolbaugh et al., 2007, Ganci et al., 2011, Koeppen et al., 2011, Murphy et al., 2011, Steffke and Harris, 2011, Vaughan et al., 2012, Blackett, 2014, Blackett, 2015, van der Meer et al., 2014, Gutiérrez et al., 2012]. A main aim is often the selection of an optimum background: this is used to contrast with an expected change or to describe a representative average state of a given neighbourhood. Subsequent processing and statistical analyses vary depending on the field of application.

Our work is a modification on aforementioned attempts to apply normalization for spatial data. We build on kernel-based approaches (for an overview see, e.g., [Canty, 2010]). We use a deconvolution matrix to select a pixel's spatial neighbourhood and normalize every pixel by the mean value of its neighbourhood set. In this way, patterns that are present in the central as well as the surrounding pixels are suppressed, and localized fluctuations are made more visible. We then apply a window-based retrieval technique to isolate these fluctuations in the normalized time series.

The added value of this approach is that spatial and spectral techniques are combined in a single algorithm and in a non-application-driven manner. The normalization can suppress regionally extended patterns at different time-scales (diurnal, seasonal, yearly etc.), based on both spatial and temporal components of the original data, and independently of the length of the dataset. Processing is run uniformly in the whole dataset and can isolate fluctuations which are not expected or known in advance. The processing chain is at the same time generic and flexible enough to be applied in different domains, and may be applied in near real-time mode.

We demonstrate our approach on a hyper-temporal geostationary thermal infra-red (TIR) dataset, recorded by the SEVIRI sensor onboard the Meteosat-9 satellite, and subset over two areas of different TIR variability in time and in space. We choose to base our case study on TIR data because of the wide range of earth-science related TIR applications: monitoring of fire and volcanic activity, geothermal exploration, and others [Sobrino et al., 2013; Ulusoy et al., 2012]. The resolution of the sensor (3×3 km spatial, 15-min temporal) supports temporally intensive monitoring. In real-life applications, it rarely happens that the same well-known fluctuation is repeated in different conditions. Thus, to be able to evaluate the performance of our algorithm in different contexts and with better control, we carry out experiments with

known synthetic fluctuations imposed in real data. We increase a small number of consecutive brightness temperature (BT) values in the original data and show that the normalization makes these increases more visible. We retrieve the fluctuations in the normalized time series and evaluate retrieval in reference to the intrinsic signal variability.

We then compare our findings with results of the application of widely used statistical and spectral methods. We decompose the TIR signal using Seasonal-Trend Decomposition based on LOESS (STL), following Cleveland et al. [1990] and Hafen [2010]. In this way we remove the dominant daily and seasonal component of the data and we try to detect the artificially imposed fluctuation in the remainder. As an alternative, we use Fast Fourier Transform to define all principal frequencies of the signal. Similar work has been done, for example, by Humlum et al. [2011] to study periodic oscillations in climatic records and by Wang et al. [2014] to study water level fluctuations. We remove the defined patterns and re-construct the signal based on the remaining frequency components, in order to trace the imposed fluctuation there. Finally, we apply the Standard Normal Variate method version which Jimenez-Munoz et al. [2013, 2015] used to detect anomalous thermal episodes over the Amazon. The technique rescales the data using their mean and standard deviation; thresholds are then applied to detect anomalous standardized values. The comparison between results of the different approaches concludes the performance evaluation of our method.

## **3.2 Methodology**

In the first part we present our method. We then shortly present the three spectral and statistical approaches we applied to evaluate its performance.

### **3.2.1 Proposed method**

Our proposed method consists of three steps: pre-processing, normalization and retrieval (Figure 3.1).

#### **Pre-processing**

Image pre-processing is intended to discard measurements that have been disturbed by factors other than the variable of interest. This step requires consistently defined a priori knowledge on the presence of disturbances. Pre-processing is specific to each application and dataset, and is not covered here in detail. In the case of most satellite imagery- based studies, the main factor interfering with the signal from the earth's surface is the atmosphere, and especially the presence of cloud cover.

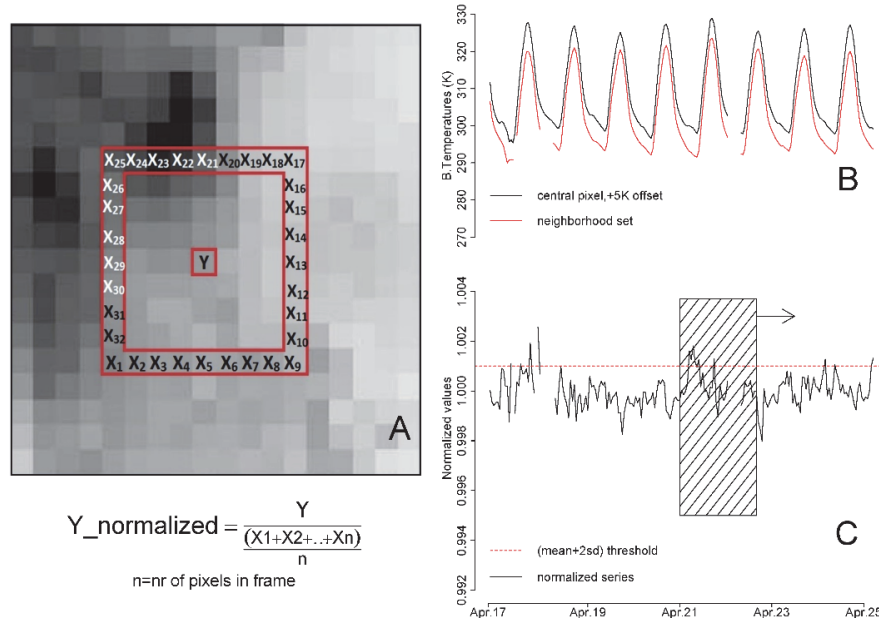


Figure 3.1. Methodology. (A) With normalization, every pixel value in the image is divided by the average value of a frame of neighbouring pixels. The pixels between the frame and the central pixel are not included to ensure that a localized fluctuation is not averaged out. (B) Original time series from central Pixel Y (black line) and the average value of its neighbourhood set (red line), with offset for clarity. (C) The same series after normalization. The daily pattern common in the central and neighbouring pixels is no longer dominant in the data. During retrieval, normalized values above the threshold are flagged. A temporal moving window counts the number of flags within a specific time period.

### Normalization

In this step we use spatial information to suppress temporal patterns common between a pixel and its neighbourhood, without explicitly modelling them. As a result, fluctuations which affect only the central pixel stand out.

We define a square, single-pixel-wide frame of neighbouring pixels (from here on, “neighbourhood set”, see Figure 3.1A). The neighbourhood set lies at a defined distance around a central pixel (Figure 3.1A, details on choosing the distance are provided later). The inside of the frame is not included in the neighbourhood set. The underlying assumption is that a localized fluctuation has a spatial extent smaller than the area framed by the neighbourhood set; it is thus contained in the central pixel, but not in its neighbours. We divide the central pixel value by the average value of the neighbourhood set (Figure 3.1B). The process is repeated for all time-steps and results in a spatially based normalized time-series, which retains the temporal resolution and serial dependence of the original time-series.

The normalized time series shows the degree of dissimilarity between a pixel and its surroundings. In that sense, it expresses local heterogeneity and the way this changes over time. When the central and neighbouring pixels are similar, the expected normalized value is approximately 1 and constant throughout time (Figure 3.1C). Normalized values deviate from 1 as the differences between the central pixel and the frame increase. These differences reflect the natural variability in the signal, which may be due to differential heating of the earth surface, atmospheric effects, differences in reflective/emissive properties of different land covers, soil moisture etc. An anomalous fluctuation in this context is defined as an unusual change in the relation between the pixel and its surroundings. This translates to normalized values largely deviating from the series mean.

Natural variability of the signal is statistically described by its standard deviation ( $\sigma$ ). Normalization is designed to minimize this natural variability by suppressing signal patterns which are common between the pixel and its neighbours. Short-term localized anomalous fluctuations are better visible when the rest of the normalized series has minimal variability. We thus use  $\sigma$  of the normalized series to evaluate performance of normalization and to choose the distance between the normalization frame and the central pixel. The  $\sigma$  of the normalized series is used to evaluate performance of normalization and to choose the distance between the normalization frame and the central pixel. We first calculate  $\sigma$  for each normalized time series separately. We then calculate the average  $\sigma$  of all normalized series in the dataset to describe variability for the whole study area ( $\sigma_{\text{dataset}}$ ). We choose as optimal the normalization distance which results to the lowest  $\sigma_{\text{dataset}}$ . Once set, the distance between central pixel and frame remains the same for all the pixels of the dataset. This setting is not built-in the code because every study area has different natural variability, potentially changing through time. By allowing the distance to be user-defined, we ensure flexibility of the method.

The statistical metric of  $\sigma_{\text{dataset}}$  is the main criterion to choose frame size in the case of unknown fluctuations. However, it should be critically assessed as well, especially if there is more information available on an expected fluctuation. The distance between central pixel and frame should be large enough that a localized fluctuation is not included in the neighbourhood set. If the aim is to detect an oil spill extending over hundreds of km<sup>2</sup>, it would be pointless to choose a 1-km distance (even if this resulted to the lowest  $\sigma_{\text{dataset}}$ ). Larger sets increase statistical consistency in the normalization process and may decrease the variability. However, if the size of the set is so large that the central pixel and its neighbours are incomparable (for example, if they belong to different climatic regions), there may be hardly any common patterns left for the normalization to suppress. In the cases mentioned above, the optimal choice would be the set which fulfils all provisions: lowest  $\sigma_{\text{dataset}}$  possible; and, frame



size larger than the expected spatial extent of the fluctuation but within the limits of the same climatic region.

Data availability may influence the result of normalization. In each time-step different parts of the normalization set may be missing. In case of extensive cloud cover there is a possibility that the few pixels remaining after cloud-masking are not representative of the complete neighbourhood set. A threshold is set to ensure that normalized values of consequent time-steps are comparable. The threshold discards normalized values which have been calculated with less than a specific portion of the pixels of the neighbourhood set. This threshold is optional. Statistical testing is applied to evaluate if its application is needed and to define its level.

### **Retrieval**

Retrieval highlights and temporally isolates periods in time when the normalized signal is dominated by values strongly deviating from the mean. Fluctuations which are not only localized but also temporally persistent are retrieved using a window-based approach. A temporal moving window can locate fluctuations in time and its application is independent of the process that may be generating the fluctuation [Chandola et al., 2012].

A mean +  $2\sigma$  threshold is applied on the normalized time series to distinguish the values which most strongly deviate from the mean. All values exceeding this threshold are flagged (Figure 3.1C) and from here onwards they will be referred to as anomalies. Consequently, a temporal moving window counts the anomalies falling within the specified duration of the window (Figure 3.1C). The temporal length of the moving window is defined by the user, depending on the application, the temporal resolution of the sensor and the desired level of detail in temporally locating the fluctuation. The resulting time series consists of the total number of anomalies per window and represents temporal clustering of highly deviating normalized values.

Also for this step the results of retrieval may be affected by data availability. Lack of data within the temporal window may result in low numbers of flags; these may be mistakenly considered to reflect low values in normalized data. To compensate for this effect, the number of anomalies in the temporal window is divided by the ratio existing observations/theoretical number of observations. Furthermore, a threshold is used to discard values which were calculated with less than a minimum number of available observations. This threshold is optional. Statistical tests determine if its application is needed and define its level (see following sections for more details).

### 3.2.2 Traditional methods applied for evaluation

We apply three commonly used spectral and spatial approaches.

First we use the Seasonal-Trend decomposition by LOESS (STL) [Cleveland et al., 1990; Hafen, 2010; Wang et al., 2014] which defines the daily and yearly components in the data. These components are subtracted by the original values. The remainder is examined to detect the imposed +3 K fluctuation.

We then apply Fast Fourier Transform (FFT) to define all principal frequency components in the dataset [Humlum et al., 2011; Wang et al., 2014]. FFT requires continuous series; Singular Spectrum Analysis (SSA) is used to eliminate missing data. This is a gap-filling method which preserves periodic patterns of the signal [Buttlar et al., 2014; Korobeynikov, 2010; Kondrashov and Ghil, 2006]. After performing forward-FFT, we remove the most dominant frequencies and use inverse-FFT to reconstruct the signal. We expect to detect the imposed +3 K increase in the reconstructed signal.

Finally, a version of Standard Normal Variate method is applied, as recently presented in Jimenez-Munoz et al. [2013, 2015]. Following this approach, standard scores are calculated from original values to show how many standard deviations is the distance between each observation and the mean of the series:

$$BT_{standardized} = \frac{BT_{original} - BT_{mean}}{\sigma}$$

The authors classify standard scores to the following levels of warming (standardized score range in brackets): abnormal [+0.5, +0.8], moderate [+0.8, +1.3], severe [+1.3, +1.6], extreme [+1.6, +2.0] and exceptional [+2.0 and higher]. The probability of a score being anomalous is up to 57.6% for the abnormal warming level, up to 80.4% for moderate, up to 86% for severe, up to 95.4% for extreme and more than 95.4% for exceptional warming levels [Jimenez-Munoz et al., 2015].

We apply this method in the three recommended scales: monthly (June), seasonal (May, June and July), and yearly (2011).

## 3.3 Application and evaluation

In the first part of this section we demonstrate application of our method on two spatial subsets of a year-long satellite TIR dataset. The two subsets represent study areas of different homogeneity (e.g. in local weather, land cover, and anthropogenic activities, see Figure 3.2). We impose in both datasets artificial fluctuations in the form of increased brightness temperatures (see Table 1 for details). We describe how we defined the settings for the

processing. We evaluate the choices in terms of their effect on single-series  $\sigma$  and  $\sigma_{dataset}$ . We then retrieve the imposed anomalies and evaluate performance of the method. In the second part of the section we process the same data with three traditional approaches and compare results with our method to evaluate it further.

### 3.3.1 Input

We used TIR imagery acquired from the Spinning Enhanced Visible and Infra-Red Imager (SEVIRI) onboard EUMETSAT's Meteosat-9 geostationary satellite, positioned at  $0^\circ$  /36,000 km. The instrument has a nominal spatial resolution of  $3 \times 3 \text{ km}^2$  at nadir and a sampling rate of 15 minutes in the TIR channels. We used two spatial subsets of a year-long (2011) whole-disk dataset from channel 9, registered at a wavelength range of  $9.8\text{--}11.8 \mu\text{m}$  ( $\lambda_{cent} = 10.8 \mu\text{m}$ ). Original top-of-atmosphere radiance values are converted to Brightness Temperatures (BT, in K, following Clerbaux [2006]). The first study area (spatial extent:  $327 \times 303 \text{ km}$ , or  $109 \times 101$  pixels) is located in the desert in Niger, and serves as an example of a homogeneous background with low spatial and temporal variability in the TIR. The second study area is located in Kenya (spatial extent  $309 \times 318 \text{ km}$ , or  $103 \times 106$  pixels) and is very diverse in terms of geomorphology and land cover, representing a very heterogeneous background.

### 3.3.2 Pre-processing

We masked cloud-affected pixels using masks of EUMETSAT's Climate-Monitoring Satellite Application Facility (CM-SAF). This dataset was produced with software of the Nowcasting SAF [Derrien and Le Gléau, 2005]. We excluded cloud-filled and cloud-contaminated pixels from further processing. As some of the clouds in the dataset were not detected by the available masks, we further discarded remaining pixels with values lower than the lowest recorded temperatures in historical archives.

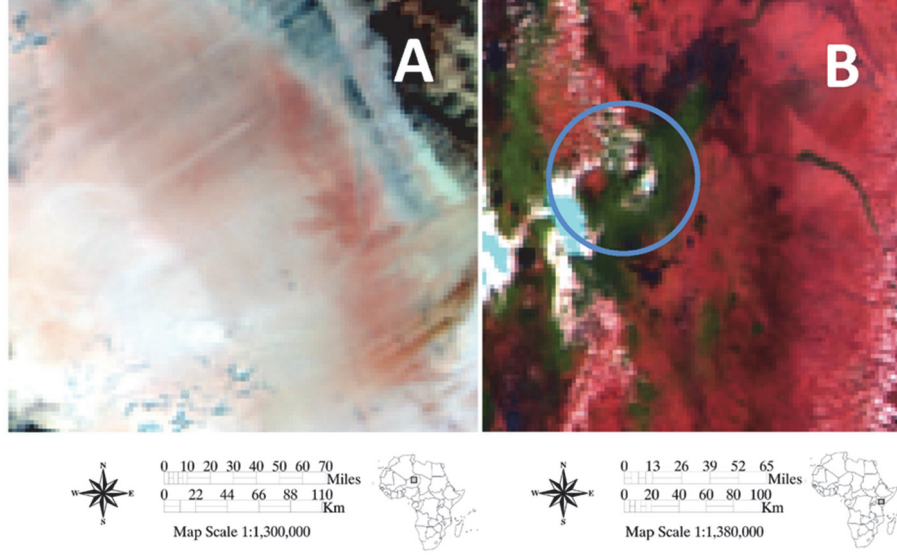


Figure 3.2. Study areas. Natural color RGB images from the study areas in Niger (A) and Kenya (B). The area in Niger is almost completely covered by desert consisting of sand and gravel, with the exception of the rock formations in the NE corner of the image. Kenya has a more complex cover including bare soil, vegetated land, rock formations, water bodies, urban areas and mountains. The main feature is Mount Kenya circled in blue, in this image partly covered by clouds (white and turquoise).

### 3.3.3 Results and evaluation of our method

We first present and evaluate the choice of settings applied in the processing; then we evaluate performance of retrieval. All the performed tests are described in Table 3.1. Experiments 1–4 (Table 3.1) were designed to test the following settings: size of the neighborhood set, data availability thresholds, and length of retrieval window. These settings are application-dependent; the choices we present are specific for this case study. Experiments 5–8 (Table 3.1) were designed to test performance of normalization and retrieval of synthetic fluctuations.

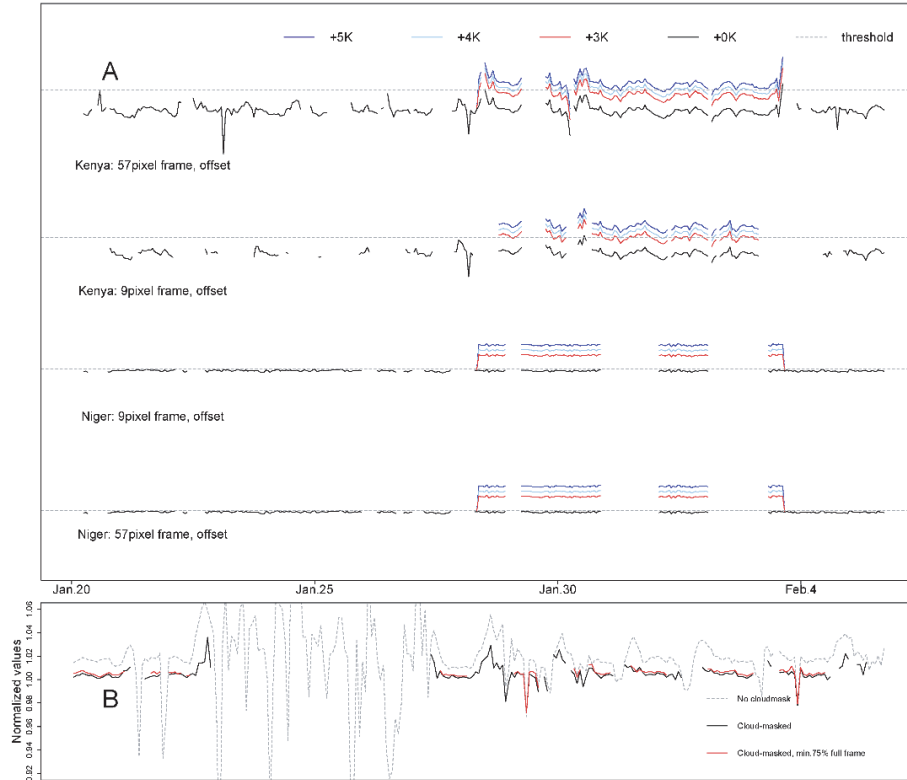
#### Settings

The first parameter to test was the optimal size of the normalization set (Experiment 1, Table 3.1). We performed normalization with different sizes of neighborhood sets, their frame sides ranging from 9 to 73 pixels (respectively 12–108 km shortest distance between each side and the central pixel). We selected the size with the minimum  $\sigma_{dataset}$ . For Niger and Kenya the statistics were different but in both cases we had minimum  $\sigma_{dataset}$  when applying a 57-

pixel frame side. The  $\sigma_{dataset}$  was increased from 0.001 (57-pixel frame side) to 0.003 (9-pixel frame side) in Niger, and from 0.003 to 0.006 respectively in Kenya (Fig. 3.3A).

Table 3.1 Experiments and related choices

Objective	Experimental setting
<b>Choice of settings</b>	
<b>1. Size of normalization set: Prefer the size which results to normalized series with the lowest standard deviation</b>	Use frame side of 9, 17, 25, 33, 41, 49, 57, 65, 73 pixels, corresponding to approximate radii ranging between 13 and 108 km.
<b>2. Retrieval with temporal windows of different length: Prefer windows of similar duration as the fluctuation of interest</b>	5-day long artificial fluctuation retrieval with 2-, 4-, 7-, 10- and 20-day windows.
<b>3. Data availability threshold for normalization: Choose the minimum data availability that preserves the distribution of values in the set</b>	Normalize with sets of minimum 30%, 50%, 60%, 70%, 75%, 80%, and 100% data availability. Record standard deviation and distribution of values within the set.
<b>4. Data availability threshold for retrieval: Choose the minimum data availability that preserves the distribution of values in the series leaving enough data for analysis</b>	Retrieval with minimum 10%, 20%, 30%, 35%, 40%, 50%, 60%, and 80% of data available.
<b>Experiments on normalization and retrieval</b>	
<b>5. Spatial extent of fluctuations suppressed by the normalization</b>	Impose +3 K fluctuation in (a) 1 pixel, (b) 10 pixels and (c) 7411 pixels (67% of the image).
<b>6. Evaluate intensity of retrieved fluctuations</b>	Impose and retrieve fluctuations of +1, 2, 3, 4, 5, and 6 K.
<b>7. Retrieve fluctuations in different locations in the image and in the time series</b>	Impose and retrieve the same +3 K fluctuation. (a) In timeslots with/without gaps (February, June, September, and November) (b) At locations with high/low variability in the normalized series.
<b>8. Effect of clouds on normalization</b>	Process data with and without cloud masking.



**Figure 3.3. Effect of algorithm settings and missing data. (A) Effect of frame size.** The panel shows the result of normalization of series with different imposed fluctuations, using different frame sizes, in both study areas. The series are displayed with offset for clarity. Normalized series in the heterogeneous area of Kenya (first two rows) are more variable than in homogeneous Niger (last two rows). Imposed fluctuations of +(3–5) K exceed the detection threshold regardless of frame size, and are more visible when the variability is low. Artificial fluctuations increase linearly the normalized values. **(B) Effect of clouds and missing data.** The presence of clouds increases variability in the original time series (grey dashed line); even after cloud masking (black line), cloud remnants introduce spikes. Lack of data in the neighborhood set also increases variability of cloud-masked normalized series. In contrast, the series normalized with at least 75% of the data available (red solid line) is the least variable.

We tested retrieval of a 5-day artificial fluctuation using temporal windows of 2–20 days. The fluctuation was retrieved as an anomaly in all of the cases. The main differences in retrieval were the timing, the width and the size of the resulting peak. For all windows the anomaly starts to be visible at the same time, but the center of the peak shifts to later times for larger windows. The peak is higher for longer windows since more anomalous values are captured in a single window. The optimum window length is roughly the same as the anticipated length of the anomaly. If this is not known in advance, it is

recommended to test multiple windows and assess the persistence of peaks between short(er) and long(er) windows.

The application of data availability thresholds is not relevant for datasets which are not affected by missing values. We calculated the  $\sigma_{dataset}$  at different data availability levels to see if the use of these thresholds would be needed in this study. We found that a 20% decrease in data availability doubled the  $\sigma_{dataset}$  of Niger; in Kenya the corresponding  $\sigma_{dataset}$  became five times higher. We thus decided that it would be relevant to apply the thresholds in order to minimize this variability. The choice on these thresholds is based on two criteria: (a) preserve the distribution of values in the normalization set/temporal window and (b) make sure that enough data remain for analysis. We decided that a minimum of 75% data was needed for normalization, and a corresponding 25% for retrieval. The reduction in variability of normalized series after application of a 75% threshold can be seen in Fig. 3.3B.

The examples we show in the figures are obtained with the following optimal settings: a normalization set of a 57-pixel frame side, with at least 75% of the pixels present; and a 7-day long temporal window, with at least 25% of the observations present.

### Retrieval performance

With Experiment 5 (Table 3.1) we tested the hypothesis that fluctuations of extent smaller than the normalization set are detectable. We imposed the same +3 K fluctuation in areas of different spatial extent. As expected, the fluctuation was visible when its extent was smaller than the area framed by the normalization set but not when its extent was larger. In Fig. 3.4, artificial fluctuations of +3 K which cover both the central pixel, coincide after retrieval with the +0 K imposed fluctuation (black line).

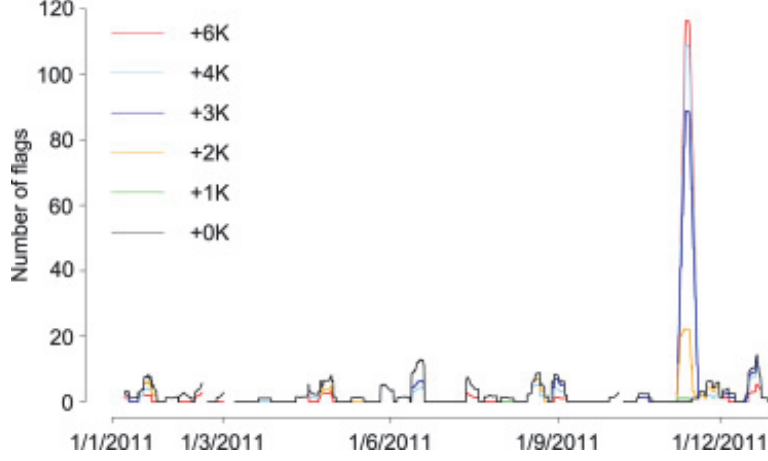


Figure 3.4 Detection of synthetic anomalies of different magnitude (Niger). Localized artificial fluctuations are retrieved as peaks of different magnitude.

In Experiment 6 (Table 3.1) we imposed fluctuations of magnitude (+1) to (+6) K in the data (Fig. 4 for Niger). Regardless of the size of the fluctuation, higher fluctuations linearly increased normalized values (Fig. 3.3A). In the retrieval the effect was non-linear. The smallest imposed fluctuations we could retrieve, under conditions of high variability and low data availability, were +2 K in Niger and +3 K in Kenya (Fig. 3.4).

We imposed fluctuations in different locations in space and in time, to examine cases with higher and lower variability and with more or less missing values (Experiment 7, Table 3.1). Retrieved anomalies had different amplitudes when the same fluctuation was imposed in different times. Peaks were in general larger when the variability of the time series was lower and data availability was higher. As a result, we cannot quantify that a specific fluctuation will be always retrieved as an anomaly of corresponding amplitude; this should be evaluated in the context of each study area.

Our last experiment (Experiment 8, Table 3.1) tested the effect of cloud masking. Cloud masking reduced the variability of the normalized series approximately by a factor of 4 (Fig. 3.3B). The value of  $\sigma_{\text{dataset}}$  changed from 0.020 to 0.005 after cloud masking in Niger and from 0.037 to 0.009 in Kenya. In contrast, the increase of  $\sigma_{\text{dataset}}$  that was caused by localized fluctuations, like the ones we imposed in both datasets, was of the order of  $10^{-4}$ . This shows the importance of removing disturbances from the data prior to processing.



### 3.3.4 Comparison with other methods

We used the original dataset from Niger and imposed a +3 K 5-day increase in BT in June 2011. This had proven straightforward to retrieve with our method, so we tested detecting it with others.

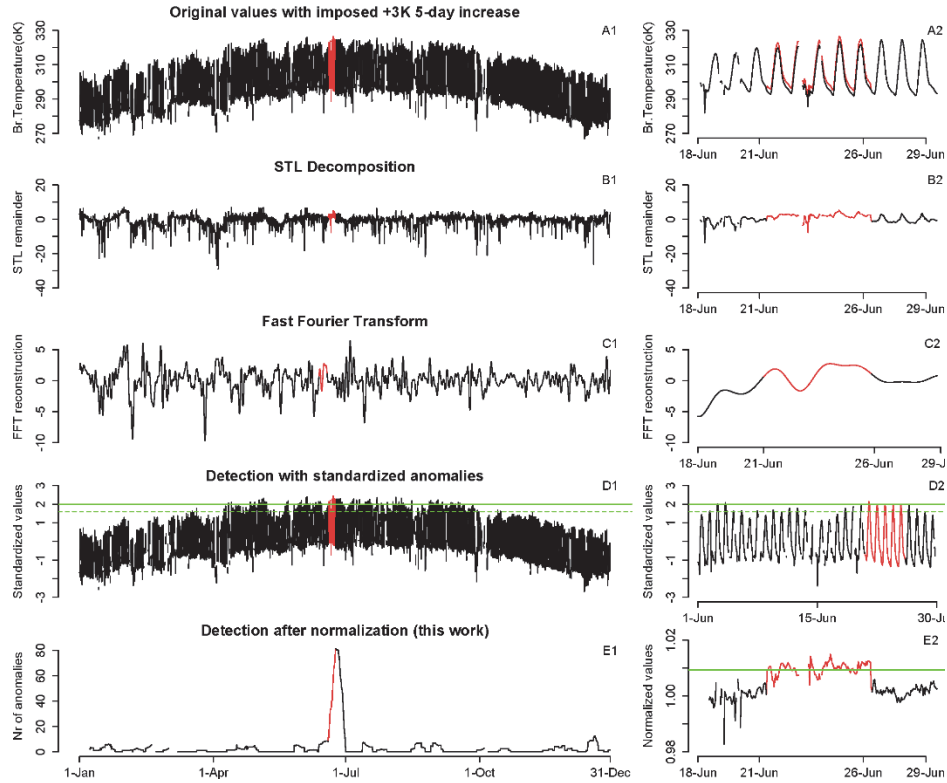


Figure 3.5 Comparisons between detection methods. Application of different methods to detect a synthetic +3 K 5-day BT increase imposed in the Niger dataset. A1: Original time series, with the imposed anomaly in red. The imposed anomaly does not stand out from the rest of the data. A2: the same (detail). B1: Remainder of STL decomposition of the original series. The anomaly is not distinguishable. B2: The same (detail). C1: FFT-reconstruction of the original series, after removing the three dominant frequency components (yearly, daily and twice-per-day). The impose anomaly does not stand out. C2: The same (detail). D1: Time series of the standard scores calculated from the original year-long dataset. The green dashed line designates extreme values, and the green solid line shows values which are classified as exceptional. The imposed BT increase is classified as anomalous but cannot be distinguished from the rest of the extreme/exceptional standardized values in the dataset. D2: The same but in monthly scale. E1: The methodology described in this paper clearly identifies the imposed anomaly as the single highest peak in the dataset. E2: Normalized dataset (detail). The imposed anomaly stands out in the normalized series, largely exceeding the detection threshold (in green).

Results from the application of all methods are summarized in Figure 3.5. The first row shows the original series with the imposed fluctuation. The second row shows the remainder of STL decomposition. This is the result of subtracting the seasonal and daily patterns from the original series. The artificial fluctuation (in red) is contained in the remaining values, but cannot be distinguished from the rest of the data (Figure 3.5B1,2). In the case of FFT, the dominant frequency components identified are yearly, daily, and twice-per-day. These components are removed and the signal is reconstructed using inverse FFT. The imposed anomaly does not stand out in the reconstructed signal (Fig.3.5C1,2).

Detection with standardized scores in a yearly scale yields extreme and exceptional anomalies throughout the period between mid-April to October (Figure 3.5D1). Standardized data retain the seasonal pattern of the original dataset. The imposed fluctuation is included in the values which are considered extreme/exceptional, but so were many other values. As a result, the imposed anomaly does not stand out. Similar results are obtained with analysis on seasonal and monthly scale (Figure 3.5D2).

In contrast, after applying the normalization described in this study, the imposed +3 K 5-day increase is detected as the single most prominent peak in the whole dataset (Figure 3.5E1.). The majority of anomalous values exceeds the detection threshold already in the normalized series (Figure 3.5E2).

### **3.4 Discussion**

Fluctuations of geophysical parameters are of interest in applications ranging from climatic studies to hazard monitoring. Detection of such fluctuations is often a complicated task, even with the variety of available satellite sensors. Depending on the parameter of interest, input from multiple parts of the spectrum may be needed; issues of saturation may arise in different scales; and fluctuations may be obscured by predominant patterns.

Localized fluctuations of interest may remain within the expected range of original values. In such cases, pixel values are normal in absolute value, and the fluctuation is easily obscured by predominant (daily, seasonal) patterns of the signal. Such fluctuations can be detectable with examination of each value in its spatial context [Byun et al., 2007]. For example, a forest fire which would saturate the TIR channels of high-resolution sensors, may increase the pixel value of low-resolution TIR imagery only by a few degrees Kelvin [Wooster et al., 2005]. This increase is localized and alters the usual relationship between the affected pixel and its surroundings. Our methodology targets such fluctuations by identifying anomalies as deviations from the usual relation between a pixel and its spatial neighbourhood. This approach facilitates

detection of contextual anomalies which remain within the normal range of values. Furthermore, it allows detection of fluctuations over constantly elevated or constantly low background. This could be the case, for example, of temporally variable volcanic activity in a permanent lava lake.

Considering the spatial context of an anomaly offers another advantage: it constrains its occurrence locally, providing insight on the relation between the anomaly and the potential underlying causes. An anomaly that extends in thousands of  $\text{km}^2$  can most likely not be attributed to a spatially limited process. In that respect, the shape of the normalization frame serves as an upper boundary for the areal extent of a detected anomaly. The use of an open frame ensures that only a spatially limited fluctuation is detectable: if the fluctuation was included in the normalization set, it would be averaged out when the value of an anomalous pixel is divided by the value of its equally anomalous surroundings. As shown in Experiment 5, anomalies are only detected when they are not covered by the frame.

In terms of the temporal dimension of the data, our method suppresses patterns not by modelling past observations, but by identifying the commonalities between signals of neighboring areas in the time when they emerge. This is especially important because predominant patterns, like seasonality, often vary in time. For example, weather extremes are becoming more frequent [Easterling et al., 2000]; by using the approach described in this Chapter, a short-lived localized Brightness Temperature increase caused by spatially finite surface processes can be distinguished from a short-term, unusually warm period affecting the whole TIR image. Present-time normalization of commonalities allowed the methodology of this Chapter to outperform approaches which do not fully address predominant patterns or rely on past observations to suppress them. Another advantage is that processing can take place in near-real-time mode. Additionally, by defining the temporal occurrence of an anomaly, it is possible to better evaluate its relation with potential causative processes.

Rather than pre-defining the settings of our method, we statistically determine specific criteria for the choices. In this way we ensure flexibility of the method and optimal application in study areas with different local conditions. Information on the characteristics of the fluctuation of interest can be used to constrain application of the methodology, but it is not required to have *a priori* knowledge of the expected fluctuation or the predominant patterns in the signal. Settings are applied uniformly to the whole dataset, and retrieval is automated, supporting objectivity and comparability of the results. The  $mean+2\sigma$  threshold that isolates anomalies is based on each single-pixel time series; that has the advantage that anomaly detection is performed considering the local conditions. Normalized series of high variability have

higher  $mean+2\sigma$  threshold; then, values need to be higher before they are declared anomalous. Datasets of long duration are recommended, to better describe the usual relation between a pixel and its surroundings.

Our results show that the proposed methodology can detect short-term (lasting 1.4% of the duration of the dataset), localized (affecting only one pixel) and low- intensity anomalies (as low as 0.7% of the signal, which was the case of +2 K over a 300 K background in Niger). This facilitates use of low spatial resolution sensors for detection of small-scale environmental changes, even with only one band available. The case study presented as an example was based on geostationary TIR input. Geostationary products offer the advantage of high temporal coverage, which is important for timely monitoring of extremes [Giglio et al., 2003]. The use of hyper-temporal datasets is further emphasized because temporal resolution is required to detect short-lived fluctuations. The approach, however, is not restricted to the input of our example; it may be applied at different scales and different types of spatiotemporal data. Higher spatial resolution may also provide more detail in locating a potential source of anomaly.

### **3.5 Conclusions**

We focus on unexpected, localized, short-term signal fluctuations and propose a methodology to detect them using single-band input. Our normalization procedure suppresses spatially extended, large-scale temporal patterns in single-pixel time series without having to explicitly model them. Data are brought to scale and localized fluctuations with an extent smaller than the defined become more obvious, regardless of the underlying causative processes. The fluctuations are distinguished from large-scale periodical patterns by analysing both the spatial and temporal dimensions of geophysical data. This can serve a wide spectrum of applications and facilitate monitoring of extremes.

## **Chapter 4. Study of Volcanic Activity at Different Time Scales Using Hypertemporal Land Surface Temperature Data<sup>2</sup>**

### **4.0 Abstract**

We apply a method for detecting subtle spatiotemporal signal fluctuations to monitor volcanic activity. Whereas midwave infrared data are commonly used for volcanic hot spot detection, our approach utilizes hypertemporal longwave infrared-based land surface temperature (LST) data. Using LST data of the second-generation European Meteorological Satellites, we study (a) a paroxysmal, 1 day long eruption of Mount Etna (Italy); (b) a prolonged, 6 month period of effusive and lateral lava flows of the Nyamuragira volcano (Democratic Republic of Congo); and (c) intermittent activity in the permanent lava lake of Nyiragongo (Democratic Republic of Congo) over a period of 2 years (2011–2012). We compare our analysis with published ground-based observations and satellite-based alert systems; results agree on the periods of increased volcanic activity and quiescence. We further apply our analysis on mid-infrared and long-infrared brightness temperatures and compare the results. We conclude that our study enables the use of LST data for monitoring volcanic dynamics at different time scales, can complement existing methodologies, and allows for use of long time series from older sensors that do not provide midwave infrared data.

### **4.1 Introduction**

In this work we explore the utility of land surface temperature data, retrieved from longwave infrared (LWIR) records of geostationary satellites, to detect different types of volcanic activity. For this purpose we apply a recently published methodology [Pavlidou et al., 2016], which builds up on kernel-based image processing approaches, to allow utilization of satellite LWIR archives for long-term volcanic monitoring.

Monitoring of remote inaccessible volcanic areas at different timescales has been made possible by the availability of synoptic satellite coverage, leading to the gradual development of the quantitative discipline of satellite volcanology [Ramsey and Harris, 2013, Blackett, 2014]. Infrared (IR) observations from volcanic targets are reported since the 1960s [Wooster and

---

<sup>2</sup> Pavlidou, E., Hecker, C., van der Werff, H., & van der Meijde, M.[2017]. Study of volcanic activity at different time scales using hypertemporal land surface temperature data. *Journal of Geophysical Research: Solid Earth*, 122. DOI: 10.1002/2017JB014317

Rothery, 2000, Wright et al., 2004]. Instruments of different dynamic range, spatial and temporal resolutions have been utilized for volcanic studies [Donegan and Flynn, 2004, Pieri and Abrams, 2004, Watson et al., 2004, Blackett, 2014]. Technical characteristics of thermal sensors commonly used in volcanic applications can be found, for example, in Pieri and Abrams [2004]; Blackett [2014], [Blackett, 2017]; Ramsey and Harris [2013].

The theoretical base of volcanic IR remote sensing stems from the relationship between the kinetic temperature of an object and its spectral radiance (energy radiated from the object per unit of wavelength), as expressed by Planck's Function [Blackett, 2017, Wooster and Rothery, 2000]:

$$L_{\lambda}(T) = \frac{2hc^2}{\lambda^5 (\exp(\frac{hc}{\lambda kT}) - 1)} \times 10^{-6} \quad (\text{Eq.4.1})$$

where:

- $\lambda$  : wavelength (m),
- T : temperature (K),
- L : spectral radiance (W /m<sup>2</sup> /sr / $\mu$ m),
- h = 6.6  $\times 10^{-34}$  Js (Planck's constant)
- k = 1.38  $\times 10^{-23}$  J/k (Boltzmann's constant)
- c = 3  $\times 10^8$  m/s (velocity of light in vacuum)

An implication of Eq.1 is that the emittance of warmer objects peaks in shorter wavelengths (known as Wien's law, e.g. in Wooster and Rothery [2000]). Active lava flows have an average temperature in the range of 500-1000 K and their thermal emissions are stronger in shorter infrared wavelengths. The midwave infrared part of the spectrum (MWIR, 3-5  $\mu$ m) is particularly suitable for detection of volcanic features, as it is less affected by solar irradiance than daytime shortwave IR and allows detection of a wider range of volcanic activity [Blackett, 2013]. Wright et al. [2002] report an example of 200% increase in spectral radiance measured in the 4 $\mu$ m region (MWIR) in the presence of magmatic material as opposed to the less pronounced channel response (up to 1% increase) in the 11 $\mu$ m region (longwave infrared, LWIR). As a result, mainstream remote sensing methods for volcanic hotspot detection utilize MWIR input.

Recent overviews of volcanic hotspot detection and monitoring of volcanic activity can be found, for example, in Harris et al. [2016] and Blackett [2017]. Steffke and Harris [2011] classify relevant algorithms in three main categories. In the first category, fixed threshold/spectral techniques utilize the difference in MWIR-LWIR band sensitivity to temperatures of molten material. A characteristic example of dual-band approach is MODVOLC, an automated

hotspot detection algorithm based on input from the Moderate Resolution Imaging Spectroradiometer (MODIS). MODVOLC (accessible through [modis.higp.hawaii.edu](http://modis.higp.hawaii.edu)) has been a point of reference for more than a decade, issuing near-real time alerts on a global scale [Wright et al., 2002, Wright, 2015]. It flags thermally anomalous pixels based on the Normalized Thermal Index (NTI) which is the ratio between the difference and the sum of MWIR and LWIR radiances [Wright et al., 2004, Wright, 2015]. Also based on MWIR input, the MIROVA system provides near-real time monitoring and radiative power time series using MODIS data (Coppola et al. [2016]; accessible through [www.mirovaweb.it](http://www.mirovaweb.it)). The HOTVOLC system uses METEOSAT data for monitoring lava, ash and SO<sub>2</sub> (Gouhier et al. [2016]; accessible through [hotvolc.opgc.fr](http://hotvolc.opgc.fr)). In the second category, contextual approaches detect hotspots by comparing target pixels to a non-volcanic background. An early example in this category is the VAST method introduced by Harris et al. [1995]. Flasse and Ceccato [1996], Murphy et al. [2011], Blackett [2013] and Carr et al. [2016] follow similar approaches. In the third category, temporal (or time-series based) methods statistically identify anomalies using past observations of the same pixel (see, for example, Di Bello et al. [2004] and Pergola et al. [2004]). Hybrid approaches have been developed to increase performance and provide better description of lava flows in terms of emplacement and radiative power. Examples include the work of Higgins and Harris [1997]; Kervyn et al. [2008]; Koeppen et al. [2010]; Ganci et al. [2011].

Even though the first space-borne volcanic observation was registered by the MWIR channel of the Nimbus-1 meteorological satellite in 1964 [Wooster and Rothery, 2000], MWIR observations are not always available [Blackett, 2014]. Meteorological satellites did not routinely provide MWIR coverage until the late 1990s. In order to obtain information from older generation sensors, LWIR data can be used instead. Moreover, as mentioned by Carter et al. [2008], LWIR input can be used to monitor subtle radiance changes over time, especially where it comes to cooler volcanic features. The spectral response of cooler features is more pronounced in longer wavelengths. Reath et al. [2016] for example, combine LWIR data from different sensors (Advanced Very High Resolution Radiometer-AVHRR and Advanced Spaceborne Thermal Emission and Reflection Radiometer-ASTER) to detect subtle thermal output preceding effusive and explosive eruptions in Kamchatka. Wessels et al. [2013] used, among others, ASTER-derived atmospherically corrected at-surface LWIR radiance and surface kinetic temperature data to detect thermal unrest prior to the 2009 eruption at Redoubt volcano. Still, atmospheric influences need to be addressed because they might obscure subtle anomalies [Watson et al., 2004, Blackett, 2014, Carr et al., 2016].

We study volcanic events at different timescales without MWIR input, based entirely on LWIR-derived Land Surface Temperature data (LST). This is

particularly relevant in view of the availability of a consistent LWIR-based LST dataset retrieved from the European Meteorological Satellites' (METEOSAT) geostationary set of sensors [Duguay-Tetzlaff et al, 2015]. LST data derived from geostationary satellite input have high temporal resolution (24- 192 images daily depending on the sensor and the period), and are corrected for atmospheric effects and clouds. Because of the coarse spatial resolution of the geostationary sensor (3-5 km at nadir, depending on the sensor), the emittance of subpixel volcanic targets is averaged over larger areas and saturation of the recorded signal is not common. Furthermore, METEOSAT sensors have been in operation already for more than 30 years covering half of the planet. These characteristics allow for the construction of time series which (a) are hypertemporal, thus suitable for monitoring transient volcanic events, and (b) span over decades, thus being more useful to study long-term characteristic volcanic cycles.

The spatial resolution and the lower sensitivity of LWIR to volcanic material require the use of a detection method able to distinguish subtle signal variations; such variations should be detected also in case of constant activity. To achieve these, the method described in Chapter 3 is applied. It combines time series analysis with contextual approaches, allowing for site-specific detection. It normalizes pixel values by their spatial neighbours, suppressing spatially extended patterns at the time they emerge, independently of past observations. As a result, subtle variations in the signal are visible even in case of persistent activity.

We show the applicability of this approach using hypertemporal LST datasets which cover three different volcanic events. The first is the short paroxysmal Mount Etna (Italy) eruption of August 12th, 2011. The second event is a series of lava flows from Nyamuragira volcano and the third, volcanic activity the permanent lava lake of Niyragongo; both are located in Virunga National park (Democratic Republic of Congo) and the sequence of volcanic activity took place between November 2011 and April 2012. This choice of events allows us to test activity evolving in different timescales. The findings are validated using satellite and ground-based studies and reports. Finally, the same methodology is applied on non-atmospherically-corrected at-sensor Brightness Temperatures (BT), derived from METEOSAT MWIR (Channel 4) and LWIR (Channel 9) data over Mount Etna. The findings are intended to explore atmospheric effects on the detection and complementarity in the use of different wavebands.



## 4.2 Materials and Methods

### 4.2.1 Data

We use the LST product of the Land Surface Analysis Satellite Application Facility of EUMETSAT (LSA-SAF) [Trigo, 2009]. The SEVIRI sensor onboard the METEOSAT Second Generation (MSG) geostationary satellites has a 3x3km nominal spatial resolution and sampling frequency of four images per hour. The LSA-SAF team retrieve LST based on clear-sky measurements following the Generalized Split-Window (GSW) algorithm [Wan and Dozier, 1996]. First, they perform cloud masking using software of the Nowcasting Satellite Application Facility (NWC SAF). Then, as described by Trigo et al [2009] they use an adjusted version of the GSW algorithm to correct the data for atmospheric effects based on the differential absorption in two LWIR bands, centred at 10.8 $\mu$ m and 12 $\mu$ m. The LST product comes with pixel-by-pixel information on estimated uncertainties of LST values. LST data are available at three levels of confidence based on these uncertainties: above nominal (estimated uncertainty less than 1K), nominal (uncertainty between 1 and 2K) and below nominal (uncertainty above 2K).

For our analysis we only use only pixels with nominal or above-nominal uncertainty levels (2 K) are used for analysis in this study. LST data are subset over the two study areas. The first is located in Sicily, Italy and covers 66x65 pixels over the volcano of Mount Etna (Figure 4.1a). The dataset is one-month long (August 2011) and covers a paroxysmal episode that took place on Mount Etna, on the 12th August 2011. The second area (122x108 pixels of the image) is subset over the Virunga mountain range National park and includes two major active volcanoes on the African continent: shield volcano Nyamuragira and stratovolcano Nyiragongo, which hosts a permanent lava lake (Figure 4.1b). This dataset extends over two years (January 2011- December 2012) and covers the Nyamuragira eruption period between November 2011 and April 2012. In this period lava flows and activity in the lava lake took place, as described in the reports of the Global Volcanism Program (GVP) of the Smithsonian Institution [GVP 2011a-d; 2013; 2014a-b]. The spatial extent of the subsets is chosen to completely contain the corresponding volcanic targets, including also non-affected areas.

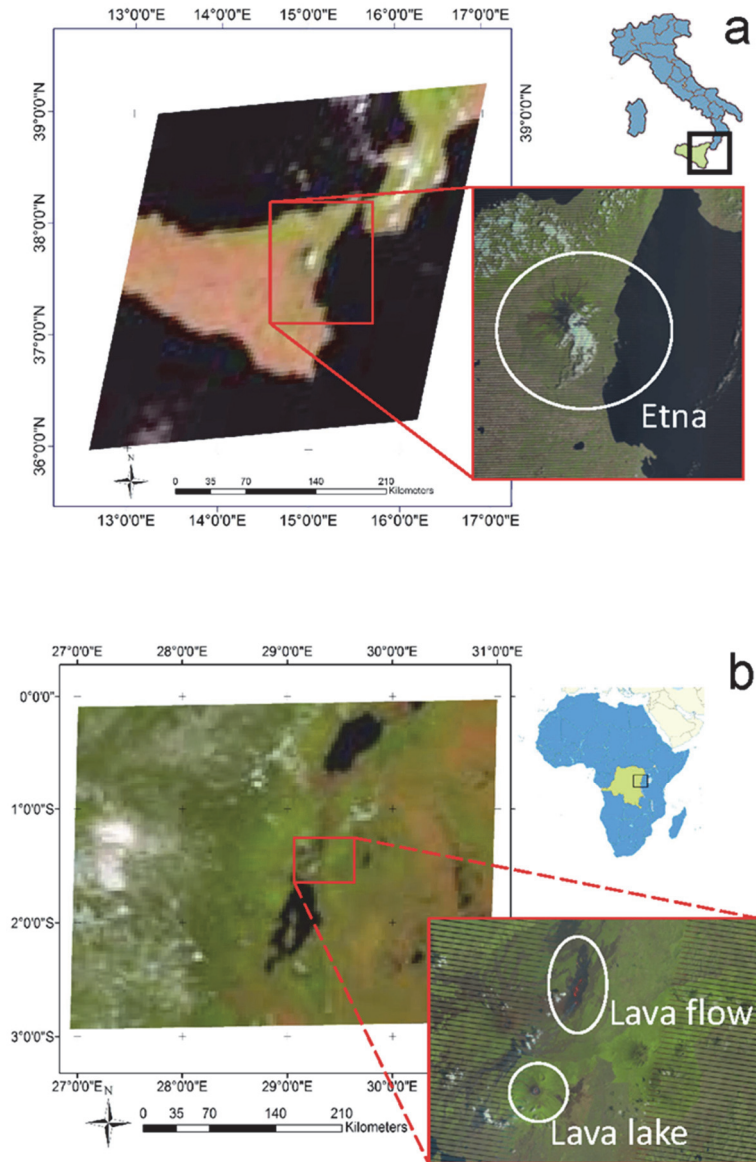
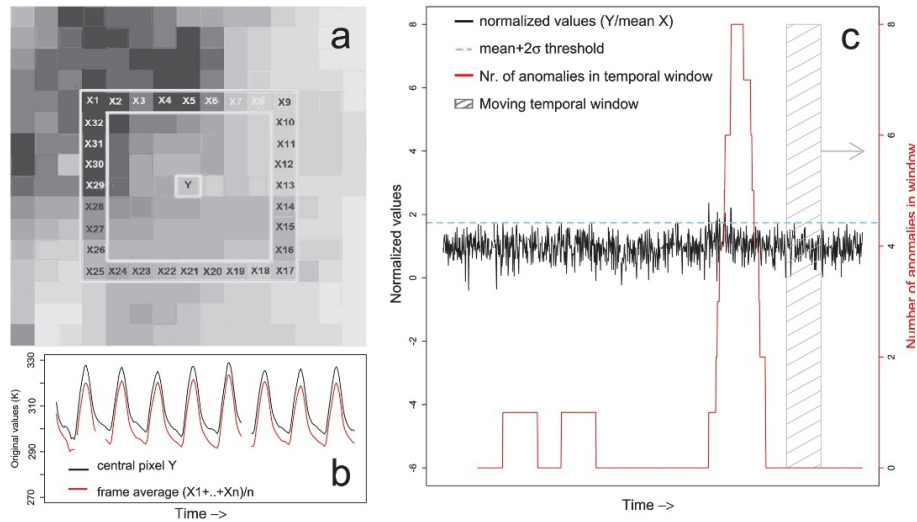


Figure 4.1. Study areas over Mount Etna (panel a) and Virunga National Park, D.R. Congo (panel b), with the general location indicated by the insets. In both panels, the images on the left are natural color RGB color composites from SEVIRI, while on the right the Landsat ETM+ color composite images show the volcanic targets with more spatial detail. High resolution imagery is only available at infrequent intervals (once every ~16 days). Green colors show vegetated areas, brown refers to bare soil, black to water bodies, white to cloud cover. The presence of lava is shown in bright red in the Landsat image over Congo, the only high-resolution image available from that period of activity (February 2012). The parallel black lines in the Landsat imagery are scan line corrector errors. Landsat imagery courtesy of USGS; SEVIRI imagery courtesy of EUMETSAT.

With this choice of events we test three types of volcanic activity, with different duration and spatial extent, and with different degree of ground-based monitoring. First, the paroxysmal eruption of 12th August 2011 in Mount Etna lasted 33hrs and comprised of Strombolian activity, lava fountaining and lava flows [Ganci et al., 2013]. Ganci et al. [2013] report ground-based observations covering the whole duration of the event, which was the strongest of the year. They use data acquired from a permanently installed thermal camera to estimate a total lava flow area of  $1.3 \times 10^6 \text{ m}^2$ , where temperatures developed between 400-1200 K during three phases (flow onset, fountaining and cooling). Secondly, the flank eruption of Nyamuragira is examined. There, for the period starting 6th November 2011 and lasting through to April 2012, the Global Volcanism Program report details volcanic activity which started as fissures and fountaining in early November. New cone formation took place in early December 2011, fissures and extended flows were registered throughout January. Finally the GVP report lava flows, degassing and lake formation in February which lasted until April 2012, with continuous degassing until June. The third volcanic target is the case of the 1200m-wide summit crater of Niyragongo. There, a lava lake of average diameter of 244m according to Burgi and colleagues [2014], shows frequent fluctuations of the level of its lava surface. It is also characterized by rapid fissure eruptions, lava fountaining, overflows, and movement of cooled lava plates within the lake [Burgi et al., 2014]. This area is studied for the years 2011-2012. Concerning the period between January and June 2011, the reports of GVP detail that the level of lava in the lake was stable and there was no lava overflow. In the period 30th May-9th June 2011, Burgi et al. [2014] conducted a field campaign in the area of the lake. Late on June 3rd, an inward drainage took place, and the surface of the lava lake was falling between 4-6th June. Spampinato et al. [2013] visited Nyiragongo in March 2012 and used a ground-based thermal camera to make approximately one-hour-long recordings of the activity in the lake for three consecutive days. They report mean apparent temperatures of 800-820K on 9th March; 740-808K on 10th March; and 730-824K on 11th March. They estimate that the mean radiative power varied between 1.00-1.20GW, 0.60-1.00GW and 0.60-1.10GW on the respective days. More detailed ground information is not available; in Congo, ground-based monitoring of both volcanic targets of this study is not systematic. On the contrary, Mount Etna is being regularly monitored from the ground [GVP, 2013].

#### 4.2.2 Image processing and hotspot detection

The method described in Pavlidou et al.[2016] applies a contextual approach to suppress patterns common between each pixel and its surroundings. Pixel values are divided by the average value of a square open frame of neighbouring pixels (normalization, see Figure 4.2). In this way, we highlight localized fluctuations that are present in the central pixel but not in the surrounding frame. Thereby, predominant patterns extending over the image are suppressed, including seasonal effects and unusual weather conditions. Normalization is run spatially, for all the pixels in the image and consequently for all images in the data set.



*Figure 4.2. Methodology. (a) Every pixel Y is normalized by the average of the neighbourhood pixels X1–X32. (b) The original time series from the central pixel (black line) and the average value of the frame pixels (red line). (c) In the normalized series resulting by the division of central pixel/frame time series, values exceeding the mean +  $2\sigma$  threshold (dashed blue line) are considered anomalous. A moving temporal window (hashed grey box) counts how many anomalous values coincide within the same time period. The procedure is carried out for every pixel time series.*

After the normalization, each pixel is analysed in its time series profile. In the normalized series, local fluctuations are flagged as anomalous in case they exceed a mean+ $2\sigma$  threshold [Pavlidou et al., 2016]. The threshold is calculated for every pixel's normalized time series. Thus, it is adjusted to each specific location and thereby facilitates detection over constantly elevated backgrounds, which is important for active volcanic areas [Koeppen et al., 2010]. A moving window runs through each pixel-based time series to count the number of anomalies that fall within the window (Figure 5.3, panel c). This procedure isolates values that exceed the set threshold and cluster in time, and it allows us to examine the temporal coincidence between a localized

anomaly and the potential underlying causative process. The analysis is carried out in every pixel of the study area, in all images of all timeslots. This allows for monitoring of unknown eruptions.

Following the procedures described in Pavlidou et al. [2016] we decide on the settings applied for processing (see Appendix 1). We set the length of normalization frame side to five pixels ( $\sim 15$  km based on the nominal spatial resolution of the sensor). This choice minimizes the variability of the normalized series in both study areas. It also ensures that the frame is wider than the maximum possible spatial extent of a potential anomaly, as described in the ground-based reports. Concerning the choice of temporal window, we show results with windows both smaller and larger than the events of interest. In particular, in Mount Etna, where the paroxysmal episode of 12 August lasted 33 h [Ganci et al., 2013], we present analysis with a 1 day temporal window. In Congo, where volcanic activity was more prolonged and was dominated by flank eruptions of variable duration (from few hours to days), we use a 5 day temporal window.

#### 4.2.3 Validation

We validate our findings with ground-based reports of the Global Volcanism Program [2011a-d; 2013; 2014a-b]; with the studies of Burgi et al. [2014] and Spampinato et al. [2013] for Congo; and with the study of Ganci et al. [2013] for Mount Etna. Radiant heat flux data were kindly provided by the authors. In the comparisons, priority is given to ground-based reports whenever they are available; satellite-based evidence is used as complementary. Agreement is considered between the available reports and our results, where (a) peaks are detected when activity is reported, (b) no peaks are detected when quiescence is reported, and (c) when detected peaks temporally coincide with reported increases in apparent temperatures and/or radiative heat flux, assuming that such increases would lead to higher numbers of detected anomalies. Findings are further compared to the MODVOLC alerts issued for the periods of interest. In this case, agreement is considered if detected peaks coincide temporally with issued alerts.

#### 4.2.4 Application on MWIR and LWIR at-sensor Brightness Temperatures

We apply same methodology is applied in the study area of Mount Etna, Sicily, using as input brightness temperatures (BT) derived (a) from the MWIR band (channel 4, centred at  $3.9\mu\text{m}$ ) and (b) from the LWIR band (channel 9, centred at  $10.8\mu\text{m}$ ) of the SEVIRI sensor. The temporal and spatial resolution of the data is the same as in the case of LST, as is the study area definition and the duration of the dataset. MWIR and LWIR BT data are neither atmospherically corrected nor cloud-masked. This analysis investigates atmospheric effects on

the result of detection and comparability and complementarity of the different inputs.

### **4.3 Results**

Results of the analysis of datasets from both study areas are presented in Figures 4.3-4.6. The peaks correspond to numbers of anomalies clustered within the moving temporal window. Comparisons are provided against MODVOLC alerts, published measurements of radiant heat flux and ground-based information on reported activity or inactivity.

In the case of Mount Etna, the eruption of the 12th of August is detected as the largest peak in the LST dataset; this is visible only in the two pixel time series affected by the flow and is approximately four times as large as the next highest peak in amplitude (Figure 4.3, Panel a). The peak coincides with MODVOLC alerts and with paroxysm #10, as reported by the GVP[Wang et al., 2016]. The paroxysm of August 12th was the largest of 2011 [Ganci et al., 2013]. Ganci et al. [2013] report a short period of increasing heat flux in the morning of the 12th of August, followed by a rapid increase during the main fountaining activity and a period of waning heat flux during cooling. These can be seen in both curves reported by the authors: they calculate the first using ground measurements and for the second they use METEOSAT MWIR observations (Figure 4.3, panel b). The detected peak is in agreement with both curves. The event is also detected when using MWIR (Figure 4.4, panel a2) and LWIR (Figure 4.4, panel b2) BT input. MWIR data saturate during the paroxysms (Figure 4.4, panel a1). The result of MWIR-based detection shows that paroxysm #9 is the largest one (Figure 4.4, panel a2), even though the ground reports describe that paroxysm #10 was instead the largest of the year. This issue is not present in LWIR-based results (Figure 4.4, panel b2).

Smaller paroxysms in the same month took place between 5-6th of August (#9), on the 20th of August (#11) and around the 29th of August (#12), according to the ground-based reports. The first two events included extensive ash emissions (dense plumes of ash and tephra). These resulted in missing values in LST data and impeded detection (Figure 4.3, panel a); similarly, they masked event #11 in LWIR data (Figure 4.4, panels b1, b2). MWIR data were less affected by the plumes and the event is detected as a peak (Figure 4.4, panel a2). Paroxysm #12 was preceded by an explosion and a series of ash emissions, between August 27-28th; these were not classified as paroxysms by GVP. The onset of the event of 29th of August (paroxysm #12) was detected as a small peak in the LST-based analysis, but it was also followed by a gap in the data when reported ash emissions took place (Figure 4.3, panel a). LWIR-based analysis shows a peak on August 27th and smaller peaks on the following

days. In MWIR-based analysis only the most intense event is visible as a peak on August 29th.

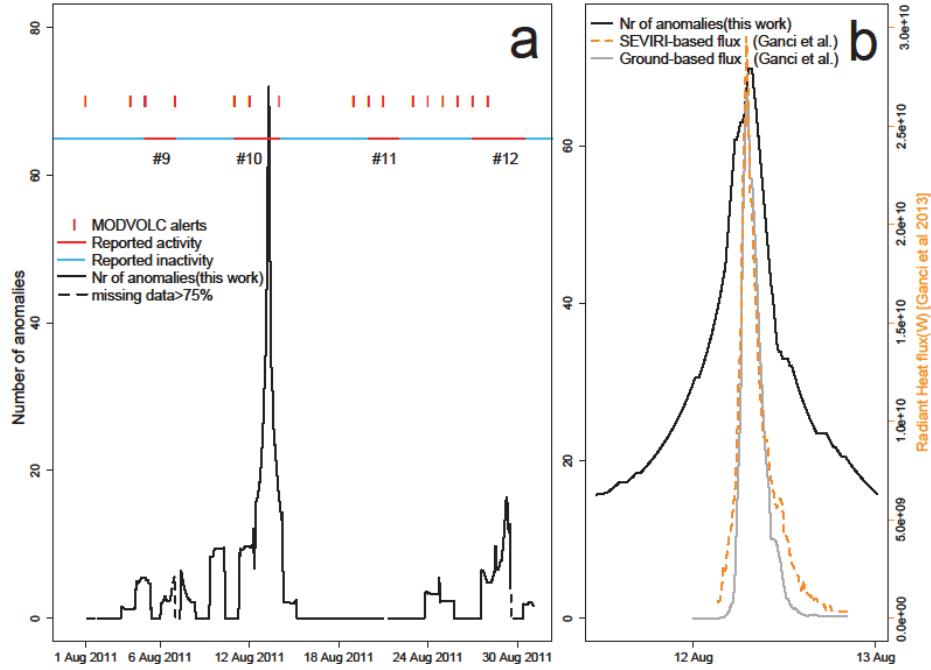


Figure 4.3. Hotspot detection at Mount Etna. Panel (a) shows detection results for the whole month of August 2011 from a pixel at the New SE crater of Mount Etna. Vertical red bars denote MODVOLC alerts. Periods of activity (paroxysms #9 to #12,) reported by the Global Volcanism Program and Ganci et al (2013), are shown in red horizontal lines. Blue lines correspond to periods of inactivity throughout August 2011. Results calculated with more than 75% data missing are shown with a dashed black line. Panel (b) shows the peak of August 12 (paroxysm #10) in comparison to the radiant heat fluxes calculated by Ganci et al (ground-based, in gray; and SEVIRI-based, in dashed brown).

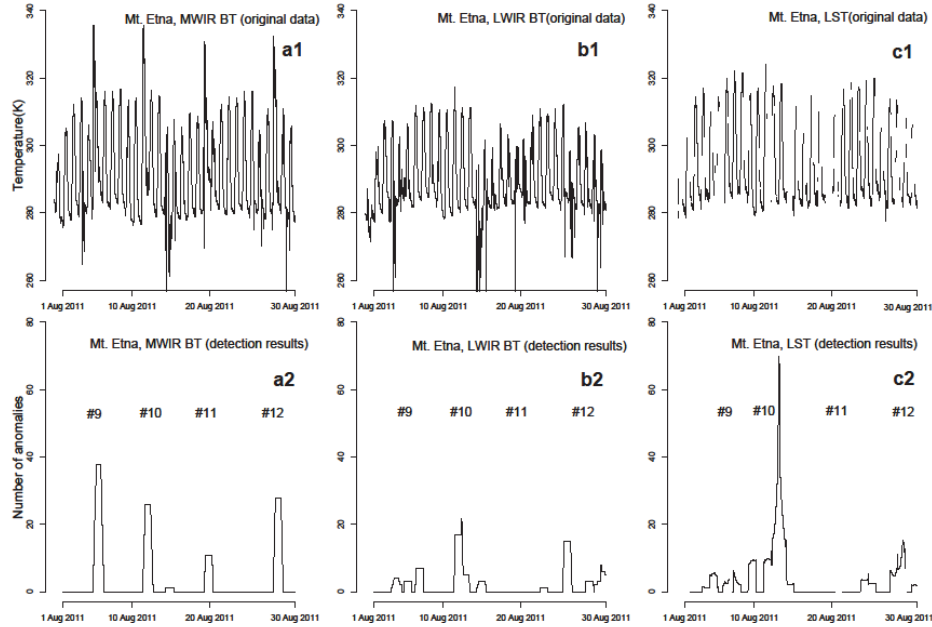


Figure 4.4. Analysis of data from different wavebands over the Mt. Etna study area. Original Brightness Temperature (BT) time series from METEOSAT MWIR band (panel a1), are compared with LWIR BT (panel b1) and LWIR-based LST time series (panel c1). Detection results are shown correspondingly in panels (a2)-(c2). Periods of reported activity are shown as paroxysm numbers.

In Congo there were two different types of activity: a series of effusive lava flows on the flanks of Nyamuragira and a permanent lava lake in Nyiragongo (Figures 4.5 and 4.5).

The Nyamuragira lava flows affected mostly six pixels (shaded white in the left panel of Figure 4.5b). We present results of the detection for all of these pixels in Figure 4.5a, using different colors. In all pixels we detect a series of peaks which are up to 6 times larger than the rest of the dataset, and they distinctively cluster in the period of volcanic activity described in the GVP report (Figure 4.5a). A high number of MODVOLC alerts were also issued during that period. The compilation of results from neighboring pixels provides insight in the spatiotemporal evolution of the event. For example, as seen in more detail in Figure 4.5b, a peak appears first in pixel 5 and then moves N-NE to neighboring pixel 4; this in in agreement to the flow described by the GVP report. Such peaks were not visible in pixels unaffected by volcanic activity (Figure 4.5c).



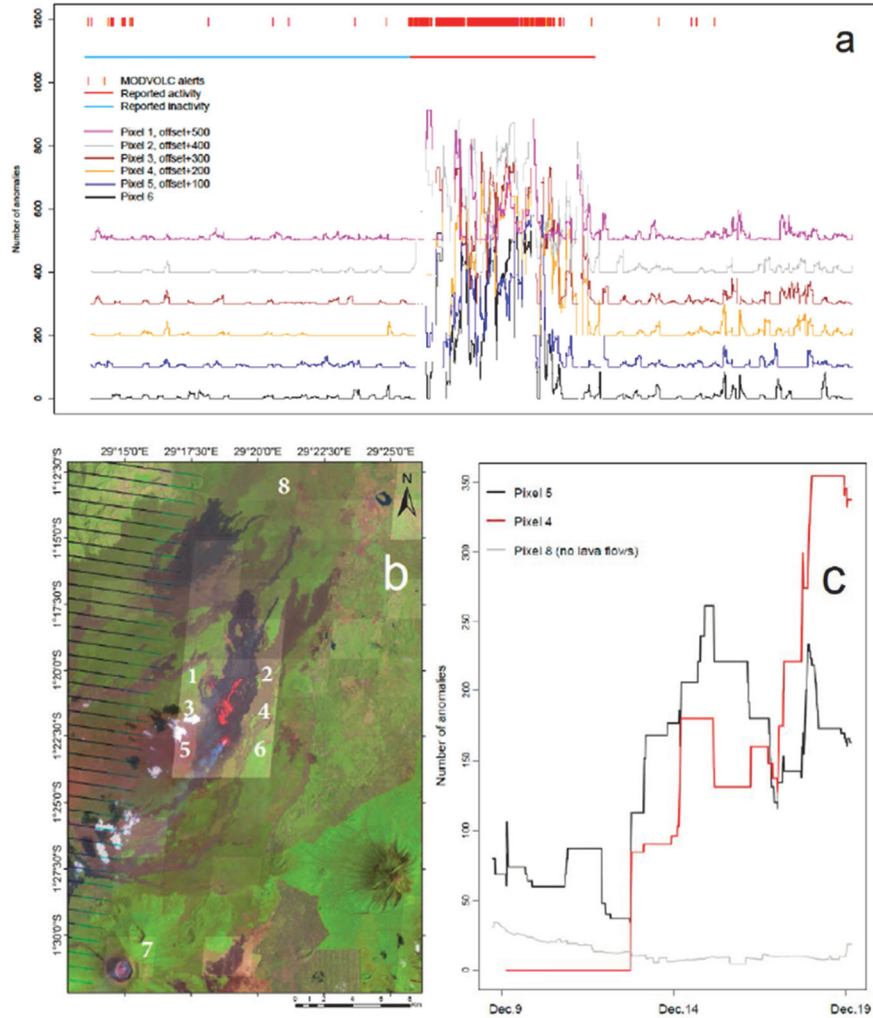


Figure 4.5. Hotspot detection in Virunga National park, D.R.Congo. Panel (a) shows detection results for the complete period 2011-2012. The number of anomalies detected in the pixels covering lava flows is shown in different color for each pixel. The location of the pixels is shown in the Landsat image of panel b. These were pixels with maximum counts of detected anomalies. The period of volcanic activity reported by the Global Volcanism Program [2014a] is denoted with a red horizontal line. Red vertical bars correspond to issued MODVOLC alerts. The period of reported inactivity is shown with a blue horizontal line. Panel b shows a detail from a Landsat ETM+ color composite image of the flows superimposed by an image showing the location of the largest peaks we detected in each pixel time series. Pixels 1-6 show the highest peaks, are shaded white and coincide with the area of fresh flows. Pixel 7 covers the lava lake of Nyiragongo. Pixel 8 is used as a reference pixel, not affected by lava flows, plotted here to show lack of variation in non-affected pixels. Panel c shows in more detail, peaks detected in pixels 4 and 5 of panel b. Peaks that are detected in pixel 5 (for example, see Dec. 15<sup>th</sup>), move temporally to pixel 4 (in the same example, Dec. 18<sup>th</sup>), indicating N-NE movement of the flows as described in the Global Volcanism Program report [2014a]. Landsat imagery courtesy of USGS.

In Nyiragongo MODVOLC alerts are present throughout the dataset, as would be expected for a permanent lava lake. The work carried out for this chapter indicates the periods when activity was more intense (Figure 4.6). The findings show quiescence between January and July 2011. Activity during this period is sporadic and of low intensity. This is followed by an increase in activity that culminated in September 2011; other periods of intense activity include the first days of March 2012 and November 2012. The GVP reports that the levels of lava in the lake were stable in the period January-June 2011 in agreement with our observations. Burgi and colleagues report an incident of convective inward drainage of the lake and a total drop of the lava surface by 33m during 3rd-6th June, 2011. The event did not involve overflows nor outward activity and no peak was detected with our analysis either. Spampinato et al. [2013] carried out SEVIRI-based calculations of radiative power, using the HOTSAT system between January and June 2012. The highest mean monthly radiative power values of this period were recorded in March. Increased hourly values were more specifically present in the beginning of March; these coincide with the highest peak we detect in March (Figure 4.6, shown with an asterisk). The authors subsequently report an oscillating falling trend in their ground-based measurements of apparent temperatures between 9th March-11th March, and that is consistent with the decreasing, low numbers of anomalies detected in these days by the analysis described in this chapter. The largest peak was detected over the lava lake was early on the 13<sup>th</sup> of September 2011, and such a peak was not present in neighboring pixels. Cloud cover was increased in this period but throughout the study area. This peak, and the rest of the detection results in Nyiragongo, could not be validated due to lack of ground-based information.

#### **4.4 Discussion**

MWIR data are commonly used in many volcanic applications and hotspot detection approaches. Still, the study of volcanic activity could benefit from the utilization of LWIR-based, archived observations. Relevant datasets are becoming increasingly available. For example, Duguay-Tetzlaff et al. [2015] report on the quality of a 30+ year-long, LWIR-based LST climate record recently compiled by EUMETSAT. GlobTemperature, an initiative of the European Space Agency, aims among others at the production and dissemination of long-term satellite LST products (<http://www.globtemperature.info/>). This study was intended to test such datasets in volcanic applications, using a recently published approach sensitive to subtle localized signal fluctuations.

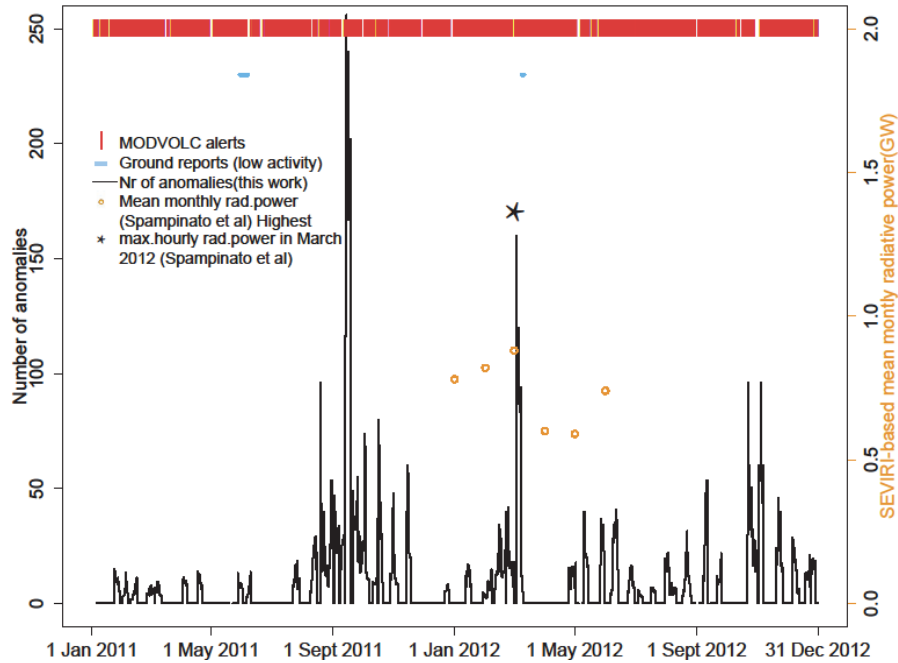


Figure 4.6 Detection results from the lava lake of Nyiragongo. MODVOLC alerts can be seen as red vertical lines and are present throughout the series, as would be expected in the case of a permanent lava lake. Our detection denotes the periods with highest activity levels. Available reports of inactivity are denoted by blue horizontal lines. Mean monthly values of radiative power, as reported by Spampinato et al (2013), are shown in the figure with brown circles. The black asterisk shows the peak that corresponds to the highest hourly radiative power values of March 2012 as calculated by the same authors.

The challenges involved in existing hotspot detection include algorithm transferability to areas with different prevailing conditions, particularly when fixed thresholds are used; and detection of subtle anomalies, commonly obscured by atmospheric conditions [Watson et al., 2004, Ganci et al., 2011, Blackett, 2014, Wright, 2015]. The temporal component of the input contains useful information for the isolation of extreme events, but detection cannot be based only on statistical characterization of past observations [Koeppen et al., 2010] especially where anomalies are obscured by predominant, temporally varying patterns. The approach described in Pavlidou et al [2016] combines contextual and temporal methods to suppress large scale patterns at the time they emerge, highlighting subtle fluctuations. The detection threshold is dynamically determined for every pixel time series. As a result, detection is inherently considering local conditions. The settings of the method are adjustable to the desired timescale and the characteristics of the volcanic target of interest (see also supplement for details). For example, if the aim is

to resolve short-term activity, shorter temporal windows can be applied as demonstrated over Mount Etna. In the case of permanent intermittent volcanic unrest, where volcanic activity is always present but at times low and at other periods high, the use of a lower threshold would describe both lower- and higher intensity events and result in a continuous series of hotspots. This is avoided in the lava lake in Nyiragongo by highlighting more intense events. It is thus possible to trace increased activity through time. Additionally, the methodology is applied uniformly with the same settings across the dataset. Unexpected fluctuations can therefore be detected even in parts of the image not known a priori, as is the case of flank eruptions in a wider volcanic area.

Our input is LST retrieved from geostationary satellite LWIR data. The advantage of geostationary sensors is that they provide frequent temporal coverage and are thus able to capture short-lived, transient events [Ganci et al., 2011, Ramsey and Harris, 2013]. Geostationary-based observations are already available for decades, supporting the possibility for long-term studies of volcanic dynamics. The datasets used are atmospherically corrected and cloud-masked. Still, the quality of addressing the prevailing atmospheric conditions remains as a limitation in volcanic monitoring; especially in the case of misclassified clouds and volcanic ash.

Cloud coverage could obscure a short-lived event. The effect of this influence is decreased in the case of geostationary-derived input, because of the frequency of sampling; however, when there is persistent cloud cover during an eruption, a transient event will not be captured. Furthermore, if clouds remain in the data there is a possibility of false detection. When frame values are very low due to the presence of clouds, the normalized value of the central pixel will be very high. This could be the case, for example, of the minor detections appearing on August 15-16th in MWIR and LWIR-based results (Figure 4, panels a2 and b2). Atmospherically-corrected LST rather than at-sensor LWIR BT are preferred as input to minimize such effects.

Cloud remnants may also affect LST retrievals. The cloud-mask used for the production of the dataset has a reported percentage of max.4% of missed cloud identification, mostly due to cloud edges and thin low clouds at night (see for details Trigo et al [2009]). LST retrievals are similarly affected by the amount of total column water vapour in the atmosphere, and possibly by the turbulent conditions prevailing over volcanic targets in unrest. These effects are partly reflected in the quality information of the LST product and partly remain in the data. For this analysis such remnants would lead to peaks non-related to volcanic activity. To minimize such effects, only LST data of at least nominal confidence level are included in the study. Any peaks resulting from non-addressed atmospheric influences, if present in the results, are not comparable to the size of the ones caused by volcanic flows. It can be argued

that the atmospheric corrections in the analysis, especially with the strict quality constraints imposed, obscured targets of interest. This is true, considering that in the case of Mount Etna, the same analysis with non-corrected LWIR BT input could detect all reported events, whereas minor events under thick ash emissions were missed using LST input. By choosing LST input, however, false detections caused by cloud remnants and atmospheric effects can be avoided, even at the expense of missing events of smaller intensity. This may be of importance for detecting activity in low-temperature targets.

Volcanic ash can also interfere with the registration of an event in the data. Watson et al. [2004] mention that the presence of SO<sub>2</sub>, principal component of volcanic plumes, attenuates the recorded IR signal. Even though an SO<sub>2</sub> plume may not mask out radiating basaltic pixels completely, as seen in the results from MWIR input, it can still attenuate the signal registered in the LWIR and is more pronounced when the plumes are thicker and contain water vapour. This is visible when comparing MWIR and LWIR data (Figure 4.4). Such attenuation in the analysis would result in a peak appearing smaller than it would be in clear sky conditions. LST retrievals do not address the potential presence of volcanic plumes. However, in spite of attenuation, detected peaks in the results could be clearly distinguished from quiescence.

Thicker ash plumes, or plumes which act as condensation nuclei and facilitate formation of ice clouds in their upward movement through the atmosphere, could lead to missing values in LST data. Such could be the case of the small paroxysm #9 in Mount Etna: the GVP reported gas emissions during the event, and in the dataset missing values did not allow detection. The percentage of missing values in the temporal window affects the size of the detected peaks. When the temporal window/normalization frame only include a few observations it is not certain that these observations are representative, and this introduces sampling uncertainty in the results. It would be thus possible to set a threshold to discard results calculated with relatively incomplete normalization frames or temporal windows. However, considering that degassing is commonly present and can be an indicator of volcanic activity (as seen also in MWIR and LWIR BT input), no data availability threshold is applied in the processing. This allows to make the most of the available data and to utilize indirect evidence of degassing. The quantification of the effect of turbulent atmospheric conditions and volcanic emissions in the future could eliminate such sources of uncertainty in LST-based results.

MWIR-based detection is not as affected by the plumes as is LWIR-based analysis. For example, paroxysm #11 is detected only using MWIR input. On the other hand, MWIR BT data reaches saturation easier and the result of the analysis does not represent the relative magnitude of the events correctly:

paroxysm #9 appears larger than #10, which is not the case according to ground-based observations. LWIR BT- based analysis is more sensitive to events of lower intensity, like the ones preceding paroxysm #12. These comparisons showed that it is possible to implement the methodology using different wavebands. This supports application of the approach further than just for utilization of LWIR LST archives. Results of analysis of different wavebands can complement each other or be used in combination with mainstream methodologies.

## **4.5 Conclusions**

We test a recently developed hot spot detection methodology to study volcanic activity. This methodology suppresses predominant patterns and detects temperature anomalies. Our analysis is done with hypertemporal LWIR-based LST data, which facilitates volcanic studies when MWIR information is not available, for example, in older sensors. The hypertemporal component of the approach allows for studying the dynamic development of volcanic activity at different timescales, from transient 1-day long events (for example in the case of Mount Etna) to longer eruptive sequences (for example in the case of Nyamuragira). The method is applied on input from different wavebands. It is adapted to local conditions by using dynamic thresholds, and as a result it can be used in different volcanic regions globally. Fluctuations in volcanic activity are detected systematically even where they occur in parts of the image not previously active and/or in areas with constant volcanic activity (for example in the case of Congo). The main findings are in agreement with the available satellite- and ground-based information. It is concluded that the approach can be applied to successfully detect subtle signal fluctuations; it can support the use of LST data for studies of volcanic activity; it can complement mainstream MWIR/LWIR methodologies, and it can prove useful when it comes to thermal targets which are not (or cannot be) sufficiently monitored from the ground.

## **Chapter 5. Time series analysis of Land Surface Temperatures in 20 earthquake cases**

### **5.0 Abstract**

Earthquakes are reported to be preceded by anomalous increases in satellite-recorded thermal emissions, but published results are often contradicting and/or limited to short periods and areas around the earthquake. We apply a methodology that allows to detect subtle, localized spatio-temporal fluctuations in hyper-temporal, geostationary-based Land Surface Temperature data. We study ten areas worldwide, covering twenty large ( $M_w > 5.5$ ) and shallow ( $< 35\text{km}$ ) land-based earthquakes. We compare years and locations with and without earthquake, and we statistically evaluate our findings with respect to distance from epicentra and temporal coincidence with earthquakes. We detect anomalies throughout the duration of all datasets, at various distances from the earthquake, and in years with and without earthquake alike. We find no distinct repeated patterns in the case of earthquakes that happen in the same region in different years. We conclude that earthquakes do not have a significant effect on detected LST anomalies.

### **5.1 Introduction**

Observations that are referred to as 'thermal anomalies' have been reported to precede earthquakes worldwide since the 1980's [Tramutoli et al., 2015]. Numerous reports describe sudden increases in Brightness Temperatures (BT) recorded by satellite sensors [Lisi et al., 2010, Ouzounov et al., 2006, Tramutoli et al., 2001]; surface temperatures derived using either satellite observations [Baral et al., 2016, Lisi et al., 2015, Akhoondzadeh, 2013b] or numerical simulations [Alvan et al., 2014, Qin et al., 2012a]; air temperatures recorded with ground-based meteorological stations [Pulinets and Dunajacka, 2007, Panda et al., 2007]; satellite-based Outgoing Longwave Radiation (OLR) [Lu et al., 2016, Ouzounov et al., 2007], Surface Latent Heat Flux (SLHF) [Cervone et al., 2006, Dey and Singh, 2003]; and increases in soil temperatures measured on-site [Rezapour et al., 2010, Liu et al., 1999]. Definitions of what is actually an earthquake related thermal anomaly vary among researchers and differ among applied methodologies. Observed anomalies also seem to differ for the same earthquake: they may appear a few hours [Akhoondzadeh, 2013b] to a few years [Yao, 2010] prior to an earthquake, and they might reappear shortly after an earthquake [Tronin, 2000a]. The spatial resolution of the data used can be as limited as point observations from a meteorological station [e.g. Jie and Guangmeng, 2013] or as coarse as gridcells of  $2^\circ \times 2^\circ$  [e.g. Singh et al., 2010]. Datasets are often

confined to limited areas and short periods of time around the earthquake [e.g. approximately one month and only in the pixel covering the earthquake epicenter in Akoondzadeh, 2013, Ouzounov et al., 2004]. Such settings do not allow for detailed examination of the spatiotemporal coincidence of earthquakes with detected anomalies and do not permit to detect false positives. Furthermore, statistical evaluation of the results is often missing (relevant discussion can be found in Eneva [2008]).

The objective of this study is to examine the presence of thermal anomalies shortly before large, shallow, land-based earthquakes. We consider that an earthquake is large when it has magnitude larger than  $M_w$  5.5 and is shallow when it has focal depth  $<35$  km. Thermal anomalies may appear for a variety of reasons other than earthquakes, including spatiotemporal variations of surface spectral emissivity [Tramutoli et al., 2005] and local atmospheric conditions, like atmospheric inversions [Qu et al., 2006]. We test the hypothesis that more anomalies would be detected at closer distances to the earthquake, shortly prior or during the earthquake, and only in years with earthquake occurrence. This hypothesis is supported by published research concluding that anomalies increase with increasing earthquake magnitude; anomalies are found predominantly near the epicenter, one day before and on the day of the earthquake; and anomalies are more easily observed during shallow earthquakes than the deep ones [Xiong and Shen, 2017]. We study twenty large, shallow, land-based earthquake cases in ten study areas around the world, with different local environmental and climatic conditions. We apply a recently published methodology which suppresses large-scale patterns in the satellite signal time series [Pavlidou et al., 2016]. This methodology isolates only spatially localized anomalies and determines the time of their occurrence. We statistically evaluate our findings, taking into account the spatial and temporal occurrence of detected anomalies and earthquakes. Other influences that may be related to the detected anomalies are also examined to support the results.

## **5.2 Datasets**

### **5.2.1 Parameter of interest**

Land Surface Temperatures (LST) derived from geostationary satellite sensors are often considered as preferred input parameter for earthquake-related studies [Jiao et al., 2017; Lisi et al., 2010]. LST products are derived from Top-of-Atmosphere Brightness Temperatures after cloud-masking and application of atmospheric corrections. As a result, they are less affected by artefacts of atmospheric origin [Lisi et al., 2010]. A review of LST retrieval algorithms can be found in Li et al. [2013]. Operational LST products of improved accuracy, compared to previous decades [Li et al., 2013, Prata et al., 1995], are consistently calculated at a near-global scale and accompanied by



uncertainty estimates [Wan, 2014, Freitas et al., 2013, Trigo et al., 2011, Freitas et al., 2010]. Geostationary sensors offer high temporal resolution, which can resolve the duration and timing of thermal anomalies. Furthermore, the spatial resolution of these sensors is higher than the spatial resolution of surface temperatures and other parameters derived from atmospheric reanalysis. Therefore, geostationary satellite-derived LST is also chosen as input for our study.

### 5.2.2 Raw data

We use the Copernicus Land Surface Temperature (LST) dataset (<http://land.copernicus.vgt.vito.be>), which is produced with the application of atmospheric corrections and cloud masking on Top-of-Atmosphere observations of three geostationary satellites (METEOSAT, GOES, MTSAT). The product is available at a nominal spatial resolution of 5 km. Its temporal resolution is 24 images per day, with the exception of the first six months of 2010 (8 images daily) [Freitas et al., 2013]. Its spatial coverage is near-global (no coverage over Central Asia and northern latitudes) and, at the time of acquisition, data were available for the period 2010-2016. The product is delivered ready to use and free of charge through the Copernicus network (Copernicus Global Land Services, <https://land.copernicus.eu/global/index.html>) along with uncertainty information in the form of error margins. Quality information is also provided based on total column moisture content and satellite positioning.

### 5.2.3 Earthquake catalogue

The study involves all the large, shallow-focus, land-based earthquake cases occurring in the temporal and spatial domain of the available Copernicus LST dataset, according to the USGS online earthquake catalogue (<https://earthquake.usgs.gov/earthquakes/search/>). All events with magnitude >5.5, registered in the same study area within a 24-hr window, are grouped together. Each of these earthquakes is covered by one year-long LST dataset, and all study areas are also tested in a year without earthquakes. When more years without earthquakes are available, we choose a year common for many study areas for practical reasons. All study cases are presented in Table 5.1, along with their magnitude, focal depth and focal mechanism, whenever available.

Table 5.1. Earthquake events studied in this work and the corresponding study areas. The spatial extent is given in km<sup>2</sup> based on the nominal spatial resolution of the LST dataset. Earthquakes occurring within 24hours are grouped together. The climate classification is based on Peel et al [2007]. The earthquake information is taken from the online earthquake catalogue of USGS (<https://earthquake.usgs.gov/earthquakes/search/>).

Earthquake	Date	Mw	Focal depth (km)	Focal mechanism	EQ year	Control year	Study area code	Area extent (km <sup>2</sup> )	Climate
1. Ahar, Iran	11.8.2012	6.2	12	strike-slip	2012	2011	IR	415x530	Continental
2. Ahar, Iran	11.8.2012	6.4	11	strike-slip	2012	2011		415x530	Continental
3. Baja California, Mexico	4.4.2010	7.2	10	strike-slip	2010	2016	MX	1705x1665	Arid
	4.4.2010	5.7	10	thrust	2010	2016			
4. Ocotillo, Mexico	15.6.2010	5.7	8.7	strike-slip	2010	2016		1705x1665	Arid
5. Yunnan, China	10.3.2011	5.5	10	strike-slip	2011	2012	CH	840x550	Temperate
6. Shwebo, Myanmar	11.11.2012	6.8	13.7	strike-slip	2011	2012	MS	665x550	Tropical
	11.11.2012	5.8	6	strike-slip	2011	2012			
	11.11.2012	5.5	28.4	strike-slip	2011	2012			
7. Burma, Myanmar	24.3.2011	6.9	8	strike-slip	2011	2012	MB	1505x455	Tropical
	23.3.2011	5.7	10	strike-slip	2011	2012			
8. Christchurch, NZ	14.2.2016	5.8	7.6	thrust	2016	2012	NZ	1350x1400	Temperate
9. Kaikoura, NZ	13.11.2016	7.8	15	thrust	2016	2012		1350x1400	Temperate
	13.11.2016	6.5	10	thrust	2016	2012			
	13.11.2016	6.2	10	strike-slip	2016	2012			
	13.11.2016	6.1	14	strike-slip	2016	2012			
	13.11.2016	5.5	10		2016	2012			
	14.11.2016	6.5	9	strike-slip	2016	2012			
10. Amberley, NZ	22.11.2016	5.5	10	strike-slip	2016	2012		1350x1400	Temperate
11. Canterbury, NZ	3.9.2010	7	12	thrust/ strike-slip	2010	2012		1350x1400	Temperate
12. Van, Turkey	23.10.2011	7.1	18	thrust	2011	2012	TR	465x530	Continental
	23.10.2011	5.9	5	thrust	2011	2012			
	23.10.2011	5.7	5		2011	2012			
	23.10.2011	5.6	9		2011	2012			
	23.10.2011	5.6	5		2011	2012			
13. Van, Turkey	24.5.2011	5.6	14	thrust	2011	2012	TR	465x530	Continental
14. Emilia Romagna, Italy	20.5.2012	6	6.3	thrust	2012	2011	IT	655x655	Temperate
15. Emilia Romagna, Italy	29.5.2012	5.8	10.2	thrust	2012	2011		655x655	Temperate
	29.5.2012	5.5	6.8	thrust	2012	2011			
16. Norcia/ Amatrice, Italy	24.8.2016	6.2	4.4	normal	2016	2011		655x655	Temperate
	24.8.2016	5.6	3.2	normal	2016	2011			
17. Norcia/ Visso, Italy	26.10.2016	6.1	10	normal	2016	2011		655x655	Temperate
	26.10.2016	5.5	6	normal	2016	2011			
18. Norcia, Italy	30.10.2016	6.6	8	normal	2016	2011		655x655	Temperate
19. Oklahoma, USA	3.9.2016	5.8	5.6	strike-slip	2016	2012	UO	1100x1100	Temperate
20. Oklahoma, USA	6.11.2011	5.7	5.2	strike-slip	2011	2012		1100x1100	Temperate
21. Louisa, USA	23.8.2011	5.8	0.02	thrust	2011	2012	UL	1155x955	Continental

### 5.3 Methodology

The approach followed in this study consists of the following steps: pre-processing, normalization, anomaly detection, and calculation of numbers of anomalies at different time periods and distances from the earthquake. Statistical analysis is performed to examine if more anomalies are detected prior to earthquakes, and closer to earthquakes in space.

### 5.3.1 Preprocessing

The study involves only pixels with at least nominal quality (i.e. with estimated errors lower than 2K). Clouds, missing observations, water bodies and pixels with error margin  $>2K$  appear as missing data. The data are spatially subset ensuring that the study areas fully cover the earthquake-struck areas and provide reference of the signal of non-earthquake struck regions at the same time period. In total, ten study areas are defined. Following the Koeppen-Geiger climate classification [Peel et al., 2007], in four of the study areas the climate is temperate; in three, continental; in one, arid, and in two tropical. After spatial subsetting, images are stacked sequentially to construct time series. Missing images are included as gaps in the dataset. One-year datasets are constructed including all available timeslots throughout the day, allowing to consider annual, seasonal and daily fluctuations. Stacks of the same area are prepared in years with and without occurrence of large earthquakes. Years are chosen to cover all main events in the same area separately, so for example there are three datasets covering Italy: one for year 2012 (Emilia-Romagna earthquakes), one for year 2016 (Amatrice/Norcia events) and one for year 2011 (reference year, no large earthquake).

### 5.3.2 Normalization

The normalization step suppresses patterns common between the central pixel and an open frame of neighbouring pixels. Such patterns could be the diurnal cycle, seasonal effects, humid conditions extending in large areas or a heat wave. Every pixel value is divided by the average value of a frame of neighbours (Figure 5.1). The process is repeated for every pixel of every image of the dataset. After normalization, only localized fluctuations remain and are highlighted in the resulting normalized time series.

An identical frame size is used in all study areas to obtain comparable results. The assumption is that anomalous thermal emissions caused by stress accumulation would be more pronounced over a stressed area. We therefore use larger frames than that area to highlight potential anomalous emissions by applying the normalization. Following Saradjian and Akhoondzadeh [2011a], the area of stress accumulation is estimated based on the definition of earthquake preparation area given by Dobrovolsky et al. [1989]. Dobrovolsky et al. [1989] describe a spherical 'region of earthquake preparation' defined as a function of the magnitude of an impending earthquake. Rupture length is also taken into account for the decision on frame size, because it could give a first indication of the spatial extent of stress accumulation. Moment magnitude can be used to approximate rupture length based on empirical relationships [Wells and Coppersmith, 1994], even though this relationship may vary across different locations.

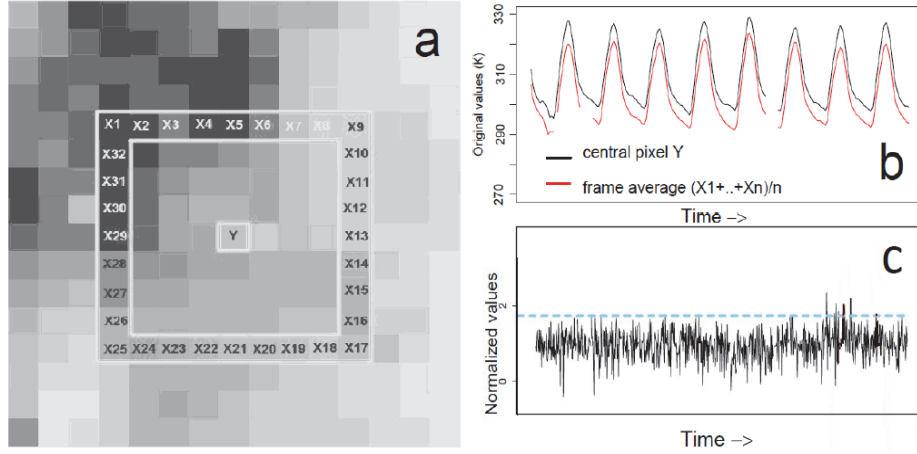


Figure 5.1. An example of anomaly detection. The value of every pixel of the image is divided by the average of an open frame of neighboring pixels (panels a, b). The procedure is repeated for all images of the dataset, resulting in a normalized time series for every pixel (panel c, black solid line). Then a mean+2sigma threshold (blue dashed line, panel c) is applied to each pixel's normalized time series to distinguish extreme values that are then regarded as anomalies.

We set the frame size based on the largest event in our earthquake list: the Kaikoura Mw 7.8 thrust rupture in New Zealand. Its preparation area according to Dobrovolsky et al. [1989] has a diameter of 68km, slightly smaller than 14 pixels in our image datasets. The calculated rupture length based on Wells and Coppersmith [1994] is 113km ( $\sim 23$  pixels), and is reported for this particular earthquake by Shi et al. [2017] as approximately 116km on-land, estimated across at least 12 major fault sections. The frame is therefore chosen to have a side length of 29 pixels (145km), which exceeds the extent of both the estimated earthquake preparation area and the rupture length in all earthquake cases of this study.

### 5.3.3 Anomaly detection

In the normalized time series, values which exceed or are equal to a threshold equal to the mean plus twice the standard deviation of the series ( $\mu+2\sigma$ ) are flagged as anomalies. This definition is consistently applied in all datasets.

The choice of  $\mu+2\sigma$  threshold is in accordance with previous research on detection of earthquake-related anomalies [Daneshvar et al., 2014, Ouzounov et al., 2006, Qin et al., 2009, Tronin et al., 2002]. The threshold is determined by each pixel's series separately, and is thus dynamically adjusted to the local conditions. Furthermore, it is applied uniformly in the whole image, allowing for detection of anomalies without *a priori* knowledge on where anomalies may appear. Sensitivity analysis of the original method showed that this threshold

could capture fluctuations as low as +2K/+3K compared to surroundings of lower/higher background variability [Pavlidou et al, 2016]. To control for the possibility that this threshold may also capture lesser environmental influences distorting the statistical evaluation of the results, we repeat the same analysis with a stricter  $\mu+3\sigma$  threshold.

#### 5.3.4 Anomaly density calculations

This step includes counting the number of anomalies per unit time, for four spatial zones describing different distances from an earthquake epicenter, and for different time periods (see Table 5.2 for spatial and temporal details).

A distinction between pre-, post- and co- seismic periods is suggested by previous research, especially considering the co-seismic period [Eneva et al., 2008]. Pre-seismic periods are considered as a time of stress accumulation. Potential anomalies related to this stress accumulation are expected to appear closer to the time of the rupture [Barkat, 2018, Xiong and Shen, 2017]. A co-seismic period is considered to be the time of rupture and stress release. Thermal anomalies appearing in this time would be related more to friction and the rupture itself rather than stress accumulation. Post-seismic periods follow the rupture and have less stress than the pre-seismic periods, since part of the accumulated stress has been released during the rupture. We define co-seismic periods as 24 hours before and 24 hours after an earthquake. All events of  $M_w > 5.5$  that happen more than 24 hours after another event are seen, in this definition, as separate. Pre- and post- seismic periods of equal duration (two months) are defined for comparability and based on literature on the appearance of earthquake-related anomalies [e.g. Qin et al., 2013; Wei et al., 2009; Wu et al., 2008]. In periods later and earlier than these, earthquake related anomalies should be unlikely.

Table 5.2 Definitions of spatial zones, temporal periods and adjusted co-seismic periods applied for anomaly density calculations.

Spatial zones	
Zone	Zone definition
<b>Zone 1</b>	within a radius of 7 pixels (35km) from the epicenter (up to half the frame side)
<b>Zone 2</b>	A concentric ring with a radius between 7 and 14 pixels (35-70km) from the epicenter (half the frame side until the edge of the frame side)
<b>Zone 3</b>	A concentric ring with a radius between 14 and 28 pixels (70-140km) from the epicenter (between one and two times the frame side)
<b>Zone 4</b>	Rest of the study area with a radial distance larger than 28 pixels (>140km) from the epicenter (larger than twice the frame side)
Temporal periods	
Period	Period definition
<b>Period 1 (pre-earthquake)</b>	beginning of the dataset until 2 months prior to the co-seismic period
<b>Period 2 (2month pre-earthquake)</b>	two months before the co-seismic period
<b>Period 3 (co-seismic)</b>	24 hours before and 24hrs after the earthquake
<b>Period 4 (2month post-earthquake)</b>	two months after the co-seismic period
<b>Period 5 (post-earthquake)</b>	two months after the co-seismic period until the end of the dataset
Extended co-seismic periods	
Earthquake sequence	Duration of adjusted co-seismic period
Canterbury, New Zealand [Quigley et al., 2016]	2/9/2010-31/12/2010
Kaikoura, New Zealand [Shi et al., 2017]	12/11/2016-31/12/2016
Baja California, Mexico [Hauksson et al., 2011]	3/4/2010-16/6/2010
Italian sequence of 2016 [Chiaraluce et al., 2017]	23/8/2016-31/10/2016
Van, Turkey [Doğan and Karakaş, 2013, Elliot et al., 2011]	22/10/2011-10/11/2011
Emilia-Romagna, Italy [Govoni et al., 2014]	19/5/2012-30/5/2012

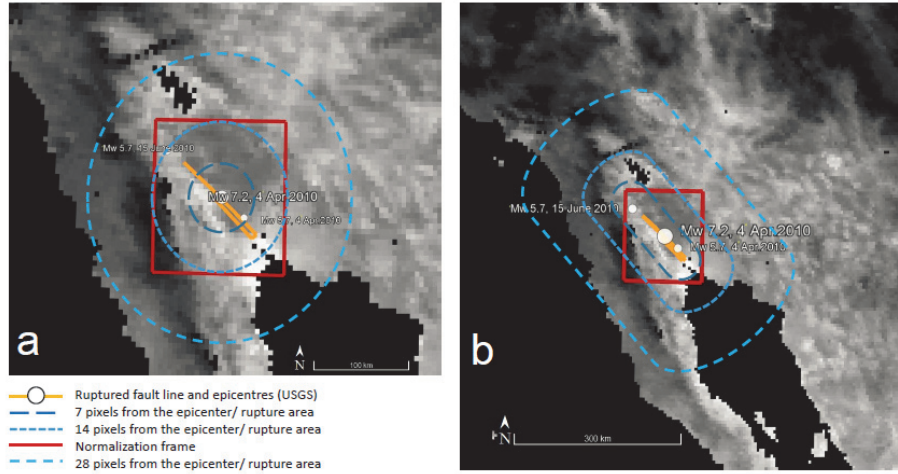


Figure 5.2. Definition of distance zones in the case of Baja California, Mexico. The images show the Copernicus LST image of April 4th, 2010, 09:00 hrs (day of the earthquake). Without adjustment of the co-seismic period, earthquakes which occur more than 24 hours apart are considered separate events. For example, panel a shows definition of distance zones based only on the earthquake of April 4th. The distance zones (dashed lines) are defined around the epicenter. When the whole earthquake sequence is considered (in the case of California, one day before the Apr.4th event until one day after the June 15th event), distance zones are defined with regard to the ruptured area, which includes the ruptured fault and the epicentres of all large earthquakes in the sequence (panel b). The size of the normalization frame is shown with squares in both panels.

This study includes earthquakes which may be considered as parts of longer earthquake sequences. The sequences are characterized by a higher frequency of low-medium magnitude events than in the rest of the earthquake catalogue. It could be argued that, in these cases, stress is being released not in a single rupture but through a sequence of ruptures. To account for such prolonged sequences, the analysis is done twice: once with a 48hr-long co-seismic period and once with a prolonged co-seismic period, adjusted to cover the duration and spatial extent of each sequence. When prolonged co-seismic periods are used, distance zones are defined with regard to the ruptured area, which includes the ruptured fault and epicentres of all large earthquakes in the sequence (Figure 5.2). Analysis is performed on both standard and adjusted co-seismic periods in the cases described in Table 5.2.

The total number of anomalies in each zone and period is divided by the total number of available observations in the corresponding zone/period. The result is anomaly density in different periods at different distances from each earthquake. The same processing (normalization-anomaly detection-calculation of anomaly density) is repeated in years with and without earthquake occurrence.

### 5.3.5 Statistical analysis

Statistical analysis follows to evaluate detection results based on the hypothesis that a higher anomaly density should be found in areas spatially closer to the earthquake, and in periods shortly before or during the earthquake and only in the years with an earthquake.

A three-way mixed Analysis Of Variance (ANOVA) analysis allows to test for significant combined effects among three factors of interest: time period, distance zone and earthquake/no earthquake year. The analysis is repeated for the following experimental configurations: use of different thresholds and with/without adjustment of the co-seismic period.

There are several possible outcomes of such an analysis. One is that there is a statistically significant combined effect of time period and distance zone only in the earthquake year. This is expressed as a significant three-way interaction effect between time period, distance zone and earthquake year. Such a finding can confirm the hypothesis only if the interaction between them shows that there is a higher anomaly density before or during the earthquake. If the three-way interaction shows a higher anomaly density after the earthquake, then the hypothesis is rejected and anomaly density cannot be considered as a precursory or co-seismic phenomenon. If no significant three-way interaction is found, the hypothesis is rejected. Another possible outcome is that there is a statistically significant combined effect of any two factors, regardless of the third factor. For example, the combination of time period and distance zone may have a statistically significant effect also in the non-earthquake year. Such an effect may be present for reasons other than an earthquake occurrence, and as a result this finding does not confirm the hypothesis. This condition is expressed as a two-way interaction effect, only between two factors. A third possible outcome is that any of the factors (time period, distance zone, earthquake year) separately affect anomaly density, regardless of the effect of their combination. For example, anomaly density may differ between time periods in the whole study area, regardless if there is an earthquake or not. This is expressed as a statistically significant simple main effect, separately for each factor. Such effects indicate that the period, the year and/or the location affect anomaly density regardless of earthquake occurrence, and as a result this finding does not confirm the hypothesis. The last possible finding is that there is no statistically significant interaction between the factors (they have no combined effect on anomaly density) or that there is no significant simple effect of each factors separately. The absence of simple and/or two-way interaction effects does not provide information for the hypothesis. The only meaningful finding that can confirm or reject the hypothesis is the presence or absence of a significant three-way interaction.



## 5.4 Results

### 5.4.1 Anomaly detection

This section presents results of the anomaly detection. Examples, highlighting the results of the spatial and temporal processing, are shown for the earthquakes of Emilia Romagna, Italy, in 2012 (Figure 5.3) and Baja California, Mexico in 2010 (Figure 5.4). These results are compared to results of the same periods and regions in 2011 and 2016, respectively, when no earthquakes of  $M_w > 5$  are recorded in these areas.

#### 5.4.1.1 Spatiotemporal distribution of anomalies

The spatial distribution and the number of anomalies detected in the co-seismic period of the first earthquake of 2012 (19<sup>th</sup>-21<sup>st</sup> May, 48 time observations in total, see details in Table 5.1) is shown in Fig. 5.3a. The earthquake epicentres of Emilia Romagna are shown with red stars in the north. The cluster of anomalies closest to the earthquake epicenter appears around 50km to the east. This cluster (designated *pd* in Figure 5.3) is approximately 100km long and 50km wide. The highest number of anomalies in this cluster, for the whole co-seismic period, is six of a total of 48 observations. The anomalies of this cluster appear in the morning and early afternoon, mostly between 07:00 and 14:00hrs and then disappear. Scattered anomalous pixels can be seen at the same distance from the earthquake, but towards the west. At a distance of around 180 km to the northeast of the earthquake epicenter, another, less spatially extended, cluster of anomalies (*u*) can be seen. The core of this cluster covers 25x25km and has the highest number of anomalies in time for the whole study area (8 anomalies in 48 observations). More clusters of anomalies appear towards the south of the earthquake. One cluster (*ab*) appears at approximately 330 km SE of the epicenter, close to the area where the earthquakes of Norcia, Visso and Amatrice occur over four years later (shown in Figure 5.3 with 3 red stars in the south). The highest number of anomalies in cluster (*ab*) is also 6 out of 48 observations, like in the cluster closest to the earthquake. At 500 km SE of the earthquake, a more extended complex of clusters (*n*) is visible with a high number of anomalies (8 anomalies in the 48 observations, the highest of the whole study area). Finally, a cluster of anomalies (*c*) appears on the southwestern end of the island of Corsica, approximately 600 km SW of the earthquake epicenter. It has less anomalies (4 in 48 observations) and extends to an area as big as the cluster closest to the earthquake.

When the detection threshold is increased to  $\mu + 3\sigma$  (Figure 5.3b), only the highest normalized values are considered anomalous and less anomalies overall are detected. Smaller clusters of anomalies are visible in the study area. The cluster (*pd*) closest to the earthquake epicenter is still visible and has 4

anomalous values in 48 observations, similar to the clusters of Corsica (c) and central Italy (ab) which are located further than 300km from the earthquake epicenter. The highest number of anomalies recorded in time is 5 anomalies in 48 observations and can be found in cluster (n), 500 km SE from the earthquake. Cluster (u) is not visible.

When a longer co-seismic period is considered (19<sup>th</sup>-31<sup>st</sup> May, using the  $\mu+2\sigma$  threshold), to cover all earthquakes of the 2012 sequence, more anomalies are detected (Figure 5.3c). The previously identified clusters (pd), (c), (n), (u) remain but are now more spatially extended. The most extended, and new, cluster is located 200 km northwest of the earthquake epicenter. This cluster (m) shows the highest number of anomalies in the image, 42 in the 240 observations compared to the up to 26 anomalies out of 240 observations close to the epicenter. An area of anomalies (shown as pv, with 200km maximum length and 500km maximum width) connects clusters (m) and (pd). More clusters of anomalies appear along the east (ap, r, tv) and the west (am) coast of Italy, and on the islands of Corsica (c) and Sardinia (just south of cluster c).

When the same processing is applied for the co-seismic period of 19<sup>th</sup>- 21<sup>st</sup> May in a year without large earthquakes (2011, Figure 5.3d), similar patterns are observed as compared to the year with earthquake (2012, Figures 5.3a,c). Previously identified clusters in the north and south (m, pd, u, n) are again visible. The anomalies in the north are connected in an extended cluster with maximum length of  $\sim 100\text{km}$  and a width of  $\sim 500\text{km}$  (pv). Clusters appear in the centre of Italy (at) and at the east coast (at, am). Scattered smaller clusters are visible throughout central Italy. The maximum number of anomalies in the year without earthquake is higher than in the year of the earthquake (17 out of 48 observations in 2011, compared to 8 out of 48 observations in 2012) and is found in cluster (m).

Anomalies have been reported to appear earlier than one day before the earthquake. For the earthquakes of the example from Italy, Qin et al [2012] report "local temperature enhancements around the epicentres on the night of May 12, 2012, i.e., 8 days before the May 20, 2012 earthquake". They present an anomaly on an image corresponding to 20:00hrs on May 12<sup>th</sup> 2012 (shown in Figures 5.3e-g as the shaded area). They find anomalies at the same location in an image showing daily average temperatures of May 12<sup>th</sup>, 2012. We detect no anomalies in the image of 20:00hrs (Figure 5.3f). In terms of daily averages, Figure 5.3e shows the total numbers of anomalies we detected for the full day of May 12<sup>th</sup>, 2012. A cluster of anomalies (pd) is visible close to the earthquake epicenter, which is the most spatially extended cluster in the image. Half of the area of the cluster coincides spatially with the anomaly reported by Qin et al [2012]. The highest number of anomalies in this cluster (but not in the whole image) is six out of a total of 24 observations, and all six

appear between the morning and the early afternoon. A cluster of anomalies of the same intensity but covering a slightly smaller spatial extent than the one close to the earthquake, is visible in Figure 5.3e, over the island of Corsica (c). Clusters (n) and (ab) are also visible but have smaller spatial extent than in the co-seismic period. The highest number of anomalies in the image, that is seven out of a total of 24 observations, appears in cluster (mp) on the west coast of Italy.

Cluster (pd), which is visible to the east of the earthquake epicentres in the pre- and co- seismic period of 2012, is also visible in other years and days. An example from a post-seismic period (27<sup>th</sup> of September 2011) is shown in Figure 5.3g. On that day the cluster covers the earthquake epicenter and partially coincides with the area where Qin et al [2012] found an anomaly in 2012. Clusters (m), (at), (pv), (u), (at) and (c) are also visible.

We aggregate and plot numbers of anomalies for longer periods to show how detected anomalies vary with time during the year. An example is shown in Figure 5.4 for the study area of Baja California. In the figure, anomalies are presented as three-month averages for two years: 2010 (year of the earthquake) and 2016 (year without earthquake). The area (c) closest to the earthquake shows higher numbers of anomalies between April and June. This happens in the earthquake year 2010 (earthquake epicenter shown with a red star) and also in year 2016 with no  $M > 5.5$  earthquake. The mountainous areas (a) in Arizona (Grand Canyon, Mongolon Rim) show a higher numbers of anomalies from the beginning of the year until June, in both years (less pronounced in 2016). The mountain range of the Sierra Madre Occidental (b) produces anomalies throughout the year, most prominently between September and December.

In the examples above, but also in all studied areas, localized anomalies are found throughout the duration of the datasets and throughout the spatial extent of the study areas. Clusters of anomalies with high intensity are present in the years with and without an earthquake, and at different periods of the year.

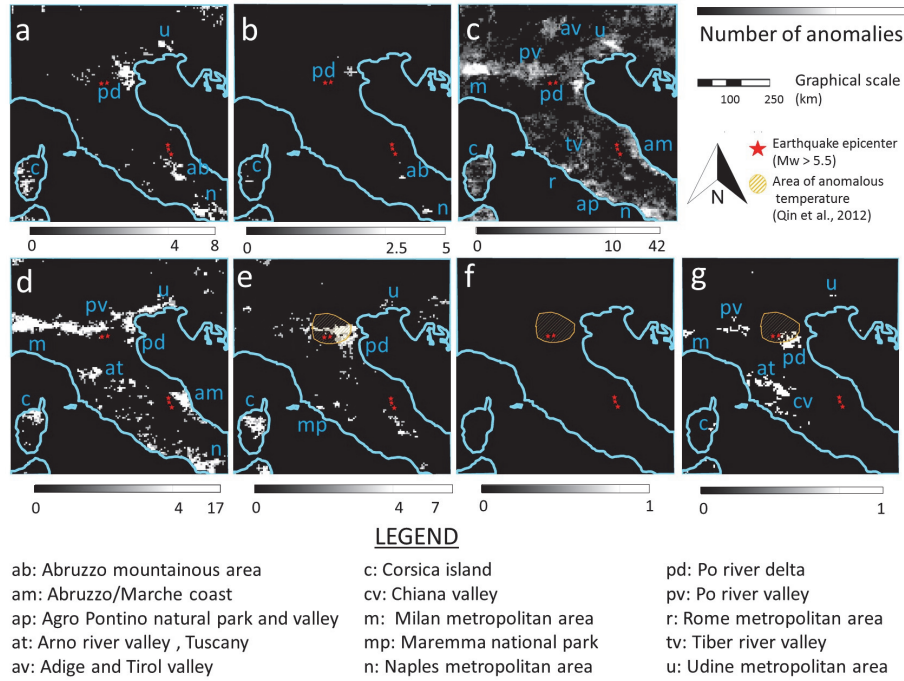


Figure 5.3. Anomalies detected in the study area of Italy. Earthquake epicentres (stars) in the north correspond to the Italian Mw>5 earthquakes of 2012. Earthquake epicentres (stars) in the southeast correspond to the Italian Mw>5 earthquakes of 2016. Panel (a): detection results for 19<sup>th</sup> -21<sup>st</sup> May 2012 (co-seismic period, earthquake year) using a  $\mu+2\sigma$  detection threshold. Panel (b): the same as in (a) but using a  $\mu+3\sigma$  threshold. Panel (c): detection results for 19<sup>th</sup> -30<sup>th</sup> May 2012 (prolonged co-seismic period, earthquake year). Panel (d): detection results for 19<sup>th</sup> -21<sup>st</sup> May 2011 (year without earthquakes). Panel (e): detection results for the whole day of May 12<sup>th</sup>, 2012. Panel (f): detection results for May 12<sup>th</sup>, 2012, 20:00hrs, when an anomaly is reported by Qin et al. [2012] (shaded area in the North). Panel (g): detection results for the 27<sup>th</sup> of September, 2011 (year with no earthquakes).

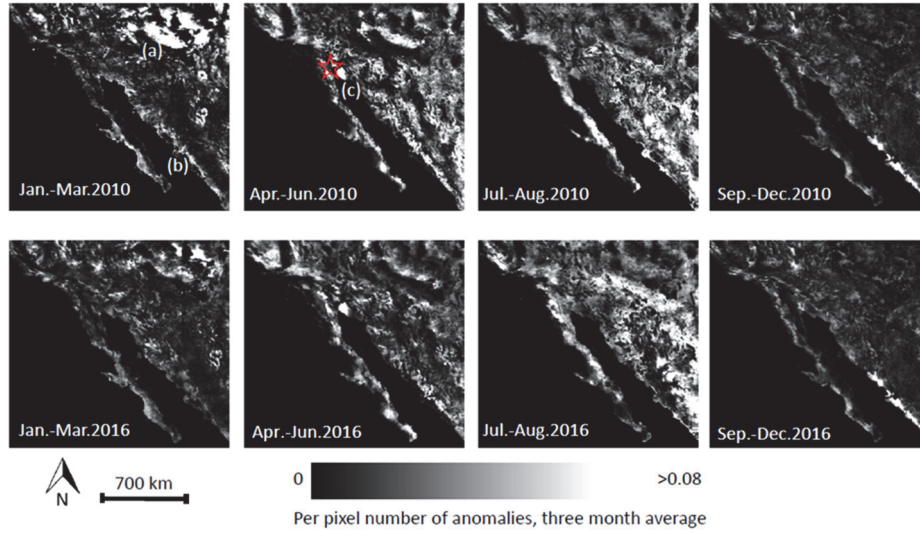


Figure 5.4. Per pixel numbers of anomalies, averaged every three months, in the study area of Baja California (details in Table 5.1) for the years 2010 (upper row, with earthquake) and 2016 (lower row, without earthquake). Areas with higher numbers of anomalies are shown in white tones. The red star denotes the epicentre of the 2010 earthquake.

#### 5.4.1.2 Anomaly density

We calculate anomaly density values for each spatial zone and period, to have comparable results in periods of different duration (for example, the co-seismic period of 48 hours and the pre-seismic period of 2 months). Anomaly density values for all earthquakes are provided in Appendix 2. The mean anomaly density of spatial zones further from the earthquake (distance  $>70\text{km}$ ) is higher than in zones closer to the earthquake (Figure 5.5a,b). In terms of time, post-earthquake periods have the highest mean anomaly density and the co-seismic period has the lowest (Figure 5.5c-f).

#### 5.4.2 Statistical analysis

Statistical analysis is performed to conclude if there is a significant relation between the occurrence of earthquakes and detected anomalies. Such a relation would be established only if anomaly density is higher close to earthquake epicentres, before/during the earthquake and only in the years with an earthquake. The analysis is repeated for the following experimental configurations: use of different thresholds and with/without adjustment of the co-seismic period. Additionally, the analysis shows the statistical significance of the differences of average anomaly density between different zones/periods (Fig. 5.5). Table 5.3 presents the results of each ANOVA test, in the different experimental configurations. A technical description of the tests is provided in Appendix 3.

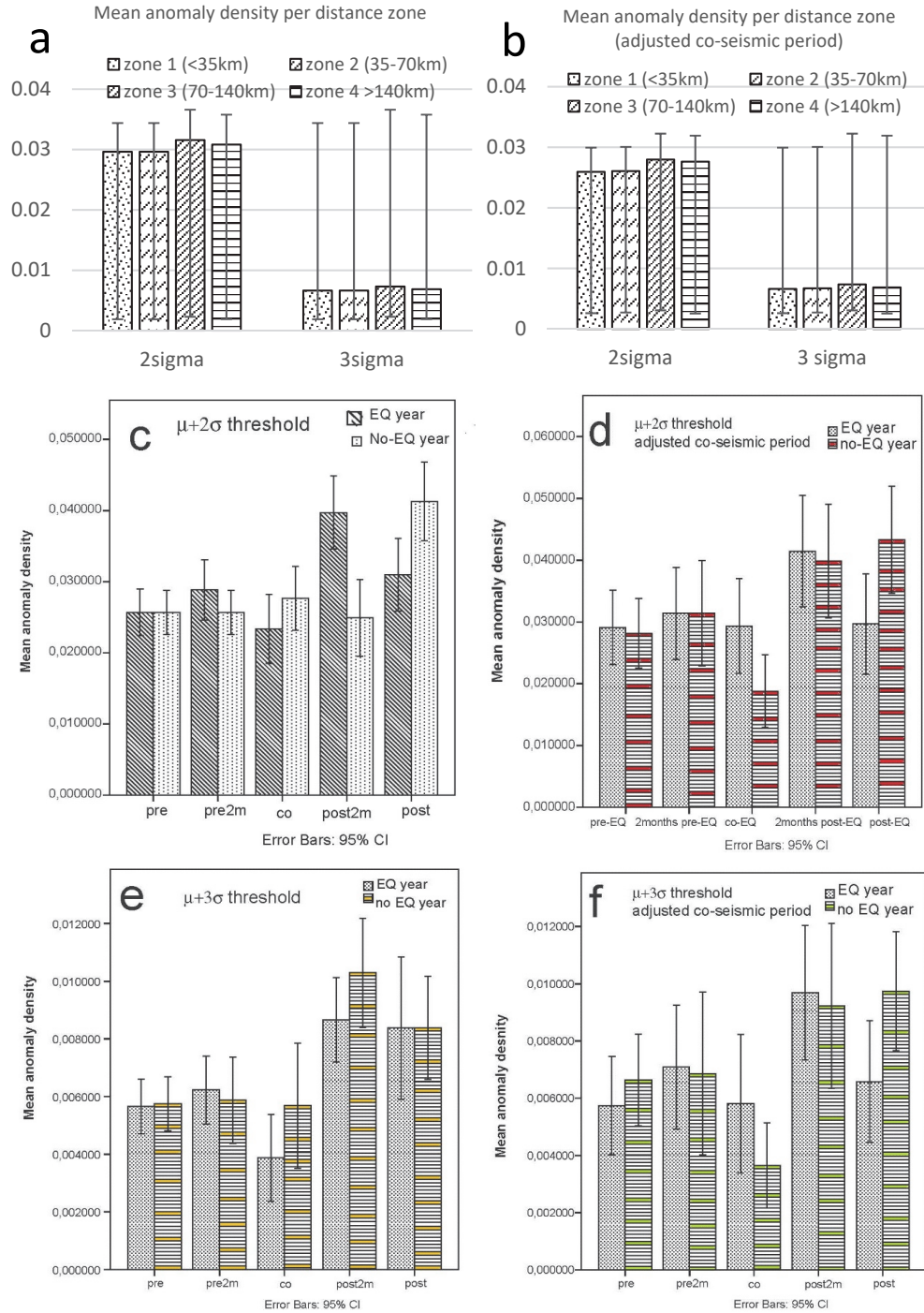


Figure 5.5 Anomaly density for all studied earthquakes in earthquake and no-earthquake years, summarized by distance zone (panels a,b) and by period (panels c,d,e,f). Results are shown for all different configurations ( $\mu+2\sigma$  and  $\mu+3\sigma$  threshold; with and without adjustment of the co-seismic period).

interaction between earthquake year, period and distance zone (Line 1, Table 5.3). This means that there is no significant difference in anomaly density found only in the year of the earthquake. It is thus not possible to say that there are more anomalies close to the earthquake, before/during the earthquake, and only in the year of the earthquake. Any differences between the anomaly density in different zones and periods are present in earthquake and no-earthquake years alike, and they cannot be attributed to the occurrence of the earthquake. This is the only statistical result related to the hypothesis that there is a relation between earthquakes and anomalies, and the findings reject the hypothesis.

Statistical analysis further provides indications of how anomaly density differs in different years, locations and periods, regardless of the earthquake (Lines 2-7 of Table 5.3). Regardless of the experimental configuration, location relative to the earthquake epicentre does not affect anomaly density at any time (lines 2, 4). Distance from the earthquake by itself has no significant effect on anomaly density (line 6). Anomaly density differs with period and year, depending on the experimental configuration (line 3). Period by itself has a statistically significant effect in anomaly density of different years and distance zones (line 5). This effect is found in all cases except when an adjusted co-seismic period is applied with a  $\mu+3\sigma$  threshold. Year by itself has a significant effect on anomaly density only when adjusted co-seismic periods are applied in combination with a  $\mu+3\sigma$  threshold (line 7).

Overall, the result of the statistical analysis is that regardless of the applied threshold and the definition of the co-seismic period, there is no significant effect of the earthquakes in the anomaly density measured across distance zones and periods.

Table 5.3. Summary of the results of ANOVA statistical tests. *F*-values show to what extent the variance observed in the sample results from a specific effect/interaction, rather than from chance. The *p*-value shows the probability of obtaining the calculated values in the absence of a specific effect/interaction. Lower *p*-values show higher probability that an effect/interaction is statistically significant. In the Table, "NONE" denotes that no statistically significant effect/interaction is found; "YES" shows that a statistically significant effect/interaction is found. Significance levels for these effects are denoted with \* ( $p < .05$ ) or \*\* ( $p < .001$ ). Partial  $\eta^2$  is a measure of the size of an observed statistically significant effect, and denotes how much the effect contributes to the total variation in the sample. More information on ANOVA can be found, for example, in Maxwell and Delaney [2004].

Effect/ interaction	$\mu+2\sigma$ threshold		$\mu+3\sigma$ threshold	
	48-hr co-seismic period	Adjusted co-seismic period	48-hr co-seismic period	Adjusted co-seismic period
Effect of earthquake across distance zones and periods	<b>NONE</b> $F(6.094, 134.060) = .162$ , $p = .987$ , partial $\eta^2 = .007$	<b>NONE</b> $F(6.767, 121.798) = .596$ , $p = .753$ , partial $\eta^2 = .032$	<b>NONE</b> $F(5.210, 116.355) = .148$ , $p = .983$ , partial $\eta^2 = .007$	<b>NONE</b> $F(8.507, 155.958) = .422$ , $p = .915$ , partial $\eta^2 = .022$
Interaction effect of period and distance, regardless of year	<b>NONE</b> $F(8.914, 196.116) = 1.645$ , $p = .106$ , partial $\eta^2 = .070$	<b>NONE</b> $F(6.832, 122.971) = .574$ , $p = .772$ , partial $\eta^2 = .031$	<b>NONE</b> $F(8.905, 198.888) = 1.224$ , $p = .282$ , partial $\eta^2 = .052$	<b>NONE</b> $F(7.604, 139.413) = .788$ , $p = .608$ , partial $\eta^2 = .041$
Interaction effect of period and year, regardless of distance	<b>NONE</b> $F(2.031, 66) = .959$ , $p = .387$ , partial $\eta^2 = .014$	<b>YES**</b> $F(2.256, 54) = 6.954$ , $p = .001$ , partial $\eta^2 = .114$	<b>NONE</b> $F(1.737, 67) = .934$ , $p = .384$ , partial $\eta^2 = .014$	<b>YES**</b> $F(2.836, 55) = 6.485$ , $p < .001$ , partial $\eta^2 = .105$
Interaction effect of year and distance, regardless of period	<b>NONE</b> $F(3, 66) = .844$ , $p = .475$ , partial $\eta^2 = .037$	<b>NONE</b> $F(3, 54) = 2.279$ , $p = .090$ , partial $\eta^2 = .112$	<b>NONE</b> $F(3, 67) = .008$ , $p = .999$ , partial $\eta^2 < .001$	<b>NONE</b> $F(3, 55) = .626$ , $p = .601$ , partial $\eta^2 = .033$
Simple effect of period	<b>YES**</b> $F(2.971, 196.116) = 11.738$ , $p < .001$ , partial $\eta^2 = .151$	<b>YES*</b> $F(2.277, 122.971) = 4.787$ , $p = .007$ , partial $\eta^2 = .081$	<b>YES**</b> $F(2.968, 116.355) = 9.927$ , $p < .001$ , partial $\eta^2 = .129$	<b>NONE</b> $F(2.535, 139.413) = 6.631$ , $p < .001$ , partial $\eta^2 = .108$
Simple effect of distance	<b>NONE</b> $F(3, 66) = .333$ , $p = .801$ , partial $\eta^2 = .015$	<b>NONE</b> $F(3, 54) = .060$ , $p = .980$ , partial $\eta^2 = .003$	<b>NONE</b> $F(3, 67) = .084$ , $p = .969$ , partial $\eta^2 = .004$	<b>NONE</b> $F(3, 55) = .468$ , $p = .706$ , partial $\eta^2 = .025$
Simple effect of year	<b>NONE</b> $F(1, 66) = .284$ , $p = .596$ , partial $\eta^2 = .004$	<b>NONE</b> $F(1, 54) = 1.967$ , $p = .167$ , partial $\eta^2 = .035$	<b>NONE</b> $F(1, 67) = 1.834$ , $p = .180$ , partial $\eta^2 = .027$	<b>YES**</b> $F(1, 55) = 14.914$ , $p < .001$ , partial $\eta^2 = .213$



## 5.5 Discussion

### 5.5.1 Spatial extent of anomalies

In published research, the spatial extent of detected anomalies is often not clear. For example, Akhoondzadeh [2013a,b] and Pulinets et al. [2014] study the 6.4Mw earthquake of Ahar, NW Iran (2012) using data from meteorological stations. Such data have very limited spatial cover and cannot describe the full extent of detected anomalies. Even when Akhonzadeh et al [2013b] use satellite data (MODIS LST), they confine their anomaly detection to an area 5x5km over the earthquake epicenter. Other researchers study larger areas using data with coarse spatial resolution. For example, Alvan et al. [2014], who study the event of Baja California (2010), report anomalies in gridcells of a spatial extent of  $2^{\circ} \times 2^{\circ}$ . For the same event Jie and Guangmeng [2013] find anomalies in an area of  $1^{\circ} \times 1^{\circ}$  over the epicenter. Qin et al [2012a, b] and Wu et al [2012], who report anomalies in New Zealand (2010) and Italy (2012), use data with a  $2^{\circ} \times 2^{\circ}$  spatial resolution. When reported anomalies extend to degrees latitude/longitude, it is not possible to trace back in detail their exact origin and their spatial relation to faults and other potential causative processes. Anomaly detection in this study is based on gridded satellite data of 5x5km spatial resolution, which allows a more detailed description of the location of anomalies with respect to the earthquake. For example, reported anomalies in Italy and Mexico, with an extent of  $2^{\circ} \times 2^{\circ}$  [Qin et al., 2012; Jie and Guangmeng, 2013] spatially coincide in our results with anomalies scattered over different locations. The finer spatial resolution allows for a better understanding of potential causative phenomena of the detected anomalies, like the existence of wetlands or agricultural areas (Figure 5.3).

Furthermore, the normalization procedure we apply ensures that detected anomalies have a spatial extent less than the normalization frame (up to 18225km<sup>2</sup>). It is possible to distinguish localized anomalies from larger-scale weather patterns and climatic trends. That could explain the cases when other researchers find anomalies and our results show none. In our findings, extended patterns, like the thermal inversion described by Qu et al [2006] as a more probable reason for the appearance of a pre-earthquake anomaly in Mongolia, are suppressed. An unusually warm day in the whole study area is not linked to the earthquake, and anomalies are not masked by unusually cold days.

### 5.5.2 Distance between anomalies and earthquakes

In early literature on the topic, this distance was not taken into account; anomalies were traced as far as 500km away from the epicenter [Tronin, 2000] or further [Tramutoli et al., 2005, Qiang et al., 1997]. Piroddi and Ranieri [2012] argue that observable phenomena further than 60km from the

earthquake, even if they were spatially associable to the seismic event, they would not be practically useful as precursors because the potential alarm areas would be too big. Yang and Guo [2010] trace anomalies in the range of 65-100km from the epicenter of a past earthquake and state that, if thermal anomalies are to be used to estimate a possible future epicenter, the biggest error will be less than 100 km. We address the issue of the distance between anomalies and epicentres by classifying all detected anomalies in four different spatial zones, based on their distance from the earthquake. The two zones closest to the earthquake correspond to the suggestion of Piroddi and Ranieri [2012] and are used to distinguish areas in the more or less immediate vicinity of the earthquake. In our study areas, anomaly density is on average higher in zones further away from the earthquake, and the analysis shows that even this difference is not statistically significant. In our results, no difference is found in anomaly density between the zones. The location of the earthquake has no effect on anomaly density.

#### 5.5.3 Temporal resolution

Previous studies use daily averages or one image per 24 hours for detection [e.g. Zhang et al., 2010; Akhoondzadeh, 2013a, b]. We use one image every hour and construct hyper-temporal time series to allow the study of transient anomalies. The temporal persistence of an anomaly can provide information to explore potential causes of the anomaly. In the example of Italy, the anomaly closest to the earthquake is found only in six images in the co-seismic period of 48 observations. It appears early in the day (between 07:00 and 14:00) and it is recurrent in different days and years. There is no physical evidence to support why earthquake effects appear, disappear and reappear within hours at preferred times of the day, and persist in years without an earthquake. An alternative explanation for fluctuations at this scale could be evaporation supported by increasing incoming solar radiation at the beginning of the day and/or early morning ground fog, which in the specific example is plausible because the anomaly is located over the intensely cultivated delta of a river.

#### 5.5.4 Temporal coincidence with the earthquakes

We examine complete years, and the same area is examined with and without earthquake occurrence. This allows us to test for presence of anomalies in periods without earthquakes. Previous studies test a variety of periods just before and after the earthquake but often not full years [Alvan et al., 2014, Jie and Guangmeng, 2013, Akhoondzadeh, 2013a]. This leaves the possibility that similar anomalies may be occurring regularly throughout the year(s) without any earthquake effect, as found by Blackett et al. [2011]. Our results show that anomalies in the vicinity of earthquake epicentres, as well as everywhere in the study areas, appear throughout the years, regardless of an earthquake occurrence.

### 5.5.5 Statistical analysis

We investigate the link between earthquakes and anomalies with the application of statistical analysis following the suggestions of Eneva et al. [2008]. We take all anomalies into account. The statistical analysis shows that detected anomalies appear regardless of earthquake occurrence, in terms of time and of distance from earthquake epicenter. These results are in agreement with Eneva et al. [2008], who applied a similar statistical design reaching the same conclusions. Zhang et al. [2013], following a different statistical scheme, and also conclude that in their data the correlation between occurrence of earthquakes and detected anomalies was very low. Changing the anomaly threshold or the period of investigation (to a larger co-seismic period) in our work, does not change the results. Furthermore, no recognizable patterns are found in the case of earthquakes that repeatedly appear in similar locations (Italy, New Zealand and Oklahoma).

### 5.5.6 Atmospheric and other environmental influences

In our results, period affects anomaly density, even though this effect is present regardless of earthquake occurrence. A possible reason why time can affect anomaly density could be localized temporal variations of atmospheric parameters (as found, for example, by Jie and Guangmeng [2014]; Qu et al. [2006]). Water vapour, in particular, influences LST retrievals [Freitas et al., 2010]. An atmospheric analysis can indicate potential relation of detected anomalies with local atmospheric influences, like water vapour and rain water. We discuss here, as example, the four earthquake cases with the highest magnitude, that belong to areas of different climates: (1) the Mw7.8, 2016 earthquake in Kaikoura, New Zealand; (2) the Mw7.2, 2010 earthquake in Baja California, Mexico; (3) the Mw7.1, 2011 earthquake in Van, Turkey and (4) the M6.9, 2011 earthquake in Myanmar. We extract monthly data from the ERA5 climate Reanalysis (<https://climate.copernicus.eu/products/climate-reanalysis>), at hourly temporal resolution and at a spatial resolution of 31km) for the following variables: total column water vapour (TCWV), Total Column Rain water (TCRW), Air Temperature ( $T_{\text{air}}$ ) and Surface Temperature ( $T_{\text{skin}}$ ). We compare monthly reanalysis data with monthly numbers of anomalies detected over the epicentral areas. We observe an inverse relation between the monthly number of anomalies and total column rain water (TCRW) in all study areas (Figure 5.6). This could be because no TIR signal can be recorded by the satellite sensor while it is raining, and as a result no anomalies can be detected. In Mexico (Fig. 5.6b) and in Turkey (Fig. 5.6c), high numbers of anomalies appear in periods with high values of water vapour (TCWV) and high Skin and Air temperatures. In Myanmar, TCWV,  $T_{\text{skin}}$  and  $T_{\text{air}}$  are inversely related to the number of anomalies (Fig. 5.6d). This may indicate that increased water vapour or skin and air temperatures do not vary locally but in more extended spatial scales. In the case of New Zealand (Fig. 5.6a) the relation between

$T_{skin}$ ,  $T_{air}$ , TCWV and number of anomalies is unclear because the latter is fluctuating throughout the year. Overall, these observations indicate a potential relation with localized atmospheric effects. With the exception of TCRW, these relations are site-specific and cannot be generalized. Other influences may also be present and should be considered in combination with atmospheric influences. For example, as seen in Figure 5.3, in the case of Italy, strong anomalies spatially coincide with wetlands, cultivated land, urban/industrial areas, river valleys and complex terrain (mountainous or mountainous and coastal areas). Further study of LST variability may provide more information but is out of the scope of this work.

### 5.5.7 Limitations

In order to set the detection threshold and to define co-seismic periods, we used information from literature on earthquake preparation area definition, fault lengths and earthquake sequences. However, we acknowledge that existing literature is still exploratory. To this moment, there is no physically-based detection threshold or definition of a co-seismic period. In recognition of this, we test different settings to evaluate the sensitivity of our results to those settings. We find that the main results are robust to the use of different settings, and they remain the same regardless of the use of stricter detection thresholds or adjusted co-seismic period.

Regarding the size of normalization frame, we set it based on literature on the earthquake preparation area. It may be argued, however, that there may be links between earthquakes and anomalies appearing hundreds of kilometres away from the earthquake without being present in the immediate proximity of the event, or that earthquake-related influences extend over areas larger than 140x140km. In such cases, the frame we applied may be considered too strict and a future researcher could choose to apply normalization frames of different sizes. To this moment, the theoretical background behind such occurrences is not yet established and their precursory value may be limited, as discussed earlier.

We have examined 20 earthquakes. This collection is diverse in terms of fault mechanism, climatic conditions, topography and land cover, but it still is only a sample. An ideal sample would contain multiple similar large earthquakes with different focal mechanisms, in areas with different local conditions. Their occurrence should be covered by comparable satellite sensors and unperturbed years should be included for statistical analysis of anomaly occurrence. Further research, supported by the release of longer LST datasets, should be systematically extended to include more earthquake cases, of different characteristics and from different regions. This could increase representativeness of the sample and reduce the variability in the results of statistical analysis.

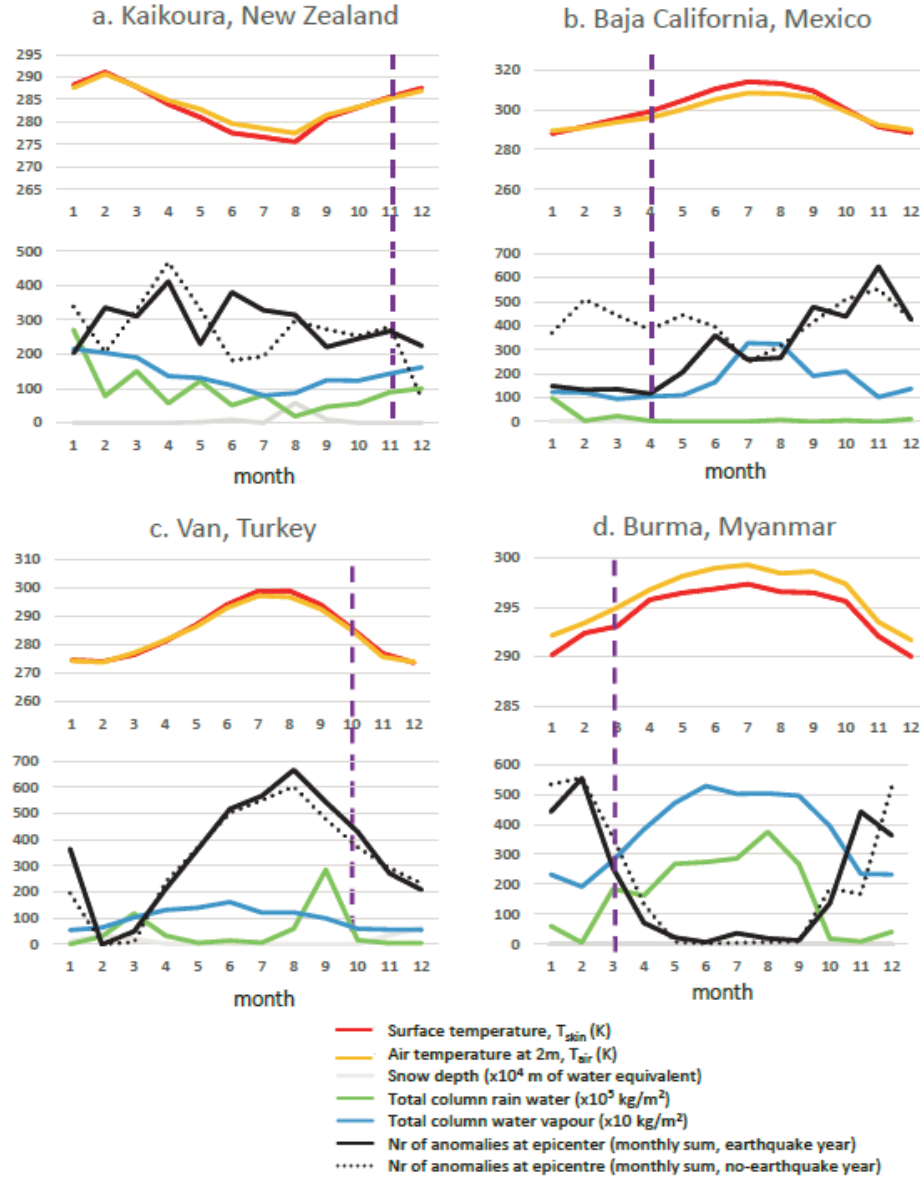


Figure 5.6. Relation between atmospheric parameters and anomalies in four study areas. The graphs show monthly values of Skin Temperature ( $T_{skin}$ ), Air Temperature ( $T_{air}$ ), Total column water vapour (TCWV), Total column rain water (TCRW) and Snow depth (SD) of ERA5 reanalysis (earthquake year); and monthly numbers of anomalies detected over the epicenter of the earthquakes of Kaikoura, NZ (panel a); Baja California, Mexico (panel b); Van, Turkey (panel c) and Burma, Myanmar (panel d). Numbers of anomalies are provided for the earthquake year (solid line) and the no-earthquake year (dashed line). The time of the earthquake event is shown with vertical purple dashed lines; there is no visible effect of the earthquake on the monthly numbers of anomalies.

## **5.6 Conclusions**

The study of hyper-temporal geostationary-based LST records to detect anomalies of spatial extent up to 18225 km<sup>2</sup> in ten earthquake-prone areas, yields anomalies throughout the year-long datasets and throughout the study areas, regardless of the occurrence of an earthquake. The results show no statistically significant difference in anomaly density between years with and without earthquake, in the same location; and no difference among locations closer or further to the earthquake, within the same study area and year. No recognizable patterns are identified when earthquakes occur in the same region in different years. It can be concluded that the twenty earthquake events tested do not significantly affect LST. The findings suggest that any earthquake-related effect, if it exists, it is weaker than other environmental influences or spatially more extensive than what is tested.

### **Acknowledgments, Samples, and Data**

LST datasets were obtained online from the LSA-SAF of EUMETSAT (<http://landsaf.meteo.pt/>). ERA-5 data sets are available through the Copernicus Climate Change Service (2017) and were retrieved through the Meteorological Archival and Retrieval System (MARS) of ECMWF. Statistical analysis was performed using the software IBM SPSS statistics for Windows, Version 24.0.

## Chapter 6. Uncertainty propagation in the normalization procedure

### 6.1 *Introduction*

"Assuming reality exists and is clearly defined, error is the difference between reality and our representation of reality" [G. Heuvelink, lecture notes on Uncertainty propagation]. Uncertainty is an expression of doubt in our representation of reality as a result of our limited knowledge and errors in measurements. All satellite measurements are affected by errors, which propagate through the process of estimating variables based on the satellite measurements. A full uncertainty budget considers different sources that may introduce error throughout the processing chain: physical effects (like the attenuation of emissions due to the presence of aerosols in the atmosphere), measurement/observational effects (related to the function or the positioning of the sensor), classification effects (for example, the classification of a pixel as cloud covered or clear-sky pixel) and effects introduced by the mathematical procedures of estimating derived parameters [Merchant et al., 2017].

It is important to quantify how the uncertainty of the LST product propagates through our processing, and to understand potential factors which play a role. This can lead to more confidence that the results of anomaly detection are not caused by the uncertainty of LST retrievals. It can further provide insight on how input LST uncertainty may influence anomaly detection.

This chapter provides a quantification of the uncertainty propagating through the normalization procedure described in Chapter 3, given the uncertainty estimates provided with the LST product. Uncertainty information is needed to determine the degree of confidence that the anomalies detected in the previous chapter do not fall within the limits of LST input uncertainty. Furthermore, it can support confidence in detection results based on the application of the normalization procedure on LST data in other case studies.

We study LST uncertainty propagation in types of areas where the lowest accuracy of LST retrievals is reported [Freitas et al, 2010, Trigo et al. 2009], in order to quantify the highest expected uncertainties. Uncertainty propagation is quantified following three different approaches. The results are compared and potential influences on uncertainty are explored. Finally, an evaluation of detection sensitivity in view of the estimated uncertainties concludes the chapter.

## **6.2 Data**

In the earthquake case study, anomalies are detected in Land Surface Temperature data from the Copernicus Global Hourly Land Surface Temperature (CGH-LST) product. The CGH-LST product provides estimates of the radiative skin temperature over land at a 5km spatial resolution, using longwave infrared instantaneous satellite observations with atmospheric correction and cloud masking. The data are accompanied by estimates of standard uncertainty. These take into account error sources such as sensor noise, effects of viewing geometry, total water vapour content (TCWV) and statistics of TCWV forecasts, and uncertainty in surface emissivity [Freitas et al., 2010, Trigo et al., 2009]. Uncertainty estimates are described as standard uncertainty (standard deviation) per pixel. According to the data provider, the types of regions of highest uncertainty in LST retrieval are arid areas, mountainous areas and areas at the edge of the METEOSAT swath [Trigo et al., 2009, Bento et al., 2017].

The LST value of each pixel can be described as  $LST \pm u$ , where  $u$  is the standard uncertainty. Four study areas are chosen because of reported high levels of uncertainty related to LST retrieval: the El Oued desert, in Algeria (arid area); the Alps, over Trentino, Italy (a mountainous area); the area over Lake Van, Turkey (at the edge of the METEOSAT swath); and an active volcanic area over Mount Etna, Sicily, Italy. The latter is of interest because volcanic masks are not considered in the retrieval of LST in the product, and this may introduce additional uncertainty in LST retrievals because volcanic eruptions alter atmospheric conditions, including aerosol and water content. All temporal subsets cover a period of one year. Data from near-by ground-based meteorological stations, retrieved online from <https://www.wunderground.com/> are used to examine the atmospheric factors that contribute to the estimated uncertainties.

## **6.3 Methodology**

Both of the uncertainty propagation approaches described by Heuvelink [1999] are applied for estimating uncertainty of normalized values, namely a Taylor series method and a Monte Carlo method. These methods are used because they are relevant for studying spatiotemporal uncertainty propagation. Other methods for uncertainty estimation, like resampling [Theodorsson, 2017] are not considered because they are not applicable in the context of this research, as they are intended for calculations of uncertainty in measurements. Additionally, a fixed value analysis is applied similarly to Haywood et al. [2002]. This is not strictly a method for estimating uncertainty propagation but is used to create deterministic bounds of expected normalized values. Results of the three methods are compared, and finally the estimated uncertainty is



examined in relation to atmospheric effects and missing values in the normalization frame.

### 6.3.1 Uncertainty of the normalization frame

Normalized values are the result of dividing the value of the central pixel with the average of its neighbours,  $T_{norm} = \frac{LST_{central}}{LST_{frame}}$  (Eq.6.1). The uncertainty of each pixel separately is readily available with the LST product, but the uncertainty of the frame of neighbouring pixels needs to be calculated.

Let the uncertainty of the normalization frame be  $u_f$ . The uncertainty of the normalization frame is calculated in all areas for a frame of eight pixels. This is the smallest possible frame that can be used for a normalization, and it consists of the immediate neighbours of a pixel. Adjacent pixels show high spatial correlation and the correlated uncertainty component is maximized. As a result, by studying a normalization frame of 8 pixels, we can quantify the highest expected levels of uncertainty.

The number of pixels in the frame is  $N = N_{ex} + N_{mis} = 8$ , where  $N_{ex}$ : existing pixels and  $N_{mis}$ : missing pixels. The variance of LST values of existing pixels in the frame is designated as  $Var_{LST}$ . Let  $u_n = u_{1,2,...,N_{ex}}$  the uncertainty of each of the existing pixels in the frame. The uncertainty of the frame ( $u_f$ ) consists of the contribution of the uncertainties of the frame pixels ( $u_n$ ) and the sampling uncertainty  $u_{samp}$ , which are calculated as follows:

- (a) the correlated uncertainties  $u$  of each of the frame pixels, with a total contribution of

$$u_{cor} = \sqrt{\frac{\sum_{ex} u_n^2}{N_{ex}}} \quad [\text{Farrance and Frenkel, 2012}], \quad (\text{Eq.6.2})$$

and

- (b) the sampling uncertainty  $u_{samp}$  which arises from the presence of missing values in the frame, and is estimated as

$$u_{samp} = \frac{N_{mis} Var_{LST}}{N_{mis} + N_{ex} - 1} \quad [\text{Dodd et al., 2016}] \quad (\text{Eq.6.3})$$

The two components described above are not correlated and can be therefore added in quadrature to provide the total uncertainty of the frame [Theodorsson, 2017]:

$$u_f = \sqrt{u_{cor}^2 + u_{samp}^2} \quad (\text{Eq.6.4})$$

### 6.3.2 Taylor series method

Following the Law of Propagation of Uncertainty (LPU, as presented in Theodorsson [2017]), the uncertainty of normalized values is expressed as a function of the uncertainties of the central pixel and the normalization frame. This function is approximated using a first order Taylor expansion [Coleman and Steele, 2009].

Let the uncertainty of the central pixel be  $u_c$  and of the frame,  $u_f$ . The correlation coefficient between  $u_c$  and  $u_f$  is denoted as  $\rho$  and LST values of the central pixel and the frame are shown as  $LST_c$  and  $LST_f$  correspondingly. Using a first order Taylor series expansion, the variance  $\tau^2$  of normalized values can be approximated as

$$\tau^2 \approx \frac{u_c^2}{LST_f^2} + u_f^2 \frac{LST_c^2}{LST_f^4} - 2\rho u_c u_f \frac{LST_c}{LST_f^3} \quad [\text{Heuvelink, 1999}] \quad (\text{Eq.6.5})$$

The square root of this variance describes the standard uncertainty of each normalized value. The calculation is repeated for all normalized values of the pixel's time series.

### 6.3.3 Monte Carlo method

According to this approach, the probability distribution of normalized values is obtained by performing  $m$  repetitions of normalization using a sample of random  $LST_{frame}$  and  $LST_{central}$ , generated from the joint distribution of  $LST_{frame}$  and  $LST_{central}$  [Heuvelink et al., 2010]. The previously estimated  $u_c$ ,  $u_f$  and  $\rho$  are used for the calculations. For a bivariate normal joint distribution for  $LST_{frame}$  and  $LST_{central}$ , it is  $X \sim N_2(\mu, \Sigma)$ , where  $\mu$  is the vector of mean  $LST_{frame}$ ,  $LST_{central}$  and  $\Sigma$  is the 2x2 covariance matrix:

$$\mu = \begin{bmatrix} \mu_{frame} \\ \mu_{central} \end{bmatrix}, \quad \Sigma = \begin{bmatrix} u_f^2 & \rho u_f u_c \\ \rho u_f u_c & u_c^2 \end{bmatrix} \quad [\text{Heuvelink, 1999}] \quad (\text{Eq.6.6})$$

The simulation is performed with 40000 repetitions in each timeslot. The result provides information not only about the mean and the standard uncertainty, but also the full probability density function of the normalized values.

### 6.3.4 Fixed value analysis

We constrain the maximum and minimum limits within which the normalized values are expected to lie, given the previously calculated  $u_c$  and  $u_f$ . The limits are defined as follows:

$$Upperlimit = \frac{LST_{central} + u_c}{LST_{frame} - u_f} \quad Lowerlimit = \frac{LST_{central} - u_c}{LST_{frame} + u_f} \quad (\text{Eq.6.7,6.8})$$

The calculation is performed in every timeslot resulting in an upper-limit time series and a lower-limit time series. The boundaries are defined using the first

interquartile of the lower-limit series and the third interquartile of the upper-limit series.

## 6.4 Results and discussion

Uncertainty estimates, following both the Taylor and Monte Carlo approaches, provide information on the uncertainty of normalized values. Fixed value analysis provides the limits within which normalized values are expected. The above are shown for all study areas in Table 6.1, where standard uncertainty is shown along with mean normalized values.

*Table 6.1. Results of the three methods. For the Taylor and Monte Carlo approaches, expected normalized values are given as mean  $\pm$  standard deviation. For the fixed value analysis, the range of expected normalized values is provided.*

Area	1 <sup>st</sup> order Taylor	Monte Carlo	Fixed value analysis
<b>Desert</b>	0.999 $\pm$ 0.0016	0.999 $\pm$ 0.0030	[0.98, 1.03]
<b>Volcano</b>	0.982 $\pm$ 0.0045	0.983 $\pm$ 0.0015	[0.94, 1.02]
<b>Mountain</b>	0.991 $\pm$ 4.04 $\times 10^{-7}$	0.991 $\pm$ 1.38 $\times 10^{-7}$	[0.99, 0.99]
<b>Edge of swath</b>	1.0 $\pm$ 1.5 $\times 10^{-7}$	1.0 $\pm$ 1.3 $\times 10^{-7}$	[0.99, 1.00]

Detailed examination of results of all methods show low levels of uncertainty in normalized values and confirm that the highest uncertainties correspond to the volcanic and the arid study area. In these two cases also the boundaries of the fixed value analysis are wider than in the other two regions (Table 6.1). The differences between the uncertainty estimates resulting from different methods are minimal. In particular, the mountainous study area and the area at the edge of the swath, the differences between 1<sup>st</sup> Taylor approximation and Monte Carlo simulation are in the order of  $10^{-7}$ . Larger differences can be seen in the volcanic and arid study areas, but still in the order of  $10^{-3}$ . Agreement between methods could be further improved by using a 2<sup>nd</sup> order approximation in future work, but this is out of the context of this study. Uncertainty estimates fall within the bounds of the fixed value analysis in all cases except for the volcanic study area. This could be due to the missing values in the volcanic time series. Fixed value analysis is performed only on existing observations, and this may explain why the expected normalized values are higher in the fixed-value estimates than following the two uncertainty propagation approaches.

A closer examination on the reasons that contribute to uncertainty shows correlation between monthly mean uncertainty and monthly missing values/monthly humidity, as recorded at ground-based meteorological stations covered by the study areas. An example is provided in Figure 6.1 for the case of the volcanic area, where uncertainty estimates were the highest. As seen in Figure 6.1, the correlation coefficient is 0.68 between monthly uncertainty

and missing values and 0.75 between monthly uncertainty and maximum humidity. Missing values lead to an increase in sampling uncertainty, while humidity increases the uncertainty of LST retrievals.

Further study of Table 6.1 shows that the differences between the uncertainty estimates resulting from the 1<sup>st</sup> Taylor approximation and the Monte Carlo simulation are minimal. In the mountainous study area and the area at the edge of the swath, the differences are in the order of  $10^{-7}$ . Larger differences can be seen in the volcanic and arid study areas, but still in the order of  $10^{-3}$ . These differences could be addressed by a 2<sup>nd</sup> order approximation in future work. Given the low levels of calculated uncertainty, it is not considered necessary to further improve agreement between the methods. Uncertainty estimates fall within the bounds of the fixed value analysis in all cases except for the volcanic study area. This could be due to the missing values in the volcanic time series. Fixed value analysis is performed only on existing observations, and this may explain why the expected normalized values are higher in the fixed-value estimates than following the two uncertainty propagation approaches.

A closer examination on the reasons that contribute to uncertainty shows correlation between monthly mean uncertainty and monthly missing values/monthly humidity, as recorded at ground-based meteorological stations covered by the study areas. An example is provided in Figure 6.1 for the case of the volcanic area, where uncertainty estimates were the highest. As seen in Figure 6.1, the correlation coefficient is 0.68 between monthly uncertainty and missing values and 0.75 between monthly uncertainty and maximum humidity. Missing values lead to an increase in sampling uncertainty, while humidity increases the uncertainty of LST retrievals.

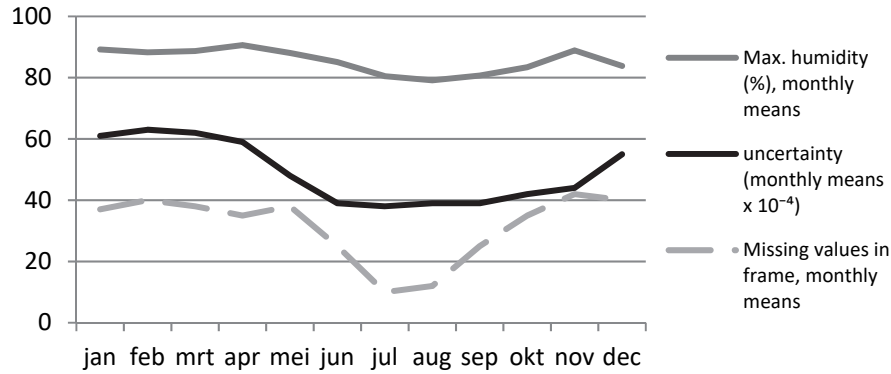


Figure 6.1. Factors influencing uncertainty propagation. The graph shows the relation between monthly uncertainty in the normalized values (in black), the number of missing values in the frame (grey dashed line) and the maximum humidity (grey solid line). Uncertainty values are from the volcanic area (Taylor approximation method). Humidity values are measured at the meteorological station of Catania and retrieved from <https://www.wunderground.com/>

An example of how LST input uncertainty propagates through the normalized series into the results of anomaly detection, is illustrated Figure 6.2. Panel (a) shows original LST series in the study area of Algeria with the uncertainty of the values shaded with blue. The normalized series is shown in panel (b), with uncertainty propagation estimated following the Taylor method (shaded blue). High levels of uncertainty can increase variability in the normalized signal, and this will increase the  $\mu+2\sigma$  detection threshold (Chapters 3-5). The hypothetical thresholds shown in panel (b) demonstrate the potential impact of the uncertainty of normalized values on anomaly detection. Anomalies which would have been detected by an error-free threshold may be missed (\*\*). Other anomalies may be detected because of their high uncertainty levels, but they do not correspond to values which would have been considered anomalous by an error-free threshold (\*). This demonstration is based on hypothetical thresholds, because in our study the uncertainty estimates for normalized values were so low that they practically did not affect the detection thresholds.

Overall, the estimated uncertainties are very low, reaching at most 0.5% of the mean of the normalized signal. The highest estimated uncertainty in the normalized series (.0045 in the volcanic case, following the 1<sup>st</sup> order Taylor approximation) reaches the levels of variability calculated in Chapter 3 for the TIR normalized series from Niger. Thus it would be expected that analysis with the use of LST data, even considering the highest estimated levels of uncertainty propagation, could reach a sensitivity comparable to the +2K that was estimated for synthetic anomaly detection in the case of Niger.

## **6.5 Conclusion**

Uncertainty of LST retrievals propagates from the satellite input to the result of anomaly detection, which is based on the normalized time series. It is relevant to examine the uncertainty of normalized values, in order to see to what extent the anomalies detected in the previous chapter are caused by LST product uncertainty. The results of this study show that uncertainty estimates of normalized values are very low, even though they are calculated for conditions of highest expected variability (areas with high LST retrieval uncertainty, spatially correlated normalization frame and central pixel). This supports confidence that the anomalies detected in the previous chapter are related to physical phenomena rather than artefacts introduced by uncertainty in LST retrievals.

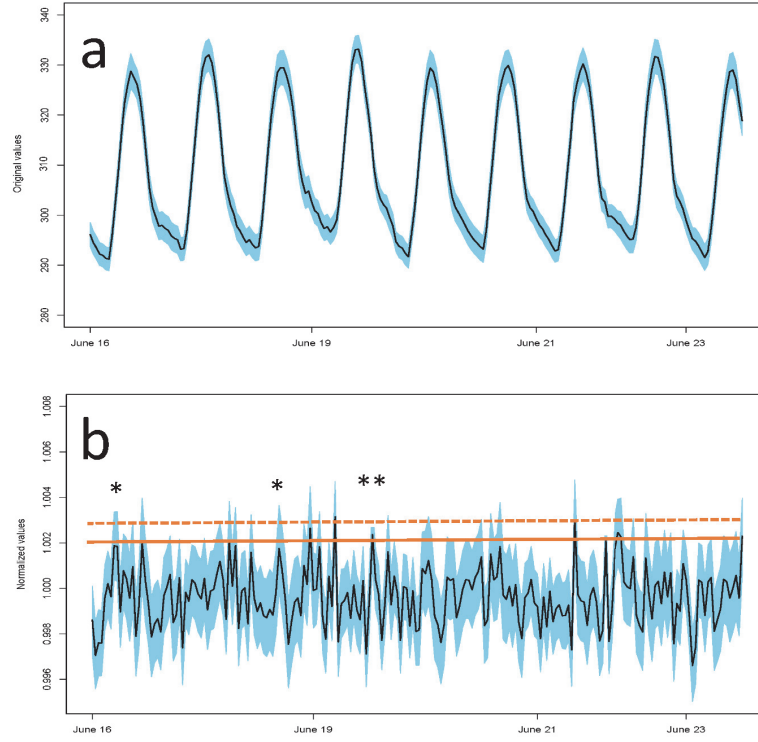


Figure 6.2. Uncertainty propagation. In panel (a), original LST values at the central pixel in the desert (Algeria) and their uncertainty  $u_c$  (in blue). In panel (b), normalized values in the same pixel. Normalization is performed using an 8-pixel-sided frame (24km). The blue areas represent uncertainty estimates based on the Taylor approximation method. The solid horizontal line represents an ideal detection threshold, based on error-free normalized values. The dashed horizontal line is a hypothetical detection threshold based on normalized values with high levels of uncertainty. When uncertainty levels in normalized values are high, detected anomalies may not correspond to values which are anomalous in reality (examples shown as \*) or actual anomalies may be missed (\*\*).





## Chapter 7. Synthesis and outlook

This thesis investigates the relation between the occurrence of earthquakes and the presence of anomalies in satellite-based land surface temperature data. The challenge in this investigation is that there is no proven physical mechanism that describes and quantifies the effect of earthquakes on thermal emissions. The hypotheses of this thesis are therefore formed based on the findings reported in current literature (as discussed in detail in Chapters 1 and 2). The experiments are also designed based on literature and multiple detection settings are tested.

The first hypothesis is that earthquakes have a traceable contribution to satellite-based LST. The second hypothesis is that more anomalies appear before/during the date of the earthquake, at closer distance to the location of the epicentre, and only on the year when the earthquake takes place. The third hypothesis is that more anomalies are expected to appear before shallower earthquakes or earthquakes of higher magnitude. These hypotheses imply that earthquake contributions to the thermal emissions recorded by the satellite sensor are (a) spatiotemporally finite, and (b) spatiotemporally coincide with the area of stress accumulation and the time before/during the earthquake. To detect anomalies with these characteristics, a suitable methodology had to be implemented that was different to the approaches previously followed in related published research.

The following section summarizes the role of this methodological procedure for the earthquake study and for broader use. A reflection follows on the application of the methodology on volcanic monitoring, as a real-life test of the method and as a self-standing application. Then, the implications of the results of the earthquake case study are analysed in relation with the hypotheses underlying the research. Finally, an outlook is provided regarding the potential of using Land Surface temperature fluctuations, as earthquake precursors and for other applications.

### **7.1 Normalization and anomaly detection**

Following the hypotheses of the study, anomaly detection should focus on spatially finite, localized patterns. As discussed in chapter 2, this was not possible using the approaches previously applied in published research on thermal earthquake precursors. Using traditional signal processing techniques, thermal emissions in any pixel of the image would be examined only based on their history. Daily/seasonal patterns which change with time would not be addressed. In contrast, the normalization procedure described in Chapter 3 suppresses commonalities at the time they emerge. Furthermore, the open frame shape of the normalization procedure ensures that only spatially finite

effects are isolated. Patterns that extend further than the normalization frame, like an extended heat wave, are suppressed, and will not be mistakenly considered as earthquake-induced anomalies.

The methodology is tested with synthetic anomalies. This procedure allows for better control and quantification of how a well-defined anomaly is seen throughout the stages of anomaly detection. Anomaly detection is binary (normalized values exceed the threshold or not) and does not give any information as to how much the normalized values exceed the threshold. Synthetic testing is therefore necessary for linking the amplitude of detected peaks with the intensity of anomaly in the original signal. The use of the dynamic  $\mu+2\sigma$  threshold has another implication: a part of the signal is always declared as anomalous. Detected peaks appear due to normal spatiotemporal variations of the signal, caused by local influences (more details in following sections) and highlighted by the normalization. The appearance of small numbers of anomalies throughout the dataset or peaks of similar size throughout the time series, should be interpreted as lack of an outstanding physical event. As shown in the synthetic testing, larger imposed anomalies are clearly discernible as peaks higher than in the rest of the image.

The advantages of normalization are relevant for the earthquake-related study, and also for the detection of subtle anomalies in other applications. The method is suitable for a broader use. It can be applied with input data that have a spatial and a temporal dimension. Tools are provided to the user to assess the effect of missing values on the procedure, when and if this is relevant for the chosen input data (see the Data Availability Thresholds described in Chapter 3). The normalization frame is adjustable to what may be considered as local in the scale of each study. For example, the detection of a burning oil spill that may reach an extent of 1500 km<sup>2</sup> (like the Deepwater Horizon oil spill of 2010) would require a larger normalization frame than the frame necessary for monitoring the thermal emissions over the Jubail Industrial City in Saudi Arabia (spatial extent: 5km<sup>2</sup> at the time of writing). Similarly, detection can be adjusted to the context of different applications: a moving temporal window may be used to visualize anomalies that coincide in a specific time interval, which may be meaningful for a particular study and not for others; and different thresholds can be used depending on the physical characteristics of target anomalies. For example, detection of slow-moving lava flows could be carried out with longer temporal windows than detection of explosive short-lived strombolian eruptions; and with higher thresholds than monitoring of low-temperature thermal fluctuations over lava domes.

## 7.2 Volcanic application

The volcanic case study is an application of the methodology of chapter 3 on a real-life target event using different satellite-based data: MWIR, LWIR and Land Surface Temperatures (LST). On one hand, the volcanic case study demonstrates detection of subtle thermal signatures that can be verified with ground-based data. On the other hand, volcanic activity is in itself interesting as a monitoring target, because it is a potential hazard to exposed populations. The methodology of Chapter 3 allows use of relatively unexploited satellite resources, like LST, to complement mainstream use of MWIR for volcanic target monitoring. This can lead to (a) the construction of longer datasets for study of volcanic dynamics and (b) the study of low-temperature volcanic targets, like lava domes, because LST data are more sensitive to low temperature emissions.

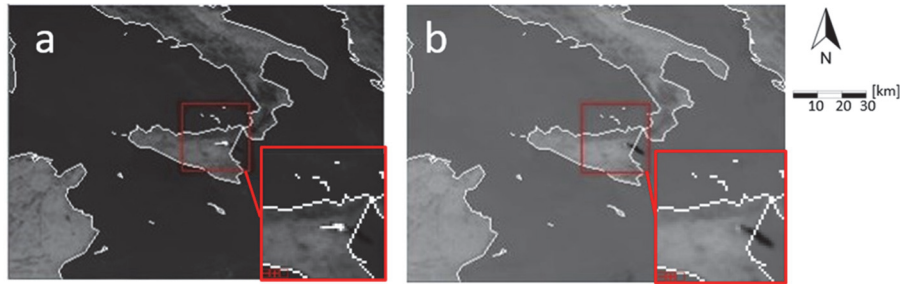


Figure 7.1. Comparison of midwave and longwave IR images. Volcanic eruption with lava flows in Mount Etna, Sicily, Italy. In panel (a), midwave infra-red image of the eruption registered by SEVIRI channel 4 ( $3.9\mu\text{m}$ ). In panel (b), the same scene as registered in the longwave-infrared channel 9 ( $10.8\mu\text{m}$ ) of the same sensor. The insets in each of the panels focus in the lava flow, which is more clearly visible in the midwave infrared (panel a) than in the longwave (panel b).

It can be argued that the intensity of IR emissions during a volcanic eruption is such, that its signature does not qualify as a 'subtle' fluctuation in satellite data. However, this is true only for specific wavelengths. A volcanic eruption which is clearly visible in the midwave infrared band 4 (MWIR,  $3.9\mu\text{m}$ ) of the SEVIRI sensor onboard the geostationary MSG satellite, cannot easily be seen in the longwave infrared band 9 (LWIR,  $10.8\mu\text{m}$ ) (an example from an eruption of Mount Etna, Italy is shown in Figure 7.1). A pixel over an area covered by 2% with a lava surface of 400K, as the ones reported by Ganci et al. [2013], shows an increase of approximately 3.5K in the Brightness Temperatures registered in band 9 (calculations based on the Thermal mixture radiance calculator of the USGS, [https://astrogeology.usgs.gov/tools/thermal-radiance-calculator/thermal\\_mixture\\_tool.html](https://astrogeology.usgs.gov/tools/thermal-radiance-calculator/thermal_mixture_tool.html)). Such increases were detected in LST data in chapter 4. Increases of similar amplitude, regardless of the cause of

the detected anomaly, should be therefore detectable in LST data in the context of other studies, like the earthquake case study.

A final observation relates to the compensation for missing data in the temporal window, which is described in chapter 3. The volcanic case study is an example of an application where missing data may be physically linked to an event and the compensation may aid detection. However, it is important to note that this compensation introduces uncertainty in the detection and should be very cautiously applied in other studies, with the use of the data availability thresholds described in chapter 3. The compensation is in any case applied only in combination with the temporal moving window, which is an optional analysis tool and is not a general requirement for anomaly detection.

## **7.3 Earthquake application**

### **7.3.1 Overview**

In the case of the earthquake case study, the LST dataset used as input is cloud-masked and atmospherically corrected. This reduces artefacts in the processing that would be caused by interferences of atmospheric origin (chapter 3). The data have high temporal and spatial resolution to allow for examination of the spatiotemporal relation between detected anomalies and the location and time of earthquake occurrence. The study includes 20 large, shallow, land-based earthquakes with different focal mechanisms, and from areas of different climatic zones. No *a priori* information is available for the physical characteristics or the timing of an earthquake-induced anomaly, so anomaly detection procedures are applied with the same settings for all pixels in the images. Similarly, all detected anomalies are considered in the analysis and included in the statistical evaluation. The study on LST retrieval uncertainty propagation (chapter 6) shows that very low levels of uncertainty propagate to normalized values, therefore detected anomalies are not likely linked to artefacts caused by input uncertainty.

### **7.3.2 Hypotheses and evaluation of results**

#### ***7.3.2.1 Hypothesis: Earthquakes have a traceable contribution to LST***

The earthquake case study showed that there is no statistically significant relation between observed anomalies and earthquakes. This hypothesis is therefore rejected, as far as this study is concerned. Either there exists no earthquake contribution to surface emissions, registered by satellite sensors, or this contribution exists but could not be traced in this study. The latter implies that the assumption of spatiotemporally finite earthquake effects, coinciding with the earthquake, may be incorrect. Two potential reasons for

this are (a) that the earthquake effect is more spatially extended than 18225 km<sup>2</sup> and/or (b) that the earthquake effect is present not only in the months prior to the earthquake but throughout the year. Potential earthquake effects may be expressed throughout the process of stress accumulation in the crust. This process varies in time and in space [Ye et al., 2018], and it has been shown that an earthquake does not necessarily release all accumulated stress (see, for example, Elliot et al., 2011 for the earthquake of Van, Turkey). Earthquake variability may account for earthquake effects being traceable in some cases and not in others.

Another potential reason why earthquake effects may remain undetected, is that earthquake effects may be much weaker than the rest of the factors influencing LSTs and the recorded TIR data. Such factors include spatiotemporal variations of surface spectral emissivity, which is highly variable over land; local geographical factors (altitude above sea level, solar exposure, geographic latitude) which lead to temporal and spatial variability in the TIR signal; and local atmospheric conditions, like humidity and aerosol content, which interfere with the registration of surface radiance by the sensor. Researchers suggest that this obstacle could be overcome by the study of anomalies in multiparametric time series, i.e. considering ionospheric perturbations and fluctuations of other parameters temporally coinciding with thermal anomalies [e.g. Pulinets, 2011]. However, the study of multiparametric time series has not proven so far the existence of pre-earthquake thermal anomalies, and anomalies in other suggested parameters have also not been verified as precursors [Jordan et al., 2011]. Non-earthquake interferences might be overcome if the physical mechanism of earthquake contributions is proven and quantified. In such a case, future anomaly detection algorithms could include physical constraints to filter only earthquake related effects.

*7.3.2.2 Hypothesis: There is a relation between earthquakes and observed anomalies (i.e. years with earthquakes and periods and locations closer to the earthquake show more anomalies)*

No statistical evidence was found to support this hypothesis. If there is any relation between earthquakes and anomalies, it is overshadowed by other influences and cannot be proven in the results of this thesis. The results of the earthquake case study show that the density of detected anomalies is statistically not related to the occurrence of earthquakes. Anomalies are detected throughout the study areas, throughout the study periods, in years with and without earthquake occurrence. No similar patterns are found in the case of similar earthquakes happening in the same area in different years. The findings persist regardless of the use of different thresholds or definitions of co-seismic period. The hypothesis is therefore rejected.

*7.3.2.3 Hypothesis: more anomalies are expected to appear before shallower earthquakes or earthquakes of higher magnitude*

Anomaly density does not differ between years with and without earthquakes. No earthquake contribution to LST could be traced in this study (see Hypothesis 1). Thus earthquakes, regardless of their magnitude and focal depth, do not affect the number of anomalies detected in this study. This hypothesis is therefore automatically rejected.

**7.3.3 Contribution of the study and potential for alternative application**

The contribution of this study is that it (a) excludes large-scale effects from anomaly detection, (b) provides fine-scale descriptions of the spatial and temporal occurrence of anomalies, relative to the time and location of earthquake occurrence, and (c) statistically evaluates the spatiotemporal coincidence of earthquakes and anomalies.

The choices made in this work consider bibliographic information, for example on the spatial extent of earthquake preparation areas. When no recommendation is available (for example, in the definition of co-seismic period) more options are tested. Application of the approach is not restricted to the settings that are chosen in this thesis and can be repeated in the future using updated criteria. For example, if a physical mechanism suggests that only specific types of earthquakes could have traceable contributions, the statistical analysis could be adjusted to accommodate such classification.

**7.4 Outlook for future studies**

The findings show that, as far as this thesis is concerned, LST anomalies could not be used as a precursor of the twenty earthquakes studied. This result should be confirmed with more earthquake cases in the future, as a larger sample could better support the statistical analysis. Anomalies related to earthquakes were either not detected because they do not exist, or because they temporally extend throughout the years, or because they are spatially more extensive than tested, or because they are overshadowed by other influences. In any of these cases their utility as precursor is limited. To conclude on the existence of earthquake-induced thermal anomalies, which could not be detected in this study, the underlying physical background should be further studied and the suggested relations should be quantified. This would support the design of an anomaly detection methodology that excludes non-earthquake related environmental influences on the recorded TIR emissions.

The research carried out shows potential for studying spatiotemporal LST variability in applications not related to earthquakes. For example, the analysis of data in more than twenty locations worldwide showed that LST time series

over urban areas have high variability (see examples in the results section of chapter 5). Further time series analysis could thus be useful for urban heat island applications and epidemiological studies. Wetlands and coastal areas also show high variability. Application of the normalization described in chapter 3 could highlight subtle, water-related fluctuations with potential utility in drought, agricultural and environmental monitoring. Finally, the methodological procedure itself could be improved in the future. For example, normalized values could be summed to produce an anomaly index that represents the intensity of an anomaly, rather than the binary presence-absence information resulting from the application of a threshold. This option was briefly examined with good results in this work, but was not further explored due to time restrictions.





## Summary

This thesis describes the application of time series analysis to satellite-derived thermal infrared imagery, in order to determine if spatiotemporally limited anomalies can be linked to earthquake occurrence as has been suggested by literature. For this purpose, a methodology is applied that can suppress large-scale patterns and highlight subtle localized fluctuations in hypertemporal satellite data. The methodology is tested on at-sensor longwave and midwave Thermal Infra-red (TIR) Brightness Temperature data and on TIR-derived Land Surface Temperature (LST) data. Experiments are first carried out to retrieve synthetic anomalies which are imposed in real-life datasets. The methodology is then applied for detection of volcanic activity, and the results are validated using published ground-based reports. The application of this methodology is found to facilitate utilization of longwave IR input for long-term volcanic monitoring, to complement existing hotspot detection techniques and to aid monitoring of lower temperature targets even in areas of constant activity. Finally, the methodology is applied to examine the presence of detectable localized (spatial extent up to 18225 km<sup>2</sup>) increases in LST prior to twenty large, shallow, land-based earthquakes worldwide. The findings show that there is no statistically significant difference between the anomaly density detected at different distances from the earthquake, at different periods (before, during and after the earthquake), and in years with and without earthquake occurrence. It is not clear if the earthquakes have no influence on the LST registered by the satellite; if the influence exists but is overshadowed by local environmental and atmospheric influences; or if influences are linked to stress accumulation rather than stress release and therefore do not coincide with the earthquake. To clarify this, more research is required on the physical background behind the links between earthquakes and TIR anomalies which are suggested by literature.

## *Summary*

---

## Samenvatting

Dit proefschrift betreft de toepassing van tijdreeksanalyse op thermische satellietbeelden, om te bepalen of spatiotemporeel beperkte anomalieën kunnen worden gekoppeld aan aardbevingen, zoals is gesuggereerd in de literatuur. Hiervoor is een methode toegepast, die grootschalige patronen in hypertemporele satellietgegevens onderdrukt en subtiele gelokaliseerde fluctuaties benadrukt. Deze methode wordt getest op at-sensor midde- en lange- golflengte thermisch infrarood (TIR) Brightness Temperature en op TIR-afgeleide oppervlaktetemperatuur (LST) gegevens. Experimenten zijn eerst uitgevoerd voor de detectie van synthetische afwijkingen die worden opgelegd in real-life datasets. De methode is vervolgens toegepast voor de detectie van vulkanische activiteit en de resultaten zijn gevalideerd met behulp van gepubliceerde rapporten. Het blijkt dat de toepassing van deze methode het gebruik van lange-golflengte-IR-input voor langdurige vulkanische monitoring ondersteunt, zoals het waarnemen van vulkanische gebieden met lagere temperaturen of met constante activiteit. Ten slotte wordt de methode toegepast om de aanwezigheid van detecteerbare gelokaliseerde (ruimtelijke omvang tot 18225 km<sup>2</sup>) verhogingen in het LST te onderzoeken voorafgaand aan twintig grote, oppervlakkige aardbevingen op het vasteland. De bevindingen tonen aan dat er geen statistisch significant verschil is tussen de waargenomen anomaliedichtheid op verschillende afstanden van de aardbeving, in verschillende perioden (voor, tijdens en na de aardbeving) en in de jaren met en zonder aardbevingen. Het is niet duidelijk of de aardbevingen geen invloed hebben op de door de satelliet geregistreerde LST; als die invloed bestaat, maar wordt overschaduwd door lokale omgevings- en atmosferische invloeden; of als die invloed verband houdt eerder met stress-accumulatie dan met stress-vrijzetting en daarom valt deze niet samen met de aardbeving. Om dit te verduidelijken, is er meer onderzoek nodig over de fysieke achtergrond achter de door de literatuur gesuggereerde verbanden tussen aardbevingen en TIR-afwijkingen.



## References

- AKHOONDZADEH, M. 2013a. An Adaptive Network-based Fuzzy Inference System for the detection of thermal and TEC anomalies around the time of the Varzeghan, Iran, (Mw = 6.4) earthquake of 11 August 2012. *Advances in Space Research*, 52, 837-852.<http://dx.doi.org/10.1016/j.asr.2013.05.024>
- AKHOONDZADEH, M. 2013b. A comparison of classical and intelligent methods to detect potential thermal anomalies before the 11 August 2012 Varzeghan, Iran, earthquake (Mw= 6.4). *Natural Hazards and Earth System Sciences*, 13, 1077
- ALEGANA, V. A., ATKINSON, P. M., WRIGHT, J. A., KAMWI, R., UUSIKU, P., KATOKELE, S., SNOW, R. W. & NOOR, A. M. 2013. Estimation of malaria incidence in northern Namibia in 2009 using Bayesian conditional-autoregressive spatial-temporal models. *Spatial and Spatio-temporal Epidemiology*, 7, 25-36.<https://doi.org/10.1016/j.sste.2013.09.001>
- 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, 303-317
- ALVAN, H. V., MANSOR, S., OMAR, H. & AZAD, F. H. 2014. Precursory signals associated with the 2010 M8.8 Bio-Bio earthquake (Chile) and the 2010 M7.2 Baja California earthquake (Mexico). *Arabian Journal of Geosciences*, 7, 4889-4897.[10.1007/s12517-013-1117-9](https://doi.org/10.1007/s12517-013-1117-9)
- BARAL, S., SHARMA, K., SARAF, A., DAS, J., SINGH, G., BORGOHAIN, S. & KAR, E. 2016. Thermal anomaly from NOAA data for the Nepal earthquake. *CURRENT SCIENCE*, 110, 150
- BHARDWAJ, A., SINGH, S., SAM, L., BHARDWAJ, A., MARTÍN-TORRES, F. J., SINGH, A. & KUMAR, R. 2017a. MODIS-based estimates of strong snow surface temperature anomaly related to high altitude earthquakes of 2015. *Remote Sensing of Environment*, 188, 1-8.<http://dx.doi.org/10.1016/j.rse.2016.11.005>
- BHARDWAJ, A., SINGH, S., SAM, L., JOSHI, P. K., BHARDWAJ, A., MARTÍN-TORRES, F. J. & KUMAR, R. 2017b. A review on remotely sensed land surface temperature anomaly as an earthquake precursor. *International Journal of Applied Earth Observation and Geoinformation*, 63, 158-166.<https://doi.org/10.1016/j.jag.2017.08.002>
- BIAU, D. J., JOLLES, B. M. & PORCHER, R. 2010. P Value and the Theory of Hypothesis Testing: An Explanation for New Researchers. *Clinical Orthopaedics and Related Research*, 468, 885-892.[10.1007/s11999-009-1164-4](https://doi.org/10.1007/s11999-009-1164-4)
- BLACKETT, M., 2013. Review of the utility of infrared remote sensing for detecting and monitoring volcanic activity with the case study of shortwave infrared data for Lascar Volcano from 2001–2005. *Geological*

## References

---

- Society, London, Special Publications, 380, 107–135.  
<https://doi.org/10.1144/sp380.10>
- BLACKETT, M., 2017. An overview of infrared remote sensing of volcanic activity. *Journal of Imaging*, 3, 13.
- BLACKETT, M., 2014. Early Analysis of Landsat-8 Thermal Infrared Sensor Imagery of Volcanic Activity. *Remote Sensing*, 6, 2282
- BLACKETT, M. 2015. An initial comparison of the thermal anomaly detection products of MODIS and VIIRS in their observation of Indonesian volcanic activity. *Remote Sensing of Environment*, 171, 75–82.<https://doi.org/10.1016/j.rse.2014.5.002>
- BLACKETT, M., WOOSTER, M. J. & MALAMUD, B. D. 2011. Exploring land surface temperature earthquake precursors: A focus on the Gujarat (India) earthquake of 2001. *Geophysical Research Letters*, 38
- BURGI, P. Y., DARRAH, T. H., TEDESCO, D., & EYMOLD, W. K. (2014). Dynamics of the Mount Nyiragongo lava lake. *Journal of Geophysical Research: Solid Earth*, 119, 4106–4122.  
<https://doi.org/10.1002/2013jb010895>
- BYUN, Y. G., HUH, Y., YU, K. & KIM, Y. I. 2007. Evaluation of graph-based analysis for forest fire detections. *Int. J. Comput. Electr. Autom. Control Inf. Eng.*, 1
- CALLE, A., GONZÁLEZ-ALONSO, F. & MERINO DE MIGUEL, S. 2008. Validation of active forest fires detected by MSG-SEVIRI by means of MODIS hot spots and AWIFS images. *International Journal of Remote Sensing*, 29, 3407–3414.<https://doi.org/10.1080/01431160701596164>
- CANTY, M. 2010. *Image Analysis, Classification, and Change Detection in Remote Sensing: With Algorithms for ENVI/IDL*, Boca Raton, Taylor and Francis.
- CANTY, M. J. & NIELSEN, A. A. 2008. Automatic radiometric normalization of multitemporal satellite imagery with the iteratively re-weighted MAD transformation. *Remote Sensing of Environment*, 112, 1025–1036.<https://doi.org/10.1016/j.rse.2007.07.013>
- CANTY, M. J. & NIELSEN, A. A. 2012. Linear and kernel methods for multivariate change detection. *Computers & Geosciences*, 38, 107–114.<https://doi.org/10.1016/j.cageo.2011.05.012>
- CARR, B. B., CLARKE, A. B., & VANDERKLUYSEN, L., 2016. The 2006 lava dome eruption of Merapi Volcano (Indonesia): Detailed analysis using MODIS TIR. *Journal of Volcanology and Geothermal Research*, 311, 60–71.  
<https://doi.org/10.1016/j.jvolgeores.2015.12.004>
- CARTER, A. J., GIRINA, O., RAMSEY, M. S., & DEMYANCHUK, Y. V. ,2008. ASTER and field observations of the 24 December 2006 eruption of Bezymianny Volcano, Russia. *Remote Sensing of Environment*, 112, 2569–2577. <https://doi.org/10.1016/j.rse.2007.12.001>
- CERVONE, G., MAEKAWA, S., SINGH, R., HAYAKAWA, M., KAFATOS, M. & SHVETS, A. 2006. Surface latent heat flux and nighttime LF anomalies

- prior to the  $M_w = 8.3$  Tokachi-Oki earthquake. *Natural Hazards and Earth System Science*, 6, 109-114
- CERVONE, G., SINGH, R., KAFATOS, M. & YU, C. 2005. Wavelet maxima curves of surface latent heat flux anomalies associated with Indian earthquakes. *Natural Hazards and Earth System Sciences*, 5, 87-99
- CHANDOLA, V., BANERJEE, A. & KUMAR, V. 2012. Anomaly detection for discrete sequences: a survey. *IEEE Trans. Knowl. Data Eng.*, 24, pp. 823-839.10.1109/TKDE.2010.235
- CHIARALUCE, L., DI STEFANO, R., TINTI, E., SCOGNAMIGLIO, L., MICHELE, M., CASAROTTI, E., CATTANEO, M., DE GORI, P., CHIARABBA, C., MONACHESI, G., LOMBARDI, A., VALOROSO, L., LATORRE, D. & MARZORATI, S. 2017. The 2016 Central Italy Seismic Sequence: A First Look at the Mainshocks, Aftershocks, and Source Models. *Seismological Research Letters*.10.1785/0220160221
- CHOUDHURY, S., DASGUPTA, S., SARAF, A. K. & PANDA, S. 2006. Remote sensing observations of pre-earthquake thermal anomalies in Iran. *International Journal of Remote Sensing*, 27, 4381-4396.10.1080/01431160600851827
- CLERBAUX, N. 2006. *Image Processing: Data Dictionary*. . Royal Meteorological Institute of Belgium.
- CLEVELAND, R., CLEVELAND, W., MCRAE, J. & TERPENNING, I. 1990. STL: a seasonal-trend decomposition procedure based on LOESS. *J. Off. Stat.*, 6, 3-73
- COOLBAUGH, M. F., KRATT, C., FALLACARO, A., CALVIN, W. M. & TARANIK, J. V. 2007. Detection of geothermal anomalies using Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER) thermal infrared images at Bradys Hot Springs, Nevada, USA. *Remote Sensing of Environment*, 106, 350-359.<https://doi.org/10.1016/j.rse.2006.09.001>
- COPPIN, P., JONCKHEERE, I., NACKAERTS, K., MUYS, B. & LAMBIN, E. 2004. Review Article Digital change detection methods in ecosystem monitoring: a review. *International Journal of Remote Sensing*, 25, 1565-1596.10.1080/0143116031000101675
- COPPOLA, D., LAIOLO, M., CIGOLINI, C., DELLE DONNE, D., & RIPEPE, M. (2016). Enhanced volcanic hot-spot detection using MODIS IR data: Results from the MIROVA system. In: A. J. L. Harris, et al. (Eds.), *Detecting, modelling and responding to effusive eruptions*. London: Geological Society, London, Special Publications.
- DANESHVAR, M. R. M., TAVOUSHI, T. & KHOSRAVI, M. 2014. Synoptic detection of the short-term atmospheric precursors prior to a major earthquake in the Middle East, North Saravan M 7.8 earthquake, SE Iran. *Air Quality, Atmosphere & Health*, 7, 29-39
- DE JONG, S. & VAN DER MEER, F. E. 2004. *Remote Sensing Image Analysis: Including the Spatial Domain*, Springer, The Netherlands.

- DEE, D. P., BALMASEDA, M., BALSAMO, G., ENGELEN, R., SIMMONS, A. J. & THÉPAUT, J.-N. 2014. Toward a Consistent Reanalysis of the Climate System. *Bulletin of the American Meteorological Society*, 95, 1235-1248. [10.1175/bams-d-13-00043.1](https://doi.org/10.1175/bams-d-13-00043.1)
- DEE, D. P., UPPALA, S. M., SIMMONS, A. J., BERRISFORD, P., POLI, P., KOBAYASHI, S., ANDRAE, U., BALMASEDA, M. A., BALSAMO, G., BAUER, P., BECHTOLD, P., BELJAARS, A. C. M., VAN DE BERG, L., BIDLOT, J., BORMANN, N., DELSOL, C., DRAGANI, R., FUENTES, M., GEER, A. J., HAIMBERGER, L., HEALY, S. B., HERSBACH, H., HÓLM, E. V., ISAKSEN, I., KÄLLBERG, P., KÖHLER, M., MATRICARDI, M., MCNALLY, A. P., MONGE-SANZ, B. M., MORCRETTE, J. J., PARK, B. K., PEUBEY, C., DE ROSNAY, P., TAVOLATO, C., THÉPAUT, J. N. & VITART, F. 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137, 553-597. [10.1002/qj.828](https://doi.org/10.1002/qj.828)
- DERRIEN, M. & LE GLÉAU, H. 2005. MSG/SEVIRI cloud mask and type from SAFNWC. *International Journal of Remote Sensing*, 26, 4707-4732. [10.1080/01431160500166128](https://doi.org/10.1080/01431160500166128)
- DEY, S. & SINGH, R. 2003. Surface latent heat flux as an earthquake precursor. *Natural Hazards and Earth System Sciences*, 3, 749-755
- DEY, S. & SINGH, R. 2003. Surface latent heat flux as an earthquake precursor. *Natural Hazards and Earth System Sciences*, 3, 749-755
- DI BELLO, G., FILIZZOLA, C., LACAVA, T., MARCHESE, F., PERGOLA, N., PIETRAPERTOSA, C., TRAMUTOLI, V. (2004). Robust satellite techniques for volcanic and seismic hazards monitoring.
- DOBROVOLSKY, I. P., GERSHENZON, N. I. & GOKHBERG, M. B. 1989. Theory of electrokinetic effects occurring at the final stage in the preparation of a tectonic earthquake. *Phys. Earth Planet. Int.*, 57, 144-156
- DOĞAN, B. & KARAKAŞ, A. 2013. Geometry of co-seismic surface ruptures and tectonic meaning of the 23 October 2011 Mw 7.1 Van earthquake (East Anatolian Region, Turkey). *Journal of Structural Geology*, 46, 99-114. <https://doi.org/10.1016/j.jsg.2012.10.001>
- DONEGAN, S. J., & FLYNN, L. P. (2004). Comparison of the response of the Landsat 7 Enhanced Thematic Mapper Plus and the Earth Observing-1 Advanced Land Imager over active lava flows. *Journal of Volcanology and Geothermal Research*, 135, 105-126. <https://doi.org/10.1016/j.jvolgeores.2003.12.010>
- DUGUAY-TETZLAFF, A., BENTO, V., GÖTTSCHE, F., STÖCKLI, R., MARTINS, J., TRIGO, I., KUNZ, H. (2015). METEOSAT land surface temperature climate data record: Achievable accuracy and potential uncertainties. *Remote Sensing*, 7, 13139.
- DUNAJECKA, M. & PULINETS, S. 2005. Atmospheric and thermal anomalies observed around the time of strong earthquakes in México. *Atmósfera*, 18, 235-247



- EASTERLING, D. R., MEEHL, G. A., PARMESAN, C., CHANGNON, S. A., KARL, T. R. & MEARN, L. O. 2000. Climate Extremes: Observations, Modeling, and Impacts. *Science*, 289, 2068-2074. [10.1126/science.289.5487.2068](https://doi.org/10.1126/science.289.5487.2068)
- ELEFThERIOU, A., FILIZZOLA, C., GENZANO, N., LACAVA, T., LISI, M., PACIELLO, R., PERGOLA, N., VALLIANATOS, F. & TRAMUTOLI, V. 2016. Long-term RST analysis of anomalous TIR sequences in relation with earthquakes occurred in Greece in the period 2004–2013. *Pure and Applied Geophysics*, 173, 285-303
- ELLINGSON, R. G., YANUK, D. J., LEE, H. T. & GRUBER, A. 1989. A Technique for Estimating Outgoing Longwave Radiation from HIRS Radiance Observations *JOURNAL OF ATMOSPHERIC AND OCEANIC TECHNOLOGY* 6
- ELLIOT, J. R., COPLEY, A. C., HOLLEY, R., SCHARER, K. & PARSONS, B. 2011. The 2011 Mw 7.1 Van (Eastern Turkey) earthquake. *Journal of Geophysical Research: Solid Earth*, 118. [10.1002/jgrb.50117](https://doi.org/10.1002/jgrb.50117)
- ENEVA, M., ADAMS, D., WECHSLER, N., BEN-ZION, Y. & DOR, O. 2008. Thermal properties of faults in Southern California from remote sensing data. San Diego, California, USA: Imagair Inc. contracted by NASA Goddard Space Flight Center, SAIC n.NNH05CC13C
- FLASSE, S. P., & CECCATO, P. (1996). A contextual algorithm for AVHRR FIRE detection. *International Journal of Remote Sensing*, 17, 419–424.
- FREITAS, S. C., TRIGO, I. F., BIOUCAS-DIAS, J. M. & GOTTSCHKE, F. M. 2010. Quantifying the Uncertainty of Land Surface Temperature Retrievals From SEVIRI/Meteosat. *IEEE Transactions on Geoscience and Remote Sensing*, 48, 523-534. [10.1109/TGRS.2009.2027697](https://doi.org/10.1109/TGRS.2009.2027697)
- FREITAS, S. C., TRIGO, I. F., MACEDO, J., BARROSO, C., SILVA, R. & PERDIGÃO, R. 2013. Land surface temperature from multiple geostationary satellites. *International Journal of Remote Sensing*, 34, 3051-3068. [10.1080/01431161.2012.716925](https://doi.org/10.1080/01431161.2012.716925)
- FREUND, F. 2003. On the electrical conductivity structure of the stable continental crust. *Journal of Geodynamics*, 35, 353-388. [https://doi.org/10.1016/S0264-3707\(02\)00154-0](https://doi.org/10.1016/S0264-3707(02)00154-0)
- FREUND, F. 2011. Pre-earthquake signals: Underlying physical processes. *Journal of Asian Earth Sciences*, 41, 383-400. <http://dx.doi.org/10.1016/j.jseaes.2010.03.009>
- FREUND, F. T. 2003b. Rocks that crackle and sparkle and glow: strange pre-earthquake phenomena. *Journal of Scientific Exploration*, 17, 37-71
- GANCI, G., JAMES, M. R., CALVARI, S., & NEGRO, C. D., 2013. Separating the thermal fingerprints of lava flows and simultaneous lava fountaining using ground-based thermal camera and SEVIRI measurements. *Geophysical Research Letters*, 40, 5058–5063. <https://doi.org/10.1002/grl.50983>
- GANCI, G., VICARI, A., FORTUNA, L. & DEL NEGRO, C. 2011. The HOTSAT volcano monitoring system based on combined use of SEVIRI and MODIS multispectral data. 2011, 54. [10.4401/ag-5338](https://doi.org/10.4401/ag-5338)

- GENZANO, N., ALIANO, C., FILIZZOLA, C., PERGOLA, N. & TRAMUTOLI, V. 2007. A robust satellite technique for monitoring seismically active areas: The case of Bhuj-Gujarat earthquake. *Tectonophysics*, 431, 197-210. <http://dx.doi.org/10.1016/j.tecto.2006.04.024>
- GHIL, M., ALLEN, M., DETTINGER, M., IDE, K., KONDRASHOV, D., MANN, M., ROBERTSON, A., SAUNDERS, A., TIAN, Y., VARADI, F. & YIOU, P. 2001. Advanced spectral methods for climatic time series. *Rev. Geophys.*, 11, 327-338
- GIGLIO, L., DESCLOITRES, J., JUSTICE, C. O. & KAUFMAN, Y. J. 2003. An Enhanced Contextual Fire Detection Algorithm for MODIS. *Remote Sensing of Environment*, 87, 273-282. [https://doi.org/10.1016/S0034-4257\(03\)00184-6](https://doi.org/10.1016/S0034-4257(03)00184-6)
- GIGLIO, L., KENDALL, J. D. & JUSTICE, C. O. 1999. Evaluation of global fire detection algorithms using simulated AVHRR infrared data. *International Journal of Remote Sensing*, 20, 1947-1984. [10.1080/014311699212290](https://doi.org/10.1080/014311699212290)
- GLOBAL VOLCANISM PROGRAM, 2011a. Report on Etna (Italy). Weekly volcanic activity report 3–9 August. In: Sennert, S.K. *Bulletin of the Global Volcanism Network*. Smithsonian Institution and USGS.
- GLOBAL VOLCANISM PROGRAM, 2011b. Report on Etna (Italy). Weekly volcanic activity report 10–16 August. In: Sennert, S.K. *Bulletin of the Global Volcanism Network*. Smithsonian Institution and USGS.
- GLOBAL VOLCANISM PROGRAM, 2011c. Report on Etna (Italy). Weekly volcanic activity report 17–23 August. In: Sennert, S.K. *Bulletin of the Global Volcanism Network*. Smithsonian Institution and USGS.
- GLOBAL VOLCANISM PROGRAM, 2011d. Report on Etna (Italy). Weekly volcanic activity report 24–30 August. In: Sennert, S.K. *Bulletin of the Global Volcanism Network*. Smithsonian Institution and USGS.
- GLOBAL VOLCANISM PROGRAM, 2013. Report on Etna (Italy). In: R. Wunderman (Ed.), *Bulletin of the Global Volcanism Network* 38:9. Smithsonian Institution. <https://doi.org/10.5479/si.GVP.BGVN201309-211060>
- GLOBAL VOLCANISM PROGRAM, 2014b. Report on Nyiragongo (DR Congo). In: R. Wunderman (Ed.), *Bulletin of the Global Volcanism Network* 39:4. Smithsonian Institution. <https://doi.org/10.5479/si.GVP.BGVN201404-223030>
- GLOBAL VOLCANISM PROGRAM, 2014a. Report on Nyamuragira (DR Congo). In: R. Wunderman (Ed.), *Bulletin of the Global Volcanism Network* 39:3. Smithsonian Institution. <https://doi.org/10.5479/si.GVP.BGVN201403-223020>
- GORNY, V., SALMAN, A., TRONIN, A. & SHILIN, B. TERRESTRIAL OUTGOING INFRARED RADIATION AS AN INDICATOR OF SEISMIC ACTIVITY. *Proceedings of the Academy of Sciences of the USSR*, 1988. 67-69.
- GOUHIER, M., GUÉHENNEUX, Y., LABAZUY, P., CACAULT, P., DECRIEM, J., & RIVET, S., 2016. HOTVOLC: A web-based monitoring system for volcanic

- hot spots. Geological Society, London, Special Publications, 426, 223–241.  
<https://doi.org/10.1144/sp426.31>
- GOVONI, A., MARCHETTI, A., DE GORI, P., DI BONA, M., LUCENTE, F. P., IMPROTA, L., CHIARABBA, C., NARDI, A., MARGHERITI, L., AGOSTINETTI, N. P., DI GIOVAMBATTISTA, R., LATORRE, D., ANSELMINI, M., CIACCIO, M. G., MORETTI, M., CASTELLANO, C. & PICCININI, D. 2014. The 2012 Emilia seismic sequence (Northern Italy): Imaging the thrust fault system by accurate aftershock location. *Tectonophysics*, 622, 44–55. <https://doi.org/10.1016/j.tecto.2014.02.013>
- GRIESER, J., TRÖMEL, S. & SCHÖNWIESE, C. D. 2002. Statistical time series decomposition into significant components and application to European temperature. *Theoretical and Applied Climatology*, 71, 171–183. <https://doi.org/10.1007/s007040200003>
- GUANGMENG, G. 2008. Studying thermal anomaly before earthquake with NCEP data. *Int Arch Photogram Rem Sens Spatial Inform Sci*, 35, 295–298
- GUTIÉRREZ, F. J., LEMUS, M., PARADA, M. A., BENAVENTE, O. M. & AGUILERA, F. A. 2012. Contribution of ground surface altitude difference to thermal anomaly detection using satellite images: Application to volcanic/geothermal complexes in the Andes of Central Chile. *Journal of Volcanology and Geothermal Research*, 237–238, 69–80. <https://doi.org/10.1016/j.jvolgeores.2012.05.016>
- HAFEN, R. 2010. Local Regression Models: Advancements, Applications, and New Methods. PhD thesis.
- HAMZA, V. M. 2001. Tectonic leakage of fault bounded aquifers subject to non-isothermal recharge: a mechanism generating thermal precursors to seismic events. *Physics of the Earth and Planetary Interiors*, 126, 163–177. [https://doi.org/10.1016/S0031-9201\(01\)00253-9](https://doi.org/10.1016/S0031-9201(01)00253-9)
- HARRIS, A. J. L., GROEVE, T. D., GAREL, F., & CARN, S. A. E. (2016). Detecting, modelling and responding to effusive eruptions. London: Geological Society of London Special Publications.
- HARRIS, A. J. L., ROTHERY, D. A., CARLTON, R. W., LANGAAS, S., & MANNSTEIN, H., 1995. Non-zero saturation of AVHRR thermal channels over high temperature targets: Evidence from volcano data and a possible explanation. *International Journal of Remote Sensing*, 16, 189–196. <https://doi.org/10.1080/01431169508954388>
- HAUKSSON, E., STOCK, J., HUTTON, K., YANG, W., VIDAL-VILLEGAS, J. A. & KANAMORI, H. 2011. The 2010 M w 7.2 El Mayor-Cucapah Earthquake Sequence, Baja California, Mexico and Southernmost California, USA: Active Seismotectonics along the Mexican Pacific Margin. *Pure and Applied Geophysics*, 168, 1255–1277. <https://doi.org/10.1007/s00024-010-0209-7>
- HAYES, G. P., MEYERS, E. K., DEWEY, J. W., BRIGGS, R. W., EARLE, P. S., BENZ, H. M., SMOCZYK, G. M., FLAMME, H. E., BARNHART, W. D., GOLD, R. D. & FURLONG, K. P. 2017. Tectonic summaries of magnitude 7 and

- greater earthquakes from 2000 to 2015. Open-File Report. Reston, VA.10.3133/ofr20161192
- HIGGINS, J., & HARRIS, A. 1997. VAST: A program to locate and analyse volcanic thermal anomalies automatically from remotely sensed data. *Computers & Geosciences*, 23, 627–645. [https://doi.org/10.1016/S0098-3004\(97\)00039-3](https://doi.org/10.1016/S0098-3004(97)00039-3)
- HUMLUM, O., SOLHEIM, J.-E. & STORDAHL, K. 2011. Spectral Analysis of the Svalbard Temperature Record 1912–2010. *Advances in Meteorology*, 2011, 14.10.1155/2011/175296
- ICHOKU, C., KAUFMAN, Y. J., GIGLIO, L., LI, Z., FRASER, R. H., JIN, J. Z. & PARK, W. M. 2003. Comparative analysis of daytime fire detection algorithms using AVHRR data for the 1995 fire season in Canada: Perspective for MODIS. *International Journal of Remote Sensing*, 24, 1669–1690.10.1080/01431160210144697
- JIE, Y. & GUANGMENG, G. 2013. Preliminary analysis of thermal anomalies before the 2010 Baja California M7.2 earthquake. *Atmósfera*, 26, 473–477.[http://dx.doi.org/10.1016/S0187-6236\(13\)71089-0](http://dx.doi.org/10.1016/S0187-6236(13)71089-0)
- JIE, Y. & GUANGMENG, G. 2014. Brief Communication: The interaction of clouds with surface latent heat flux variation before the 2011 M6.1 Russia earthquake. *Nat. Hazards Earth Syst. Sci.*, 14, 2649–2653.10.5194/nhess-14-2649-2014
- JIMÉNEZ-MUÑOZ, J. C., MATTAR, C., SOBRINO, J. A. & MALHI, Y. 2015. A database for the monitoring of thermal anomalies over the Amazon forest and adjacent intertropical oceans. 2, 150024.10.1038/sdata.2015.24
- JIMÉNEZ-MUÑOZ, J. C., SOBRINO, J. A., MATTAR, C. & MALHI, Y. 2013. Spatial and temporal patterns of the recent warming of the Amazon forest. *Journal of Geophysical Research: Atmospheres*, 118, 5204–5214.5.1002/jgrd.50456
- JING, F., SHEN, X. H., KANG, C. L. & XIONG, P. 2013. Variations of multi-parameter observations in atmosphere related to earthquake. *Nat. Hazards Earth Syst. Sci.*, 13, 27–33. DOI: 10.5194/nhess-13-27-2013
- JONSSON, P. & EKLUNDH, L. 2002. Seasonality extraction by function fitting to time-series of satellite sensor data. *IEEE Transactions on Geoscience and Remote Sensing*, 40.10.1109/TGRS.2002.802519
- KAWAMURA, H., HATANO, T., KATO, N., BISWAS, S., CHAKRABARTI, B.K. 2012. Statistical physics of fracture, friction, and earthquakes. *Reviews of Modern Physics* 84, 839. <https://doi.org/10.1103/RevModPhys.84.839>
- KERVYN, M., ERNST, G. G. J., HARRIS, A. J. L., BELTON, F., MBEDE, E., & JACOBS, P. (2008). Thermal remote sensing of the low-intensity carbonatite volcanism of Oldoinyo Lengai, Tanzania. *International Journal of Remote Sensing*, 29, 6467–6499. <https://doi.org/10.1080/01431160802167105>

- KOEPPEN, W. C., PILGER, E. & WRIGHT, R. 2011. Time series analysis of infrared satellite data for detecting thermal anomalies: a hybrid approach. *Bulletin of Volcanology*, 73, 577-593.10.1007/s00445-010-0427-y
- KUENZER, C., HECKER, C., ZHANG, J., WESSLING, S. & WAGNER, W. 2008. The potential of multidiurnal MODIS thermal band data for coal fire detection. *International Journal of Remote Sensing*, 29, 923-944.10.1080/01431160701352147
- KUENZER, C., ZHANG, J., LI, J., VOIGT, S., MEHL, H. & WAGNER, W. 2007. Detecting unknown coal fires: synergy of automated coal fire risk area delineation and improved thermal anomaly extraction. *International Journal of Remote Sensing*, 28, 4561-4584.5.1080/01431160701250432
- KUMAR, P. & FOUFOULA-GEORGIOU, E. 1997. Wavelet analysis for geophysical applications. *Reviews of Geophysics*, 35, 385-412.10.1029/97rg00427
- LABAT, D. 2005. Recent advances in wavelet analyses: Part 1. A review of concepts. *Journal of Hydrology*, 314, 275-288.<https://doi.org/10.1016/j.jhydrol.2005.04.003>
- LEE, H. T. 2014. Climate Algorithm Theoretical Basis Document (C-ATBD): Outgoing Longwave Radiation (OLR)-Daily. NOAA's Climate Data Record (CDR) Program
- LIPEROVSKY, V. A., POKHOTILOV, O. A., MEISTER, C. V. & LIPEROVSKAYA, E. V. 2008. Physical models of coupling in the lithosphere-atmosphere-ionosphere system before earthquakes. *Geomagnetism and Aeronomy*, 48, 795-806. DOI: 10.1134/S0016793208060133
- LISI, M., FILIZZOLA, C., GENZANO, N., GRIMALDI, C., LACAVA, T., MARCHESE, F., MAZZEO, G., PERGOLA, N. & TRAMUTOLI, V. 2010. A study on the Abruzzo 6 April 2009 earthquake by applying the RST approach to 15 years of AVHRR TIR observations. *Natural Hazards and Earth System Sciences*, 10, 395-406
- LISI, M., FILIZZOLA, C., GENZANO, N., PACIELLO, R., PERGOLA, N. & TRAMUTOLI, V. 2015. Reducing atmospheric noise in RST analysis of TIR satellite radiances for earthquakes prone areas satellite monitoring. *Physics and Chemistry of the Earth, Parts A/B/C*, 85-86, 87-97.<http://dx.doi.org/10.1016/j.pce.2015.07.013>
- LIU, D.-F., PENG, K.-Y., LIU, W.-H., LI, L.-Y. & HOU, J.-S. 1999. Thermal omens before earthquakes. *Acta Seismologica Sinica*, 12, 710-715
- LU, X., MENG, Q. Y., GU, X. F., ZHANG, X. D., XIE, T. & GENG, F. 2016. Thermal infrared anomalies associated with multi-year earthquakes in the Tibet region based on China's FY-2E satellite data. *Advances in Space Research*, 58, 989-1001.<http://dx.doi.org/10.1016/j.asr.2016.05.038>
- MA, T. & WU, Z. 2012. Precursor-like anomalies prior to the 2008 Wenchuan Earthquake: a critical but constructive review. *International Journal of Geophysics*, 2012.10.115/2012/583097

- MAXWELL, S. E., & DELANEY, H. D. (2004). Designing experiments and analyzing data: A model comparison perspective (2nd ed.). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers
- MEYERS, S., KELLY, B. & O'BRIEN, J. 1993. An introduction to wavelet analysis in oceanography and meteorology: with application to the dispersion of Yanai waves. *Mon. Weather Rev.*, 121, 2858-2866
- MURPHY, S. W., FILHO, C. R. D. S. & OPPENHEIMER, C. 2011. Monitoring volcanic thermal anomalies from space: Size matters. *Journal of Volcanology and Geothermal Research*, 203, 48-61. <https://doi.org/10.1016/j.jvolgeores.2011.04.008>
- NORMAN, J. M. & BECKER, F. 1995. Terminology in thermal infrared remote sensing of natural surfaces. *Agricultural and Forest Meteorology*, 77, 153-166. [http://dx.doi.org/10.1016/0168-1923\(95\)02259-Z](http://dx.doi.org/10.1016/0168-1923(95)02259-Z)
- OUZOUNOV, D. & FREUND, F. 2004. Mid-infrared emission prior to strong earthquakes analyzed by remote sensing data. *Advances in Space Research*, 33, 268-273. [http://dx.doi.org/10.1016/S0273-1177\(03\)00486-1](http://dx.doi.org/10.1016/S0273-1177(03)00486-1)
- OUZOUNOV, D., BRYANT, N., LOGAN, T., PULINETS, S. & TAYLOR, P. 2006. Satellite thermal IR phenomena associated with some of the major earthquakes in 1999–2003. *Physics and Chemistry of the Earth, Parts A/B/C*, 31, 154-163. <http://dx.doi.org/10.1016/j.pce.2006.02.036>
- OUZOUNOV, D., LIU, D., CHUNLI, K., CERVONE, G., KAFATOS, M. & TAYLOR, P. 2007. Outgoing long wave radiation variability from IR satellite data prior to major earthquakes. *Tectonophysics*, 431, 211-220. <http://dx.doi.org/10.1016/j.tecto.2006.05.042>
- PANDA, S., CHOUDHURY, S., SARAF, A. & DAS, J. 2007. MODIS land surface temperature data detects thermal anomaly preceding 8 October 2005 Kashmir earthquake. *International Journal of Remote Sensing*, 28, 4587-4596
- PAVLIDOU, E., VAN DER MEIJDE, M., VAN DER WERFF, H. & HECKER, C. 2016. Finding a needle by removing the haystack. *Comput. Geosci.*, 90, 78-86. [10.1016/j.cageo.2016.02.016](https://doi.org/10.1016/j.cageo.2016.02.016)
- PEEL, M. C., FINLAYSON, B. L. & MCMAHON, T. A. 2007. Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences Discussions*, 11, 1633-1644
- PERGOLA, N., MARCHESE, F., & TRAMUTOLI, V. 2004. Automated detection of thermal features of active volcanoes by means of infrared AVHRR records. *Remote Sensing of Environment*, 93, 311–327. <https://doi.org/10.1016/j.rse.2004.07.010>
- PIERI, D., & ABRAMS, M. 2004. ASTER watches the world's volcanoes: A new paradigm for volcanological observations from orbit. *Journal of Volcanology and Geothermal Research*, 135, 13–28. <https://doi.org/10.1016/j.jvolgeores.2003.12.018>

- PIRODDI, L. & RANIERI, G. 2012. Night thermal gradient: a new potential tool for earthquake precursors studies. An application to the seismic area of L'Aquila (central Italy). *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 5, 307-312
- PULINET, S. & DUNAJECKA, M. 2007. Specific variations of air temperature and relative humidity around the time of Michoacan earthquake M8. 1 Sept. 19, 1985 as a possible indicator of interaction between tectonic plates. *Tectonophysics*, 431, 221-230
- PULINET, S. & OUZOUNOV, D. 2011. Lithosphere–Atmosphere–Ionosphere Coupling (LAIC) model – An unified concept for earthquake precursors validation. *Journal of Asian Earth Sciences*, 41, 371-382. <http://dx.doi.org/10.1016/j.jseaes.2010.03.005>
- PULINET, S. A., MOROZOVA, L. I. & YUDIN, I. A. 2014. Synchronization of atmospheric indicators at the last stage of earthquake preparation cycle. *Research in Geophysics*, 4. doi:10.4081/rg.2014.4898
- PYAYT, A. L., KOZIONOV, A. P., MOKHOV, I. I., LANG, B., KRZHIZHANOVSKAYA, V. V. & SLOOT, P. M. A. 2013. An Approach for Real-time Levee Health Monitoring Using Signal Processing Methods. *Procedia Computer Science*, 18, 2357-2366. <https://doi.org/10.1016/j.procs.2013.05.407>
- QADER, S. H., ATKINSON, P. M. & DASH, J. 2015. Spatiotemporal variation in the terrestrial vegetation phenology of Iraq and its relation with elevation. *International Journal of Applied Earth Observation and Geoinformation*, 41, 107-117. <https://doi.org/10.1016/j.jag.2015.04.021>
- QIANG, Z.-J., XU, X.-D. & DIAN, C.-G. 1997. Case 27 thermal infrared anomaly precursor of impending earthquakes. *Pure and Applied Geophysics*, 149, 159-171
- QIN, K., GUO, G. & WU, L. 2009. Surface latent heat flux anomalies preceding inland earthquakes in China. *Earthquake Science*, 22, 555-562
- QIN, K., WU, L. X., DE SANTIS, A. & CIANCHINI, G. 2012b. Preliminary analysis of surface temperature anomalies that preceded the two major Emilia 2012 earthquakes (Italy). *Annals of Geophysics*
- QIN, K., WU, L., SANTIS, A. D., MENG, J., MA, W. & CIANCHINI, G. 2012a. Quasi-synchronous multi-parameter anomalies associated with the 2010–2011 New Zealand earthquake sequence. *Natural Hazards and Earth System Sciences*, 12, 1059-1072
- QIN, K., WU, L., ZHENG, S. & LIU, S. 2013. A deviation-time-space-thermal (DTS-T) method for global earth observation system of systems (GEOSS)-based earthquake anomaly recognition: criteria and quantify indices. *Remote Sensing*, 5, 5143-5151
- QU, C. Y., MA, J. & SHAN, X. J. 2006. Counterevidence for an earthquake precursor of satellite thermal infrared anomalies. *Chinese Journal of Geophysics*, 49, 426-431

- QUIGLEY, M. C., HUGHES, M. W., BRADLEY, B. A., VAN BALLEGOOY, S., REID, C., MORGENROTH, J., HORTON, T., DUFFY, B. & PETTINGA, J. R. 2016. The 2010–2011 Canterbury Earthquake Sequence: Environmental effects, seismic triggering thresholds and geologic legacy. *Tectonophysics*, 672, 228-274. <https://doi.org/10.1016/j.tecto.2016.01.044>
- RAMSEY, M. S., & HARRIS, A. J. L. (2013). Volcanology 2020: How will thermal remote sensing of volcanic surface activity evolve over the next decade? *Journal of Volcanology and Geothermal Research*, 249, 217–233. <https://doi.org/10.1016/j.jvolgeores.2012.05.011>
- RANDOLPH, T. 2005. Scale-based normalization of spectral data. *Cancer Biomark.*, 2
- REATH, K. A., RAMSEY, M. S., DEHN, J., & WEBLEY, P. W. (2016). Predicting eruptions from precursory activity using remote sensing data hybridization. *Journal of Volcanology and Geothermal Research*, 321, 18–30. <https://doi.org/10.1016/j.jvolgeores.2016.04.027>
- REIMANN, C., FILZMOSER, P. & GARRETT, R. G. 2005. Background and threshold: critical comparison of methods of determination. *Science of the Total Environment*, 346, 1-16. <https://doi.org/10.1016/j.scitotenv.2004.11.023>
- REZAPOUR, N., FATTAHI, M. & BIDOKHTI, A. A. 2010. Possible soil thermal response to seismic activities in Alborz region (Iran). *Nat. Hazards Earth Syst. Sci.*, 10, 459-464. [10.5194/nhess-10-459-2010](https://doi.org/10.5194/nhess-10-459-2010)
- SAJDA, P., LAINE, A. & ZEEVI, Y. 2002. Multi-Resolution and Wavelet Representations for Identifying Signatures of Disease. *Disease Markers*, DOI:18.10.1155/2002/108741
- SARADJIAN, M. & AKHOONDZADEH, M. 2011a. Prediction of the date, magnitude and affected area of impending strong earthquakes using integration of multi precursors earthquake parameters. *Natural Hazards and Earth System Sciences*, 11, 1109
- SARADJIAN, M. & AKHOONDZADEH, M. 2011b. Thermal anomalies detection before strong earthquakes ( $M > 6.0$ ) using interquartile, wavelet and Kalman filter methods. *Natural Hazards and Earth System Sciences*, 11, 1099-1108
- SARAF, A. & CHOUDHURY, S. 2005a. Cover: NOAA-AVHRR detects thermal anomaly associated with the 26 January 2001 Bhuj earthquake, Gujarat, India. *International Journal of Remote Sensing*, 26, 1065-1073
- SARAF, A. K. & CHOUDHURY, S. 2005b. Thermal remote sensing technique in the study of pre-earthquake thermal anomalies. *Journal of the Indian Geophysical Union*, 9, 197-207
- SARAF, A. K., RAWAT, V., BANERJEE, P., CHOUDHURY, S., PANDA, S. K., DASGUPTA, S. & DAS, J. 2008. Satellite detection of earthquake thermal infrared precursors in Iran. *Natural Hazards*, 47, 119-135
- SARAF, A. K., RAWAT, V., CHOUDHURY, S., DASGUPTA, S. & DAS, J. 2009. Advances in understanding of the mechanism for generation of earthquake



- thermal precursors detected by satellites. *International Journal of Applied Earth Observation and Geoinformation*, 11, 373-379. <http://dx.doi.org/10.1016/j.jag.2009.07.003>
- SCHARLEMANN, J. P. W., BENZ, D., HAY, S. I., PURSE, B. V., TATEM, A. J. & WINT, G. R. W. E. A. 2008. Global Data for Ecology and Epidemiology: A Novel Algorithm for Temporal Fourier Processing MODIS Data. *PLoS ONE* 3, 1
- SCHULZ, J., MEYWERK, J., EWALD, S. & SCHLUESSEL, P. 1997. Evaluation of Satellite-Derived Latent Heat Fluxes. *Journal of Climate*, 10, 2782-2794. [4.675/1520-0442\(1997\)010](https://doi.org/10.1175/1520-0442(1997)010)
- SHEN, X., ZHANG, X., HONG, S., JING, F. & ZHAO, S. 2013. Progress and development on multi-parameters remote sensing application in earthquake monitoring in China. *Earthquake Science*, 26, 427-437. DOI:10.1007/s11589-013-0053-9
- SHI, X., WANG, Y., LIU-ZENG, J., WELDON, R., WEI, S., WANG, T. & SIEH, K. 2017. How complex is the 2016Mw 7.8 Kaikoura earthquake, South Island, New Zealand? *Science Bulletin*, 62, 309-311. <https://doi.org/10.1016/j.scib.2017.01.033>
- SINGH, R. P., MEHDI, W., GAUTAM, R., SENTHIL KUMAR, J., ZLOTNICKI, J. & KAFATOS, M. 2010. Precursory signals using satellite and ground data associated with the Wenchuan Earthquake of 12 May 2008. *International Journal of Remote Sensing*, 31, 3341-3354
- SMITH, S. R., LEGLER, D. M. & VERZONE, K. V. 2001. Quantifying Uncertainties in NCEP Reanalyses Using High-Quality Research Vessel Observations. *Journal of Climate*, 14, 4062-4072. [10.1175/1520-0442](https://doi.org/10.1175/1520-0442)
- SOTO-PINTO, C., ARELLANO-BAEZA, A. & SÁNCHEZ, G. 2013. A new code for automatic detection and analysis of the lineament patterns for geophysical and geological purposes (ADALGEO). *Computers & Geosciences*, 57, 93-103. <http://dx.doi.org/10.1016/j.cageo.2013.03.019>
- SPAMPINATO, L., GANCI, G., HERNÁNDEZ, P. A., CALVO, D., TEDESCO, D., PÉREZ, N. M., YALIRE, M. M. (2013). Thermal insights into the dynamics of Nyiragongo lava lake from ground and satellite measurements. *Journal of Geophysical Research: Solid Earth*, 118, 5771-5784. <https://doi.org/10.1002/2013jb010520>
- STEFFKE, A. M. & HARRIS, A. J. L. 2011. A review of algorithms for detecting volcanic hot spots in satellite infrared data. *Bull Volcanol* 73. <https://doi.org/10.1007/s00445-011-0487-7>
- STOLLE, F., DENNIS, R. A., KURNIWAN, I. & LAMBIN, E. F. 2004. Evaluation of remote sensing-based active fire datasets in Indonesia. *International Journal of Remote Sensing*, 25, 471-479. DOI:10.1080/01431160310001618022
- TAKEUCHI, A., LAU, B. W. S. & FREUND, F. T. 2006. Current and surface potential induced by stress-activated positive holes in igneous rocks.

- Physics and Chemistry of the Earth, Parts A/B/C, 31, 240-247.  
<http://dx.doi.org/10.1016/j.pce.2006.02.022>
- TARY, J. B., HERRERA, R. H., HAN, J. & VAN DER BAAN, M. 2014. Spectral estimation—What is new? What is next? *Reviews of Geophysics*, 52, 723-749. DOI:10.1002/2014rg000461
- TRAMUTOLI, V. Robust AVHRR techniques (RAT) for environmental monitoring: theory and applications. *Remote Sensing*, 1998. SPIE, 13.
- TRAMUTOLI, V., ALIANO, C., CORRADO, R., FILIZZOLA, C., GENZANO, N., LISI, M., MARTINELLI, G. & PERGOLA, N. 2013. On the possible origin of thermal infrared radiation (TIR) anomalies in earthquake-prone areas observed using robust satellite techniques (RST). *Chemical Geology*, 339, 157-168.<http://dx.doi.org/10.1016/j.chemgeo.2012.10.042>
- TRAMUTOLI, V., CORRADO, R., FILIZZOLA, C., GENZANO, N., LISI, M. & PERGOLA, N. 2015. From visual comparison to Robust Satellite Techniques: 30 years of thermal infrared satellite data analyses for the study of earthquake preparation phases. *Bollettino di Geofisica Teorica ed Applicata*, 56
- 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.
- TRIGO, I., FREITAS, S., BIOUCAS-DIAS, J., BARROSO, C., MONTEIRO, I., & VITERBO, P. (2009). Algorithm theoretical basis document for land surface temperature (LST). EUMETSAT Land Surface Analysis Satellite Facility.
- TRONIN, A. 1996. Satellite thermal survey—a new tool for the study of seismoactive regions. *International Journal of Remote Sensing*, 17, 1439-55
- TRONIN, A. 2000. Thermal IR satellite sensor data application for earthquake research in China. *International Journal of Remote Sensing*, 21, 3169-3177
- TRONIN, A. 2010. Satellite Remote Sensing in Seismology. A Review. *Remote Sensing*, 2, 124
- TRONIN, A. A. 2006. Remote sensing and earthquakes: A review. *Physics and Chemistry of the Earth, Parts A/B/C*, 31, 138-142.<http://dx.doi.org/10.1016/j.pce.2006.02.024>
- TRONIN, A. A. Thermal satellite data for earthquake research. *Geoscience and Remote Sensing Symposium*, 2000. Proceedings. IGARSS 2000. IEEE 2000 International, 2000b. IEEE, 2703-2705.
- TRONIN, A. A., HAYAKAWA, M. & MOLCHANOV, O. A. 2002. Thermal IR satellite data application for earthquake research in Japan and China. *Journal of*

- Geodynamics, 33, 519-534. [http://dx.doi.org/10.1016/S0264-3707\(02\)00013-3](http://dx.doi.org/10.1016/S0264-3707(02)00013-3)
- ULUSOY, İ., LABAZUY, P. & AYDAR, E. 2012. STcorr: An IDL code for image based normalization of lapse rate and illumination effects on nighttime TIR imagery. Computers & Geosciences, 43, 63-72. <https://doi.org/10.1016/j.cageo.2012.02.012>
- UMARKHODGAEV, R. M., LIPEROVSKY, V. A., MIKHAILIN, V. V., BOGDANOV, V. V., MEISTER, C. V. & LIPEROVSKAYA, E. V. 2012. IR spectral analysis for the diagnostics of crust earthquake precursors. Nat. Hazards Earth Syst. Sci., 12, 3269-3274. DOI: 10.5194/nhess-12-3269-2012
- VAN DER MEER, F., HECKER, C., VAN RUITENBEEK, F., VAN DER WERFF, H., DE WIJKERSLOOTH, C. & WECHSLER, C. 2014. Geologic remote sensing for geothermal exploration: A review. International Journal of Applied Earth Observation and Geoinformation, 33, 255-269. <https://doi.org/10.1016/j.jag.2014.05.007>
- VAN DER WERFF, H. M. A., BAKKER, W. H., VAN DER MEER, F. D. & SIDERIUS, W. 2006. Combining spectral signals and spatial patterns using multiple Hough transforms: An application for detection of natural gas seepages. Computers & Geosciences, 32, 1334-1343. <https://doi.org/10.1016/j.cageo.2005.12.003>
- VAUGHAN, R. G., KESZTHELYI, L. P., LOWENSTERN, J. B., JAWOROWSKI, C. & HEASLER, H. 2012. Use of ASTER and MODIS thermal infrared data to quantify heat flow and hydrothermal change at Yellowstone National Park. Journal of Volcanology and Geothermal Research, 233-234, 72-89. <https://doi.org/10.1016/j.jvolgeores.2012.04.022>
- VENKATANATHAN, N. & NATYAGANOV, V. L. 2014. Outgoing longwave radiations as pre-earthquake signals: preliminary results of 24 September 2013 (M 7.7) earthquake. Current Science, 106
- WAN, Z., & DOZIER, J. 1996. A generalised split-window algorithm for retrieving land-surface temperature from space. IEEE Transactions on Geoscience and Remote Sensing, 34, 892-905.
- WANG, F., WANG, X. & ZHAO, Y. E. A. 2014. Long-term periodic structure and seasonal-trend decomposition of water level in Lake Baiyangdian, Northern China. Int. J. Environ. Sci. Technol., 11. <https://doi.org/10.1007/s13762-013-0362-5>
- WANG, L. & ZHOU, C. 1984. Anomalous variations of ground temperature before the Tangsan and Haiheng earthquakes (in Chinese). Journal of Seismological Research, 7, 649-656
- WATSON, I. M., REALMUTO, V. J., ROSE, W. I., PRATA, A. J., BLUTH, G. J. S., GU, Y., BADER, C. E., & YU, T. 2004. Thermal infrared remote sensing of volcanic emissions using the moderate resolution imaging spectroradiometer. Journal of Volcanology and Geothermal Research, 135, 75-89. <https://doi.org/10.1016/j.jvolgeores.2003.12.017>

- WEI, C., ZHANG, Y., GUO, X., HUI, S., QIN, M. & ZHANG, Y. 2013. Thermal infrared anomalies of several strong earthquakes. *The Scientific World Journal*, 2013
- WEI, L., GUO, J., LIU, J., LU, Z., LI, H. & CAI, H. 2009. Satellite Thermal Infrared Earthquake Precursor to the Wenchuan Ms 8.0 Earthquake in Sichuan, China, and its Analysis on Geo-dynamics. *Acta Geologica Sinica (English Edition)*, 83, 767-775
- WELLS, D. L. & COPPERSMITH, K. J. 1994. New empirical relationships among magnitude, rupture length, rupture width, rupture area and surface displacement. *Bulletin of the Seismological Society of America*, 84, 974-1002
- WESSELS, R. L., VAUGHAN, R. G., PATRICK, M. R., & COOMBS, M. L. (2013). High-resolution satellite and airborne thermal infrared imaging of precursory unrest and 2009 eruption at Redoubt Volcano, Alaska. *Journal of Volcanology and Geothermal Research*, 259, 248-269. <https://doi.org/10.1016/j.jvolgeores.2012.04.014>
- WEST, M. 1997. Time series decomposition. *Biometrika*, 84, 489-494. <https://doi.org/10.1093/biomet/84.2.489>
- WOOSTER, M. J., & ROTHERY, D. A. (2000). A review of volcano surveillance applications using the ATSR instrument series. *Advances in Environmental Monitoring and Modelling*, 1, 3-35.
- WOOSTER, M. J., ROBERTS, G., PERRY, G. L. W. & KAUFMAN, Y. J. 2005. Retrieval of biomass combustion rates and totals from fire radiative power observations: FRP derivation and calibration relationships between biomass consumption and fire radiative energy release. *Journal of Geophysical Research: Atmospheres*, 110, n/a-n/a. <https://doi.org/10.1029/2005jd006318>
- WOOSTER, M. J., XU, W. & NIGHTINGALE, T. 2012. Sentinel-3 SLSTR active fire detection and FRP product: Pre-launch algorithm development and performance evaluation using MODIS and ASTER datasets. *Remote Sensing of Environment*, 120, 236-254. <https://doi.org/10.1016/j.rse.2011.09.033>
- WRIGHT, R. (2015). MODVOLC: 14 years of autonomous observations of effusive volcanism from space. Geological Society, London, Special Publications, 426. <https://doi.org/10.1144/sp426.12>
- WRIGHT, R., FLYNN, L. P., GARBEIL, H., HARRIS, A. J. L., & PILGER, E. 2004. MODVOLC: Near-real-time thermal monitoring of global volcanism. *Journal of Volcanology and Geothermal Research*, 135, 29-49. <https://doi.org/10.1016/j.jvolgeores.2003.12>
- WRIGHT, R., FLYNN, L., GARBEIL, H., HARRIS, A., & PILGER, E. 2002. Automated volcanic eruption detection using MODIS. *Remote Sensing of Environment*, 82, 135-155. [https://doi.org/10.1016/S0034-4257\(02\)00030-5](https://doi.org/10.1016/S0034-4257(02)00030-5)

- WU, L., LIU, S. & WU, Y. The experiment evidences for tectonic earthquake forecasting based on anomaly analysis on satellite infrared image. Geoscience and Remote Sensing Symposium, 2006. IGARSS 2006. IEEE International Conference on, 2006. IEEE, 2158-2162.
- WU, L.-X., QIN, K. & LIU, S.-J. 2012. GEOS-2 based thermal parameters analysis for earthquake anomaly recognition. Proceedings of the IEEE, 100, 2891-2907
- XIAO, G., YUAN-SHENG, Z., MEI-JIAO, Z., WEN-RONG, S. & CONG-XIN, W. 2010. Variation characteristics of OLR for the Wenchuan earthquake. Chinese Journal of Geophysics, 53, 980-988
- XIONG, P. & SHEN, X. H. 2017. Outgoing longwave radiation anomalies analysis associated with different types of seismic activity. Advances in Space Research, 59, 1408-1415. <http://dx.doi.org/10.1016/j.asr.2016.12.011>
- XIONG, P., SHEN, X., BI, Y., KANG, C., CHEN, L., JING, F. & CHEN, Y. 2010. Study of outgoing longwave radiation anomalies associated with Haiti earthquake. Natural Hazards and Earth System Sciences, 10, 2169
- XU, W., WOOSTER, M. J., ROBERTS, G. & FREEBORN, P. 2010. New GOES imager algorithms for cloud and active fire detection and fire radiative power assessment across North, South and Central America. Remote Sensing of Environment, 114, 1876-1895. <https://doi.org/10.1016/j.rse.2010.03.012>
- YANG, G. & MI, Y. 2009. Thermal anomalies and earthquakes: evidence from Wenchuan, China. Earthquake research in China, 23
- YANG, X. & LO, C. 2000. Relative radiometric normalization performance for change detection from multi-date satellite images. Photogram. Eng. Remote Sens., 66, 967-980
- YANG, Y. & GUO, G. 2010. Studying the thermal anomaly before the Zhangbei earthquake with MTSAT and meteorological data. International Journal of Remote Sensing, 31, 2783-2791
- YAO, Q.-L. & QIANG, Z.-J. 2010. The elliptic stress thermal field prior to Ms 7.3 Yutian, and M S 8.0 Wenchuan earthquakes in China in 2008. Natural Hazards, 54, 307-322. [10.1007/s11069-009-9470-4](https://doi.org/10.1007/s11069-009-9470-4)
- YAO, Q.-L. & QIANG, Z.-J. 2012. Thermal infrared anomalies as a precursor of strong earthquakes in the distant future. Natural Hazards, 62, 991-1003. [10.1007/s11069-012-0130-8](https://doi.org/10.1007/s11069-012-0130-8)
- ZHANG, W., ZHAO, J., WANG, W., HUAZHONG, R., CHEN, L. & YAN, G. 2013. A preliminary evaluation of surface latent heat flux as an earthquake precursor.
- ZHANG, Y. S., GUO, X., WEI, C. X., SHEN, W. R. & HUI, S. X. 2011. Characteristics of seismic thermal radiation of the Japan Ms9. 0 and Myanmar Ms7. 2 earthquake. Chinese Journal of Geophysics, 54, 670-676

### *References*

---

- ZHANG, Y., GUO, X., ZHONG, M., SHEN, W., LI, W. & HE, B. 2010. Wenchuan earthquake: Brightness temperature changes from satellite infrared information. *Chinese Science Bulletin*, 55, 1917-1924
- ZHANG, Y.-B. & LIU, S.-J. 2011. Thermal radiation temperature field variation of hole rock in loading process. *Rock and Soil Mechanics*, 32, 1013-1017

## Appendix 1: Volcanic case study

This Appendix details the algorithm settings tested for processing the data. In order to define optimum choices we tested different options. We justify our final choices and provide evidence that different settings do not alter the main findings but rather optimize presentation of the results. Table A1.1 summarizes the tested settings and our final choice of settings.

*Table A1. Algorithm settings applied to produce the results.*

<b>Algorithm setting</b>	<b>Choice</b>
Normalization frame size	Side length= 15km (5 METEOSAT NG pixels)
Anomaly threshold	$mean+2\sigma$
Length of temporal window	1-day long (96 observations) in Italy; 5-day long (480 observations) in Congo
Data availability thresholds (optional)	No data availability thresholds applied

### Choice of normalization frame size

Frame size is chosen following the procedure described in Chapter 3. Time series variability is measured using the standard deviation sigma ( $\sigma$ ). The average sigma ( $\sigma$ ) of all-time series in the image is calculated and used as a metric for the choice of frame size for normalization: the frame size that minimizes the average  $\sigma$  is chosen. Different frame sizes are tested, with length of sides as follows:

*Congo*: 9km (3px), 15km (5px), 27km (9px) and 51km (17px);  
*Mount Etna*: 9km (3px), 15km (5px), 21km (7px), 33km (11px) and 51km (17px).

Note that reference to length is based on nominal spatial resolution of the data, which is 3km at nadir. Congo is located almost at nadir; in the case of Etna ground pixel size would be in reality approximately 4km.

In both study areas, an increase in frame size leads to an increase in average variability of the normalized series of the whole image (Figure A1.1a).

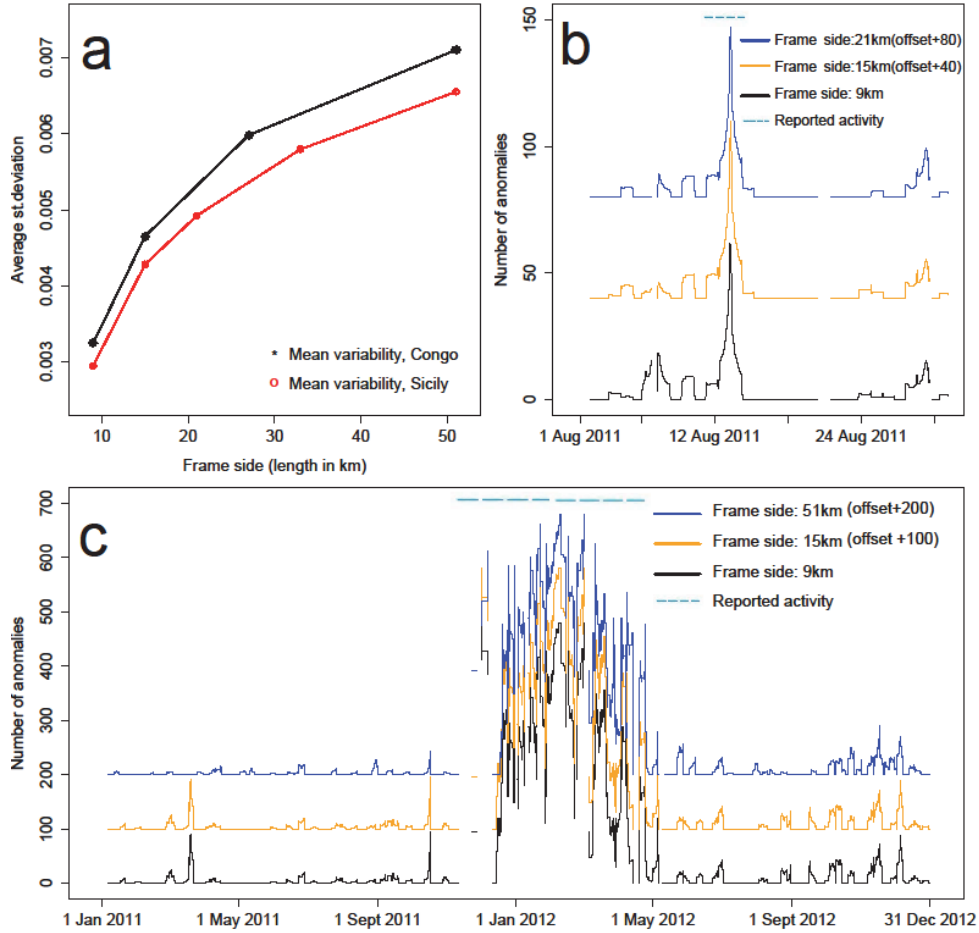


Figure A1.1. Detection using frames of different sides for normalization. Panel [a] shows the increase of variability (average  $\sigma$  of all normalized series in the image) due to the application of larger frames. Panels [b] and [c] show detection results from Mount Etna and Nyamuragira, respectively. Different colours correspond to the application of different sizes of normalization frames, and have been offset for clarity.

Lava flows may reach lengths of 10-15km and thus may extend to two adjacent pixels. In this case a hotspot would be shared between the frame and the central pixel, and would consequently be suppressed in the normalization. The smallest size of frame (3-pixel side) is therefore excluded; the next best choice is the 5-pixel-side (15 km) frame, in both study areas. Figure A1.1, panels [b] and [c], show results of detection with different frames in both study areas. The same hotspots are visible regardless of the applied frame; the main differences are seen in the distribution and size of minor peaks.



### Choice of anomaly threshold

The methodology of Chapter 3 applies a mean+2 $\sigma$  threshold to define which normalized values are considered anomalous. Researchers have been using thresholds in previous work, although following different methodologies (see, for example, Blackett [2014]). It is often discussed that the choice of optimum threshold may differ between volcanic targets. In view of that, different thresholds are tested in both study areas, namely: mean+1.5 $\sigma$ , mean+2 $\sigma$ , mean+2.5 $\sigma$ , mean+3 $\sigma$ . Figure A1.2 shows the results obtained by using different thresholds for detecting anomalies. The largest confirmed hotspots in Mount Etna and Nyamuragira are clearly detected regardless of the applied threshold. Since there was no particular focus to events of higher or lower intensities, the mean+2 $\sigma$  threshold is used as initially defined.

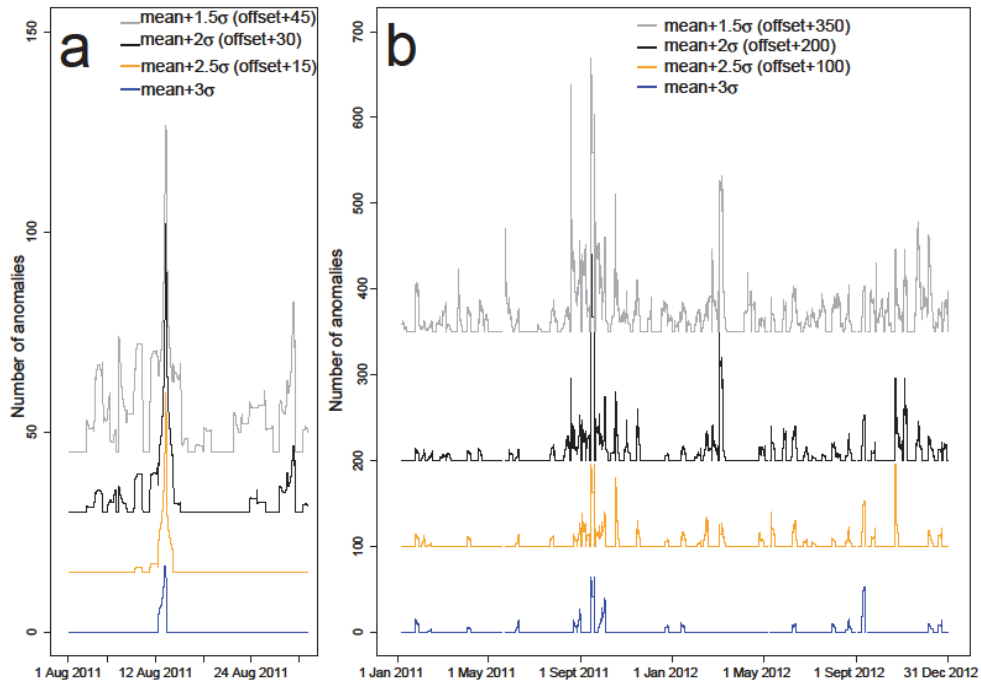


Figure A1.2. Detection with different thresholds in Mount Etna (Panel [a]) and Nyamuragongo (Panel [b]). The graphs are offset for clarity. Higher thresholds capture only the highest normalized values (see for example the case of mean+3 $\sigma$  threshold shown in blue in both areas). Lower thresholds capture also fluctuations of lower intensity (see the red line showing detection with a mean+1.5 $\sigma$  threshold).

### Choice of length of the temporal moving window

Temporal moving windows of one and five days long are tested. These are respectively shorter and longer than the events reported in Mount Etna and Congo. The peaks that correspond to volcanic activity are in both cases

detected. When a window longer than the event is used, the resulting peak extends to a longer period in the results and thus the duration of the event is not easy to interpret from the graphs. For example, see in Figure A1.3a time series from Mount Etna processed with a 5-day window to detect a 33hr-long event. Results are therefore presented based on a 1-day long window for Mount Etna. On the other hand, windows shorter than the event display a large density of small peaks, and the result does not intuitively show the duration and intensity of the event in the graph. Such is the case of a 1-day window applied to detect lava flows in the 6-month active period in Congo, shown in Figure A1.3b. Results from the application of a 5-day long window are therefore presented in the case of Congo.

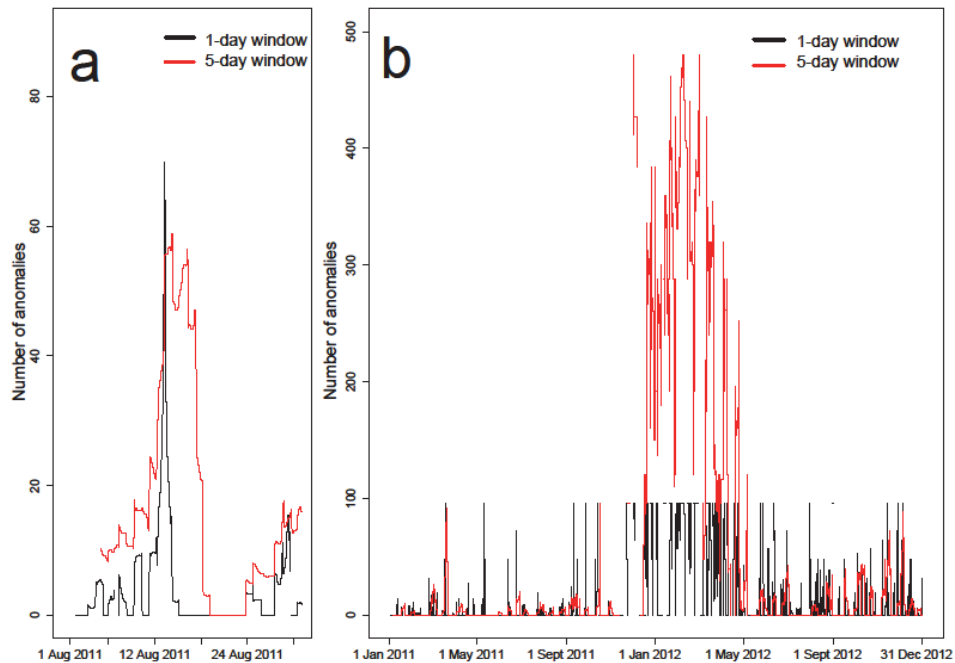


Figure A1.3. Moving temporal windows are influenced by the length of time. Panel [a] shows results from Mount Etna; panel [b] shows results from Nyamuragira lava flows in Congo.

### Optional thresholds for data availability

The original methodology allows to set thresholds for data availability. Data may be missing for three reasons: (a) the observation does not exist (registration issues with the satellite) (b) the LST value is not used because there was too much uncertainty calculated for its retrieval, (c) the observation has been masked as a cloud/water body. In this study, however, it is the case that volcanic activity may be related to missing data due to volcanic emissions;

data availability thresholds should thus not be used. This section describes the tests carried out to support this decision.

Frames are tested that are at least 2%, 26%, 50%, 74% and 98% complete, without applying any threshold for the temporal windows. Stricter requirements for data availability in the frame introduce missing data in the results (Fig. A1.4, panel [a]). Time series from Nyiragongo are more heavily affected by missing data (Fig. 5.6, panel [d]). For both Nyamuragira and Mount Etna missing data appear at the time of reported activity (Fig. A1.4, panels [b], [c]). The introduced missing data coincide with reported activity. This suggests there is a relation between volcanic activity and missing data and a data availability threshold for the normalization should not be used. This is further supported with examination of the MWIR dataset. In all tests, the main findings remain the same at least up to frames min.74% complete.

Temporal windows are tested that are at least 1%, 25%, 50%, 75% and 99% complete, without applying any threshold for the normalization frames. Similarly as before, strict requirements for full windows introduce missing data in the final results (Fig. A1.5, panel [a]). In the case of Mount Etna, the main findings remain unchanged up to windows which are min.75% complete (Fig A1.5, panel [c]). In both cases in Congo, however, a requirement above min.25% full windows leaves less than 90% of the results for study (Fig A1.5, panels [b] and [d]). In both study areas, missing data introduced by the application of the data availability threshold coincide with times of reported volcanic activity. This suggests that missing data may be related to volcanic activity (degassing); data availability thresholds on the temporal window should thus not be used in this study.

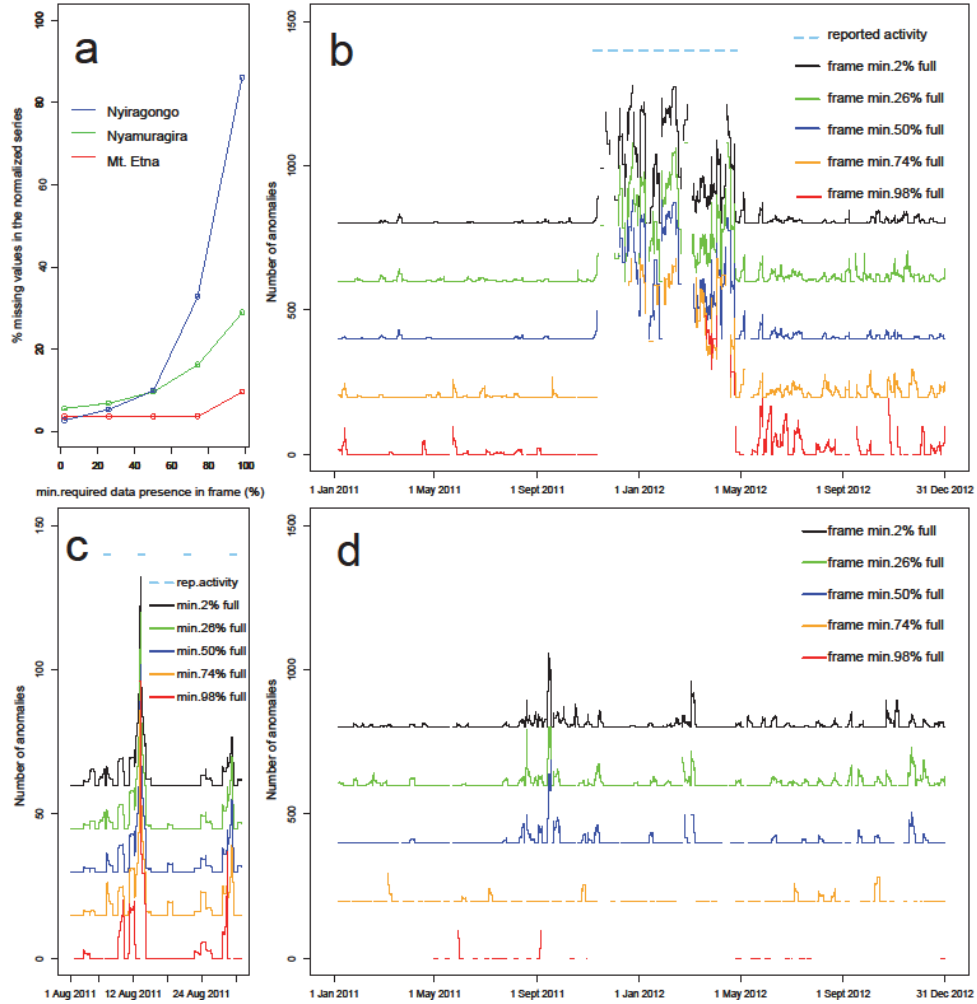


Figure A1.4. Application of data availability thresholds on the normalization frame. Panel [a] shows the increase in missing data in the results with the application of stricter thresholds. Panels [b], [c] and [d] show results from Nyamuragira, Mount Etna and Nyiragongo respectively. No data availability threshold was used in the results presented in the paper. Each series graph is offset +15 (in Mount Etna) and +200 (in Congo) from its immediate neighbour for clarity.

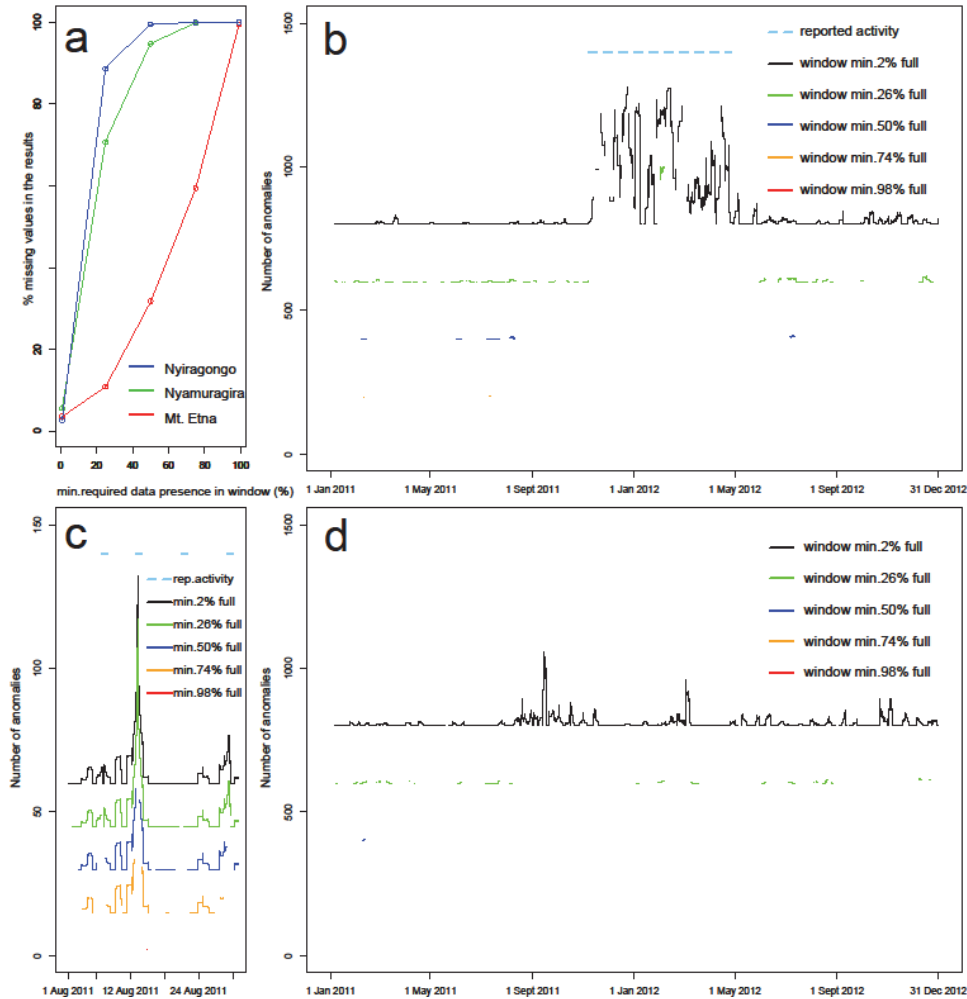


Figure A1.5. Application of data availability thresholds on the temporal window. Panel [a] shows the increase in missing data in the results with the application of stricter thresholds. Panels [b], [c] and [d] show results from Nyamuragira, Mount Etna and Nyiragongo respectively. When graphs do not show, it is due to lack of data. Each series graph is offset +15 (in Mount Etna) and +200 (in Congo) from its immediate neighbor for clarity.

## ***Appendix 2: Anomaly density***

The following tables show the anomaly density, calculated as number of anomalies per period per spatial zone, for all studied earthquakes and for all experimental configurations tested ( $\mu+2\sigma$  anomaly detection threshold,  $\mu+3\sigma$  anomaly detection threshold, 48-hr co-seismic period, adjusted co-seismic period).

A1. Anomaly Density data ,  $\mu+2\sigma$  threshold, no co-seismic period adjustment

		EQ-year					No-EQ year				
		pre-EQ	2months-pre EQ	co-EQ	2months-post EQ	post-EQ	pre-EQ	2months-pre EQ	co-EQ	2months-post EQ	post-EQ
Ahar, Iran	< 7pixels	021135	027950	017807	019092	035467	027762	027446	026400	014363	022903
	7-14 pixels	021135	027950	017807	019092	035467	027762	027446	026400	014363	022903
	14-28 pixels	034785	030002	005623	015062	044364	040471	025174	032406	015414	041806
	> 28 pixels	044776	021579	004978	012389	055273	046697	011893	017736	009955	046465
Yunnan, China	< 7pixels	004623	004923	032696	084831	095380	034712	007381	000973	074601	088128
	7-14 pixels	004623	004923	032696	084831	095380	034712	007381	000973	074601	088128
	14-28 pixels	011694	011023	014440	049838	086385	028497	014474	006287	048904	072409
	> 28 pixels	009605	013033	022653	042318	054138	030772	010373	013195	027016	052710
Burma, Myanmar	< 7pixels	015184	009850	000531	054030	042194	013649	010277	004928	050089	045142
	7-14 pixels	015184	009850	000531	054030	042194	013649	010277	004928	050089	045142
	14-28 pixels	024439	018179	077245	068335	045099	027349	013019	034041	045992	041994
	> 28 pixels	027618	029359	061891	116304	053547	029214	022831	044940	086053	063631
Shwebo, Myanmar	< 7pixels	039428	086047	075064	032355	001899	034180	114888	001857	035696	117966
	7-14 pixels	039428	086047	075064	032355	001899	034180	114888	001857	035696	117966
	14-28 pixels	056289	069766	083124	038665	006864	056310	081214	016313	034945	037292
	> 28 pixels	084570	092936	007103	015057	004718	091511	079515	040401	017232	007104
Van, Turkey	< 7pixels	032686	008359	000562	044269	087465	036204	008543	020466	043717	015539
	7-14 pixels	032686	008359	000562	044269	087465	036204	008543	020466	043717	015539
	14-28 pixels	027498	020198	012339	041648	029700	030597	020714	079663	056800	007441
	> 28 pixels	039455	018870	020467	034796	021968	038899	019628	086890	055865	011517
Van, aftershock	< 7pixels	034315	016724	021196	037304		039306	008491	087336	060261	000000
	7-14 pixels	034315	016724	021196	037304		039306	008491	087336	060261	000000
	14-28 pixels	020588	026381	013248	035850		021724	024255	028216	057024	
	> 28 pixels	038300	028611	028523	028723		038719	026324	033383	054527	003816
Baja, aftershock	< 7pixels	010916	061140	032153	128295	020738	011317	050706	030543	145943	031649
	7-14 pixels	010916	061140	032153	128295	020738	011317	050706	030543	145943	031649
	14-28 pixels	024456	027894	012243	057134	025541	021885	028709	013944	077759	035145
	> 28 pixels	020752	027633	019196	048027	017486	019064	025472	015511	055889	024482
Emilia 20May, Italy	< 7pixels	008346	031332	008017	041269	046916	002924	025992	009117	038208	041111
	7-14 pixels	008346	031332	008017	041269	046916	002924	025992	009117	038208	041111
	14-28 pixels	015483	024820	034074	034575	027476	015747	026135	015458	037274	024121
	> 28 pixels	013472	027839	047950	035261	033666	019621	025430	018480	039626	028021
Emilia 29May, Italy	< 7pixels	009508	024341	001679	053882	040130	003112	024998	018080	049704	035065
	7-14 pixels	009508	024341	001679	053882	040130	003112	024998	018080	049704	035065
	14-28 pixels	015768	030917	012333	037004	027055	014335	033682	027489	038069	023212
	> 28 pixels	013841	030566	023706	036417	033412	019022	027262	020441	043497	026577
Norcia, Italy	< 7pixels	024678	017509	015477	041275	009843	019830	018478	059328	029370	
	7-14 pixels	024678	017509	015477	041275	009843	019830	018478	059328	029370	
	14-28 pixels	033889	044265	004714	023612	012410	028637	029879	025483	029240	
	> 28 pixels	033856	036863	009922	026169	018998	028826	026868	018085	026686	
Amatrice, Italy	< 7pixels	036824	004261	000712	015974	039913	026822	010365	006450	015111	030882
	7-14 pixels	036824	004261	000712	015974	039913	026822	010365	006450	015111	030882
	14-28 pixels	033720	044636	035146	047061	023021	026550	039311	009318	031366	029790
	> 28 pixels	033236	037235	019075	034831	027185	026931	035478	009462	026045	027103
Visso, Italy	< 7pixels	024553	017282	013537	040847	002098	020879	021750	011222	029643	001182
	7-14 pixels	024553	017282	013537	040847	002098	020879	021750	011222	029643	001182
	14-28 pixels	032957	041965	052561	023882	003469	028123	027778	082557	030440	001119
	> 28 pixels	034146	034890	080918	025714	012108	028947	026546	037695	026961	007097
Louisiana, USA	< 7pixels	025126	027319	007676	038126	013220	017712	027459	016909	015348	015123
	7-14 pixels	025126	027319	007676	038126	013220	017712	027459	016909	015348	015123
	14-28 pixels	031770	046976	016674	032897	020587	027853	036526	008821	028703	017521
	> 28 pixels	031135	038348	018301	029741	019072	029389	043057	019694	020016	023721
Canterbury, NZ	< 7pixels	023806	007217	002750	039343	056814	020780	009997	083098	054323	087069
	7-14 pixels	023806	007217	002750	039343	056814	020780	009997	083098	054323	087069
	14-28 pixels	027212	011557	018151	031817	040974	026580	013491	023815	034843	036137
	> 28 pixels	025355	014514	012134	027969	042371	027626	013121	032454	032856	047765
Oklahoma 2011, USA	< 7pixels	016807	052541	014806	014744	000000	023658	018398	000000	026087	057981
	7-14 pixels	016807	052541	014806	014744	000000	023658	018398	000000	026087	057981
	14-28 pixels	028281	021293	040923	032953	005993	031414	024231	000891	029195	048336
	> 28 pixels	031627	017721	038934	023608	019578	037753	025515	005081	026220	073658
Oklahoma 2016, USA	< 7pixels	017334	012717	018098	020200	016100	022149	017769	000579	014803	023588
	7-14 pixels	017334	012717	018098	020200	016100	022149	017769	000579	014803	023588
	14-28 pixels	028627	027972	020827	024171	021221	035002	025990	022154	027374	026671
	> 28 pixels	028619	035670	016010	021847	025890	034300	042935	029883	025645	025933
Christchurch, NZ	< 7pixels	021186	018698	023716	038993	020777	083025	042089		030846	018096
	7-14 pixels	021186	018698	023716	038993	020777	083025	042089		030846	018096
	14-28 pixels	088066	047106	058815	058361	026016	010239	051162	000000	025514	030880
	> 28 pixels	047890	042723	045079	036931	019900	070097	049580	073492	026811	025541
Kaikoura, NZ	< 7pixels	026198	031689	007631	025868	011905	021326	034344	011143	076927	000000
	7-14 pixels	026198	031689	007631	025868	011905	021326	034344	011143	076927	000000
	14-28 pixels	026080	021450	064424	050960	073413	023548	033311	054571	052485	005263
	> 28 pixels	023775	026998	028069	030122	029676	029676	034836	040980	052143	026372
Baja, Mexico	< 7pixels	010206	024771	050336	050017	033106	009883	026768	008466	056307	044572
	7-14 pixels	010206	024771	050336	050017	033106	009883	026768	008466	056307	044572
	14-28 pixels	026234	018639	027677	026995	039494	011241	021444	035974	035377	055114
	> 28 pixels	018483	021758	013396	026189	026692	013296	018767	018886	024136	033907
Baja aftershock, Mexico	< 7pixels	021783	043912	012697	040674	029132	021899	024765	019868	044178	041221
	7-14 pixels	021783	043912	012697	040674	029132	021899	024765	019868	044178	041221
	14-28 pixels	029516	020068	017299	026231	044208	012863	022664	032683	032759	055973
	> 28 pixels	018446	021700	013416	026157	026679	013248	018776	018877	024723	034022

## Appendices

### A2. Anomaly Density data , $\mu+2\sigma$ threshold, with co-seismic period adjustment

		EQ-year					No-EQ-year				
		pre-EQ	2months-pre-EQ	co-EQ	2months-post-EQ	post-EQ	pre-EQ	2months-pre-EQ	co-EQ	2months-post-EQ	post-EQ
Ahar, Iran	up to 7pixels	021135	027950	017807	019092	035467	027762	027446	026400	014363	022903
	7-14pixels	021135	027950	017807	019092	035467	027762	027446	026400	014363	022903
	14-28pixels	034785	030002	003623	013062	044364	040471	025174	032406	015404	041806
	more than 28pixels										
	more than 28pixels	044776	021579	004978	017389	055273	046697	011893	017736	009955	046465
Yunnan, China	up to 7pixels	004623	004623	032696	084831	095380	034712	007381	000973	074601	088128
	7-14pixels	004623	004623	032696	084831	095380	034712	007381	000973	074601	088128
	14-28pixels	011694	011023	014440	045838	086385	023497	014474	006287	048904	072469
	more than 28pixels										
	more than 28pixels	009605	013033	022653	042318	054138	030772	010373	013195	027016	052710
Burma, Myanmar	up to 7pixels	015184	006850	000531	054030	042194	013649	010277	004628	050089	045142
	7-14pixels	015184	006850	000531	054030	042194	013649	010277	004628	050089	045142
	14-28pixels	024439	018179	077245	068335	045699	027349	013019	034341	045962	041964
	more than 28pixels										
	more than 28pixels	027618	029259	061891	116304	053547	029214	022831	044040	086053	063631
Shwabo, Myanmar	up to 7pixels	039428	086047	075064	032355	001899	034180	014888	001857	035696	117966
	7-14pixels	039428	086047	075064	032355	001899	034180	014888	001857	035696	117966
	14-28pixels	036289	069766	083124	038965	006864	056310	081214	016313	034845	037292
	more than 28pixels										
	more than 28pixels	084570	092996	007103	015057	004719	091511	079515	004041	017232	007104
Van, Turkey	up to 7pixels	034687	013330	046847	048775		046700	010027	003529	052928	
	7-14pixels	031390	020566	051548	041868		037742	021483	032276	058928	
	14-28pixels	023006	017208	050817	035123		029452	019767	040346	054512	
	more than 28pixels										
	more than 28pixels	043207	019727	042104	029087		040816	020392	045631	056242	
Van, aftershock	up to 7pixels	025126	027319	007676	038126	013220	017712	027459	016909	015348	015123
	7-14pixels	025126	027319	007676	038126	013220	017712	027459	016909	015348	015123
	14-28pixels	031770	046676	016674	032897	020587	027853	036526	008821	028703	017521
	more than 28pixels										
	more than 28pixels	031135	038348	018301	029741	019072	029389	043057	019694	020016	023721
Baja, aftershock	up to 7pixels	016807	052541	014806	014744	000000	023658	018368	000000	026087	057981
	7-14pixels	016807	052541	014806	014744	000000	023658	018368	000000	026087	057981
	14-28pixels	028281	021293	040923	032953	005693	031414	024231	000891	029195	048336
	more than 28pixels										
	more than 28pixels	031627	017721	038934	023608	019578	037753	025515	005081	026220	073658
Emilia 20May, Italy	up to 7pixels	017334	012717	018068	020200	016100	022149	017769	000579	014803	023588
	7-14pixels	017334	012717	018068	020200	016100	022149	017769	000579	014803	023588
	14-28pixels	028627	027972	020827	024711	021221	035002	025990	021554	027374	026671
	more than 28pixels										
	more than 28pixels	028619	035670	016010	021847	025890	034300	042935	029883	025645	025933
Emilia 29May, Italy	up to 7pixels	021186	018968	023716	038993	020777	083025	042089	001348	030846	018096
	7-14pixels	021186	018968	023716	038993	020777	083025	042089	001348	030846	018096
	14-28pixels	088066	047106	058815	058961	026016	010239	031162	000000	025514	030880
	more than 28pixels										
	more than 28pixels	047890	042723	054079	036931	019900	070097	049580	073492	026811	025541
Nordic, Italy	up to 7pixels	039473	034286	030333	064605		025364	048746	048111	010159	
	7-14pixels	031939	025223	026889	053023		025825	038067	033051	061457	
	14-28pixels	027607	034159	025624	041031		046979	038889	037871	057016	
	more than 28pixels										
	more than 28pixels	024216	025847	025907	030079		024032	033234	036506	051278	
Amatrice, Italy	up to 7pixels	013747	003212	012225			020780	009997	069751		
	7-14pixels	013747	003212	012225			020780	009997	069751		
	14-28pixels	023527	011040	037009			026580	013491	035345		
	more than 28pixels										
	more than 28pixels	025432	014996	036041			027626	013121	039981		
Vesio, Italy	up to 7pixels	033938	012839	080133			022732	022206	023820		
	7-14pixels	033938	022837	084232			025586	031744	027846		
	14-28pixels	032053	038558	033556			025579	038080	030862		
	more than 28pixels										
	more than 28pixels	033739	036328	030379			027084	034447	025950		
Louisiana, USA	up to 7pixels						002465	023769	026464	044233	039077
	7-14pixels						002465	023769	026464	044233	039077
	14-28pixels						015112	027617	032563	037368	023463
	more than 28pixels										
	more than 28pixels	044736	017921	013600	009933	035113	019509	025366	026450	043356	026590
Canterbury, NZ	up to 7pixels	023093	029970	056518	067285	027524	013233	023931	056790	109732	041161
	7-14pixels	027347	024345	044035	121038	030726	014298	021954	050329	130182	040564
	14-28pixels	037351	023074	031368	064130	028416	014578	023667	035878	092773	033848
	more than 28pixels										
	more than 28pixels	018138	021864	026456	048117	017260	013370	018549	026604	055737	024525
Oklahoma 2011, USA	up to 7pixels	003685	012442	000000	002454	000000	006869	003833	000000	005980	021550
	7-14pixels	003685	012442	000000	002454	000000	006869	003833	000000	005980	021550
	14-28pixels	007524	003754	010212	010707	000000	009486	004571	000000	006967	012206
	more than 28pixels										
	more than 28pixels	007801	009922	011429	004570	004903	010147	004735	000168	005491	022440
Oklahoma 2016, USA	up to 7pixels	003995	001153	000922	002875	002536	003650	006212	000000	002699	005908
	7-14pixels	003995	001153	000922	002875	002536	003650	006212	000000	002699	005908
	14-28pixels	007655	007680	002960	004595	004698	009206	008873	000132	006248	007376
	more than 28pixels										
	more than 28pixels	006526	008872	002498	004194	005195	008005	013240	009390	004727	005435
Christchurch, NZ	up to 7pixels	000000	002816	009028	006737	004071	010494	009139		007143	003483
	7-14pixels	000000	002816	009028	006737	004071	010494	009139		007143	003483
	14-28pixels	019770	012054	005538	014095	005458	010525	000000		004266	008558
	more than 28pixels										
	more than 28pixels	011035	006920	006952	006850	007752	016366	009772	013155	003648	004710
Kalkoura, NZ	up to 7pixels	005732	008265	000000	002594	000000	003273	008588	000000	022045	000000
	7-14pixels	005732	008265	000000	002594	000000	003273	008588	000000	022045	000000
	14-28pixels	004122	003511	007257	008892	000326	003729	009150	011413	009540	000000
	more than 28pixels										
	more than 28pixels	003785	004300	003162	004449	002916	003584	006696	008277	013275	015906
Baja, Mexico	up to 7pixels	000353	004596	006074	005579	006229	001523	005493	000000	011972	010548
	7-14pixels	000353	004596	006074	005579	006229	001523	005493	000000	011972	010548
	14-28pixels	002316	002984	003647	004447	007639	002603	003427	009061	007326	014084
	more than 28pixels										
	more than 28pixels	002629	003853	000876	004088	004963	001378	002522	003122	004526	008413
Baja aftershock, Mexico	up to 7pixels	003485	008730	000000	006588	005306	003574	004681	000000	010325	009843
	7-14pixels	003485	008730	000000	006588	005306	003574	004681	000000	010325	009843
	14-28pixels	003321	003637	001857	003885	009236	002317	003158	005892	006501	014640
	more than 28pixels										
	more than 28pixels	002610	003830	000879	004081	004958	001374	002513	003122	004524	008426



A3. Anomaly Density data ,  $\mu+3\sigma$  threshold, no co-seismic period adjustment

		EQ-year				No-EQ year			
		pre-EQ	2months-preEQ	co-EQ	2months-post-EQ	post-EQ	2months-pre-EQ	co-EQ	2months-post-EQ
Ahar, Iran	up to 7pixels	005277	004619	000152	001106	013578	005750	001373	007555
	7-14 pixels	005277	004619	000152	001106	013578	005750	001373	007555
	14-28 pixels	010502	004950		001482	014850	012947	004101	004881
	more than 28 pixels								
	more than 28 pixels	015067	003020	000102	001654	013566	017085	001599	003107
Yunnan, China	up to 7pixels	001911	001197	000595	016992	024132	008633	001748	000000
	7-14 pixels	001911	001197	000595	016992	024132	008633	001748	000000
	14-28 pixels	001567	001452	001336	012342	020057	004748	002874	000357
	more than 28 pixels								
	more than 28 pixels	000415	001034	001346	007775	010624	005985	001396	000966
Burma, Myanmar	up to 7pixels	002842	000908	000000	010159	010885	001133	000520	000000
	7-14 pixels	002842	000908	000000	010159	010885	001133	000520	000000
	14-28 pixels	001975	001549	009543	011704	006154	006184	001361	010583
	more than 28 pixels								
	more than 28 pixels	005497	005635	015716	035327	011144	005629	003317	013618
Shwabo, Myanmar	up to 7pixels	013503	026158	024472	009812	000000	009268	037721	000000
	7-14 pixels	013503	026158	024472	009812	000000	009268	037721	000000
	14-28 pixels	016118	021672	009594	008483	000145	017226	021151	000887
	more than 28 pixels								
	more than 28 pixels	019031	017681	000331	001817	000320	021113	010617	000170
Van, Turkey	up to 7pixels	009587	009951	000000	015046	032475	009551	001995	003946
	7-14 pixels	009587	009951	000000	015046	032475	009551	001995	003946
	14-28 pixels	007981	002464	000965	014459	011341	007433	002261	001684
	more than 28 pixels								
	more than 28 pixels	012883	003185	002874	009551	005360	010958	002913	028691
Van, aftershock	up to 7pixels	008249	004489	000000	010887	058027	008328	001055	031718
	7-14 pixels	008249	004489	000000	010887	058027	008328	001055	031718
	14-28 pixels	006087	007080	003891	012976	012815	005181	005297	033065
	more than 28 pixels								
	more than 28 pixels	012179	006207	005199	007922	031925	010734	005989	017451
Baja, aftershock	up to 7pixels	001981	010296	000000	034544	003620	002332	010814	000000
	7-14 pixels	001981	010296	000000	034544	003620	002332	010814	000000
	14-28 pixels	003829	004911	001229	012927	005465	003388	006139	002699
	more than 28 pixels								
	more than 28 pixels	003353	004448	002319	010487	002585	002895	004695	001518
Emilia 20May, Italy	up to 7pixels	001262	006742	000000	006355	007066	000444	004856	000000
	7-14 pixels	001262	006742	000000	006355	007066	000444	004856	000000
	14-28 pixels	002881	005999	011892	008095	007525	003079	007240	000810
	more than 28 pixels								
	more than 28 pixels	003387	006555	019360	006982	007937	004512	005749	002296
Emilia 29May, Italy	up to 7pixels	001615	005983	000000	010656	007225	001123	005013	000141
	7-14 pixels	001615	005983	000000	010656	007225	001123	005013	000141
	14-28 pixels	002806	007976	000380	009037	007239	002650	008551	004899
	more than 28 pixels								
	more than 28 pixels	003311	007539	002321	007324	008008	004255	005900	002262
Nordia, Italy	up to 7pixels	005638	005787	001801	006551	000000	008607	003544	007505
	7-14 pixels	005638	005787	001801	006551	000000	008607	003544	007505
	14-28 pixels	008702	015030	000000	006122	000259	003330	007051	003628
	more than 28 pixels								
	more than 28 pixels	007949	009981	000123	005716	010192	002736	006310	002027
Amatrice, Italy	up to 7pixels	008454	000497	000000	005533	006622	005249	002019	000000
	7-14 pixels	008454	000497	000000	005533	006622	005249	002019	000000
	14-28 pixels	009523	010640	005927	016007	006417	006552	007778	000327
	more than 28 pixels								
	more than 28 pixels	005528	007771	000801	008979	005588	006317	006672	000297
Visso, Italy	up to 7pixels	005983	005397	000000	005884	000000	003810	004108	001887
	7-14 pixels	005983	005397	000000	005884	000000	003810	004108	001887
	14-28 pixels	008160	013869	010506	006425		006514	005845	016298
	more than 28 pixels								
	more than 28 pixels	008047	009122	029068	005508	006614	006266	004726	008645
Louis, USA	up to 7pixels	003746	006145	001667	010598	001651	003451	005177	001348
	7-14 pixels	003746	006145	001667	010598	001651	003451	005177	001348
	14-28 pixels	006474	013550	003721	008012	004335	006504	008226	000774
	more than 28 pixels								
	more than 28 pixels	005553	007182	001014	006771	003889	006465	010004	002765
Canterbury, NZ	up to 7pixels	003633	001301	000000	005844	012690	002743	001766	030548
	7-14 pixels	003633	001301	000000	005844	012690	002743	001766	030548
	14-28 pixels	003984	001742	005235	006594	007803	005551	002637	002613
	more than 28 pixels								
	more than 28 pixels	003892	001537	001522	004689	007243	004238	001413	005399
Oklahoma 2011, USA	up to 7pixels	003685	012442	000000	002454	000000	009899	003833	000000
	7-14 pixels	003685	012442	000000	002454	000000	009899	003833	000000
	14-28 pixels	007524	003754	010212	010707	000000	009486	004571	000000
	more than 28 pixels								
	more than 28 pixels	007801	003622	011420	004570	004603	010147	004723	000168
Oklahoma 2016, USA	up to 7pixels	003995	001153	001922	002875	002536	003650	006212	000000
	7-14 pixels	003995	001153	001922	002875	002536	003650	006212	000000
	14-28 pixels	007555	007680	002960	004595	004698	009206	008873	000132
	more than 28 pixels								
	more than 28 pixels	005526	008872	002498	004194	005195	008005	013240	009330
Christchurch, NZ	up to 7pixels	000000	002816	005028	006737	004071	010494	008139	007143
	7-14 pixels	000000	002816	005028	006737	004071	010494	008139	007143
	14-28 pixels	019770	012054	005538	014095	005488	010625	000000	004266
	more than 28 pixels								
	more than 28 pixels	011035	006920	006952	006890	002752	016396	009272	013155
Kaikoura, NZ	up to 7pixels	005732	008265	000000	002594	000000	003273	008588	000000
	7-14 pixels	005732	008265	000000	002594	000000	003273	008588	000000
	14-28 pixels	004122	003511	007257	008892	003326	003729	009150	011413
	more than 28 pixels								
	more than 28 pixels	003785	004300	003162	004449	002916	003584	006696	008277
Baja, Mexico	up to 7pixels	003353	004196	006074	005579	006229	001523	005493	000000
	7-14 pixels	003353	004196	006074	005579	006229	001523	005493	000000
	14-28 pixels	002316	002584	003647	004447	007639	002603	002427	009061
	more than 28 pixels								
	more than 28 pixels	002629	003853	000876	004088	004963	001378	002522	003122
Baja aftershock, Mexico	up to 7pixels	003485	008730	000000	006588	005306	003574	004681	000000
	7-14 pixels	003485	008730	000000	006588	005306	003574	004681	000000
	14-28 pixels	003121	003637	001857	003895	009236	002317	003158	005892
	more than 28 pixels								
	more than 28 pixels	002610	003830	000879	004081	004958	001374	002513	003122

## Appendices

### A4. Anomaly Density data , $\mu+3\sigma$ threshold, with co-seismic period adjustment

		EQ-year				No-EQ year					
		pre-EQ	2months-preEQ	co-EQ	2months-postEQ	post-EQ	pre-EQ	2months-pre EQ	co-EQ	2months-post EQ	post-EQ
Ahar, Iran	up to 7pixels	005277	004619	000152	001106	013578	005750	001573	007555	000578	004402
	7-14 pixels	005277	004619	000152	001106	013578	005750	001573	007555	000578	004402
	14-28 pixels	010502	004950	000152	001482	014850	012947	004101	004881	001881	013896
	more than 28 pixels	015067	003020	000102	001654	019566	017085	001599	003107	001505	014867
	up to 7pixels	001511	001157	000595	016992	024132	018633	001748	000000	019187	021766
Yunnan, China	7-14 pixels	001911	001157	000595	016992	024132	018633	001748	000000	019187	021766
	14-28 pixels	001567	001432	001356	012342	020057	004748	002824	000267	011776	016412
	more than 28 pixels	000415	001034	001346	007775	010624	005985	001396	000966	005578	013014
	up to 7pixels	002842	000908	000000	010159	010885	001133	000520	000000	010858	009990
	7-14 pixels	002842	000908	000000	010159	010885	001133	000520	000000	010858	009990
Burma, Myanmar	14-28 pixels	001975	001549	029543	011704	009154	006184	001361	010585	013748	012197
	more than 28 pixels	005497	005635	015716	035327	011144	005629	003317	013618	022568	013173
	up to 7pixels	013503	026158	024472	009812	000000	009268	037721	000000	004457	012333
	7-14 pixels	013503	026158	024472	009812	000000	009268	037721	000000	004457	012333
	14-28 pixels	016118	021672	009594	008483	000145	012726	021151	000887	004842	003011
Shwabo, Myanmar	more than 28 pixels	019031	017681	000331	001817	000320	021113	010617	000170	002343	000566
	up to 7pixels	009690	001587	018830	017104	009804	009694	004697	016400		
	7-14 pixels	008297	002175	020366	015967		008972	002751	008763	012677	
	14-28 pixels	008290	002636	018431	013582		008571	002350	010955	019690	
	more than 28 pixels	014190	003234	010019	007844		011306	003132	013635	018207	
Van, aftershock	up to 7pixels	008497	001874	006799			004588	003790	004694		
	7-14 pixels	010673	004040	008781			006965	006332	006924		
	14-28 pixels	008034	008456	009397			006356	007931	007719		
	more than 28 pixels	008646	007140	007275			006249	006441	005422		
	up to 7pixels	003285	005821	011962	016002	005266	001813	005168	013631	026997	011727
Baja, aftershock	7-14 pixels	003760	004135	007890	032084	006269	001981	003286	011428	038820	009768
	14-28 pixels	005577	004506	005492	014849	006163	001969	003148	007668	027903	007742
	more than 28 pixels	002610	003912	004125	010607	002575	001401	002530	005246	016082	005321
	up to 7pixels	000441	007788	000469	009495	006391	000576	003809	005649	008340	005437
	7-14 pixels	000441	007788	000469	009495	006391	000576	003809	005649	008340	005437
Emilia 20May, Italy	14-28 pixels	002927	006743	007410	009417	006877	002893	008008	003647	009026	005048
	more than 28 pixels	003370	006656	007846	007275	007985	004487	005706	003519	010005	005173
	up to 7pixels	003746	006145	011667	010598	001651	003451	005177	001348	003513	003202
	7-14 pixels	003746	006145	011667	010598	001651	003451	005177	001348	003513	003202
	14-28 pixels	006474	013550	001721	008012	004335	006504	008226	000774	006602	003060
Emilia 29May, Italy	more than 28 pixels	006553	007192	002104	006771	003989	006465	010004	007745	003331	004162
	up to 7pixels	002875	000253	001074			002742	001766	024620		
	7-14 pixels	002875	000253	001074			002742	001766	024620		
	14-28 pixels	005979	001938	007327			005551	002637	009112		
	more than 28 pixels	004062	001636	006504			004238	001413	008524		
Abruzzo, Italy	up to 7pixels	003685	012442	000000	002454	000000	006869	003833	000000	005680	021550
	7-14 pixels	003685	012442	000000	002454	000000	006869	003833	000000	005680	021550
	14-28 pixels	007524	003754	010212	010707	000000	009486	004571	000000	006967	012206
	more than 28 pixels	007801	003622	011429	004570	004603	010147	004733	000168	005451	022440
	up to 7pixels	003995	001153	000922	002875	002536	003650	006212	000000	002699	005908
Visso, Italy	7-14 pixels	003995	001153	000922	002875	002536	003650	006212	000000	002699	005908
	14-28 pixels	007655	007680	002960	004595	004698	009206	008873	000132	006248	007376
	more than 28 pixels	005526	008872	002498	004194	005195	008005	013240	009330	004727	005435
	up to 7pixels	000000	002816	009028	006737	004071	010494	009139	001348	007143	003483
	7-14 pixels	000000	002816	009028	006737	004071	010494	009139	001348	007143	003483
Louisia, USA	14-28 pixels	019770	012054	005538	014095	005458		010525	000000	004266	008518
	more than 28 pixels	011035	006920	006952	006850	002752	016366	009272	013155	008648	004710
	up to 7pixels	008203	014286	006849	000000		004230	011958	010783	033854	
	7-14 pixels	008203	014286	006849	000000		005056	006072	007118	016512	
	14-28 pixels	008203	014286	006849	000000		004825	008942	007356	016336	
Canterbury, NZ	more than 28 pixels	009256	013846	013023	002315		003638	006276	007127	012362	
	up to 7pixels	003685	012442	000000	002454	000000	006869	003833	000000	005680	021550
	7-14 pixels	003685	012442	000000	002454	000000	006869	003833	000000	005680	021550
	14-28 pixels	007524	003754	010212	010707	000000	009486	004571	000000	006967	012206
	more than 28 pixels	007801	003622	011429	004570	004603	010147	004733	000168	005451	022440
Oldahoma 2016, USA	up to 7pixels	003995	001153	000922	002875	002536	003650	006212	000000	002699	005908
	7-14 pixels	003995	001153	000922	002875	002536	003650	006212	000000	002699	005908
	14-28 pixels	007655	007680	002960	004595	004698	009206	008873	000132	006248	007376
	more than 28 pixels	005526	008872	002498	004194	005195	008005	013240	009330	004727	005435
	up to 7pixels	000000	002816	009028	006737	004071	010494	009139	001348	007143	003483
Christchurch, NZ	7-14 pixels	000000	002816	009028	006737	004071	010494	009139	001348	007143	003483
	14-28 pixels	019770	012054	005538	014095	005458		010525	000000	004266	008518
	more than 28 pixels	011035	006920	006952	006850	002752	016366	009272	013155	008648	004710
	up to 7pixels	005732	008265	000000	002594	000000	003273	006588	000000	022045	000000
	7-14 pixels	005732	008265	000000	002594	000000	003273	006588	000000	022045	000000
Kalkoura, NZ	14-28 pixels	004122	003511	007257	008892	000326	003729	009150	011413	009540	000000
	more than 28 pixels	003785	004300	003162	004449	002916	003584	006696	008277	013275	015906
	up to 7pixels	000353	004196	006074	005579	006229	001523	005469	000000	011972	010548
	7-14 pixels	000353	004196	006074	005579	006229	001523	005469	000000	011972	010548
	14-28 pixels	002316	002984	003647	004447	007639	002603	003427	009061	007326	014084
Baja, Mexico	more than 28 pixels	002639	003853	000876	004088	004963	001378	005252	003122	004536	008413
	up to 7pixels	003483	008730	000000	006588	005306	003574	004681	000000	010325	009843
	7-14 pixels	003483	008730	000000	006588	005306	003574	004681	000000	010325	009843
	14-28 pixels	003121	003637	001857	003885	009236	002317	003158	005862	006501	014640
	more than 28 pixels	002610	003830	000879	004081	004698	001374	002513	003122	004524	008426

### **Appendix 3: Statistical analyses**

Statistical analysis of anomaly density data,  $\mu+2\sigma$  threshold, no adjustment of the co-seismic period

In the data there were 14 outliers, as assessed by examination of studentized residuals for values greater than  $\pm 3$ . The outliers are traced back in the anomaly density of Myanmar (2012), Baja California (June 2010), and New Zealand (2010) and were spread across distance zones and periods. They were kept in the data because they did not materially affect the results as assessed by a comparison of the results with and without the outliers. Anomaly density was not normally distributed in all cases, as assessed by a Shapiro-Wilk's test. Transformation of the data did not establish normality in all cases so analysis was performed on the original values. There was homogeneity of variances, as assessed by Levene's test for equality of variances ( $p > .05$ ). Mauchly's test of sphericity indicated that the assumption of sphericity was violated for the two-way interaction,  $\chi^2(9) = 113.503$ ,  $p < .001$ . Thus, the Greenhouse-Geisser correction was applied (following Maxwell and Delaney [2004],  $\epsilon < .75$ ).

No significant interaction was found between earthquake occurrence, period and distance from the earthquake,  $F(6.094, 134.060) = .162$ ,  $p = .987$ , partial  $\eta^2 = .007$ . There was no statistically significant two-way interaction,  $F(8.914, 196.116) = 1.645$ ,  $p = .106$ , partial  $\eta^2 = .070$  (between period and distance zone);  $F(3,66) = .844$ ,  $p = .475$ , partial  $\eta^2 = .037$  (between earthquake year and distance zone);  $F(2.031, 66) = .959$ ,  $p = .387$ , partial  $\eta^2 = .014$  (between earthquake year and period). There was a significant simple main effect of period,  $F(2.971, 196.116) = 11.738$ ,  $p < .001$ , partial  $\eta^2 = .151$  which was present regardless of location or earthquake occurrence. In particular, the 2-month post-earthquake period had the highest average anomaly density ( $.040 \pm .026$ ). The lowest mean anomaly density ( $.023 \pm .021$ ) was found in the co-seismic period.

Average anomaly density from all earthquake cases is shown in Figure A3.1, grouped per period, zone and earthquake year.

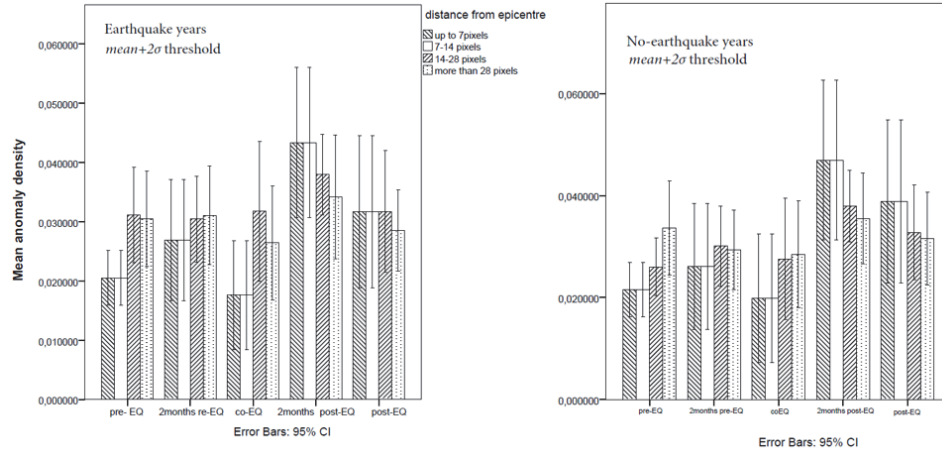


Figure A3.1. Anomaly density using a  $\mu+2\sigma$  threshold. Results from all earthquake cases grouped per time period, spatial zone and earthquake year.

#### Statistical analysis of anomaly density data, $\mu+2\sigma$ threshold, with adjustment of the co-seismic period

In the data there were 18 outliers, as assessed by examination of studentized residuals for values greater than  $\pm 3$ . The outliers are traced back in anomaly density of Myanmar (2012 and 2011), China (2010), Italy (2012) and New Zealand (2010) and were spread across distance zones and periods. They were kept in the data because they did not materially affect the results as assessed by a comparison of the results with and without the outliers. Anomaly density was not normally distributed in all cases, as assessed by a Shapiro-Wilk's test. Transformation of the data did not establish normality in all cases so analysis was performed on the original values. There was homogeneity of variances, as assessed by Levene's test for equality of variances ( $p > .05$ ). Mauchly's test of sphericity indicated that the assumption of sphericity was violated for the two-way interaction,  $\chi^2(9)=82.177$ ,  $p < .001$ . Thus, the Greenhouse-Geisser correction was applied (following Maxwell and Delaney [2004], with  $\epsilon < .75$ ).

No significant interaction was found between earthquake occurrence, period and distance from the earthquake,  $F(6.767, 121.798) = .596$ ,  $p = .753$ , partial  $\eta^2 = .032$ . There was no statistically significant two-way interaction between period and distance zone,  $F(6.832, 122.971) = .574$ ,  $p = .772$ , partial  $\eta^2 = .031$ ; none between earthquake year and distance zone,  $F(3,54) = 2.279$ ,  $p = .090$ , partial  $\eta^2 = .112$ ; and there was significant two-way interaction between earthquake year and period,  $F(2.256,54) = 6.954$ ,  $p = .001$ , partial  $\eta^2 = .114$  regardless of distance to the earthquake. There was also a significant simple main effect of period,  $F(2.971, 196.116) = 11.738$ ,  $p < .001$ , partial  $\eta^2 = .151$ . In particular, the post-earthquake period in no-earthquake years had the highest average anomaly density ( $.043 \pm .026$ ). The lowest mean

anomaly density ( $.019 \pm .018$ ) was found in the co-seismic period of no-earthquake years.

Average anomaly density from all earthquake cases is shown in Figure A3.2, grouped per period, zone and earthquake year.

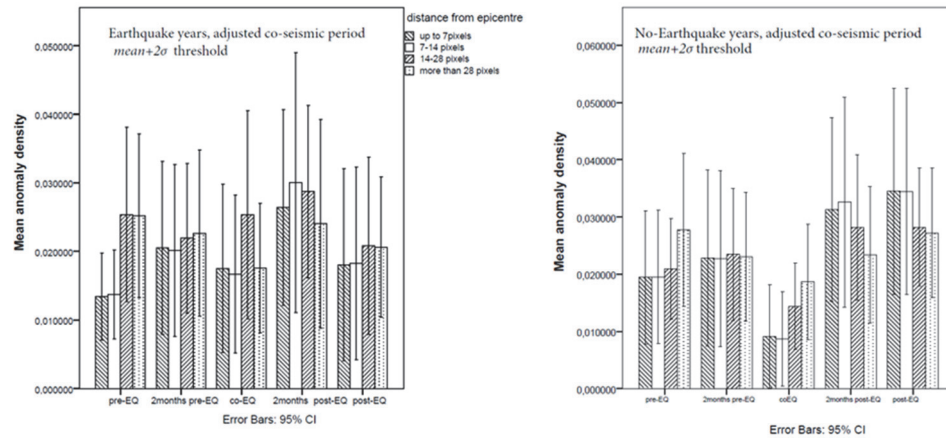


Figure A3.2. Anomaly density using a  $\mu+2\sigma$  threshold with co-seismic period adjustment for long seismic sequences. Results from all earthquake cases grouped per time period, spatial zone and earthquake year.

#### Statistical analysis of anomaly density data, $\mu+3\sigma$ threshold, no adjustment of the co-seismic period

In the data there were 14 outliers, as assessed by examination of studentized residuals for values greater than  $\pm 3$ . The outliers are traced back in anomaly density of Myanmar (2012 and 2011), Baja California (June 2010), Van (2012) and Ahar (2011) and were spread across distance zones and periods. They were kept in the data because they did not materially affect the results as assessed by a comparison of the results with and without the outliers. Anomaly density was not normally distributed in all cases, as assessed by a Shapiro-Wilk's test. Transformation of the data did not establish normality in all cases so analysis was performed on the original values. There was homogeneity of variances, as assessed by Levene's test for equality of variances ( $p > .05$ ) except for the case of the pre-earthquake period of non-earthquake years ( $p = .028$ ). Mauchly's test of sphericity indicated that the assumption of sphericity was violated for the two-way interaction,  $\chi^2(9) = 188.861$ ,  $p < .001$ . Thus, the Greenhouse-Geisser correction was applied (following Maxwell and Delaney [2004], with  $\epsilon < .75$ ).

No significant interaction was found between earthquake occurrence, period and distance from the earthquake,  $F(5.210, 116.355) = .148$ ,  $p = .983$ , partial  $\eta^2 = .007$ . There was no statistically significant two-way interaction,  $F(8.905,$



198.888) = 1.224,  $p = .282$ , partial  $\eta^2 = .052$  (between period and distance zone);  $F(3,67) = .008$ ,  $p = .999$ , partial  $\eta^2 < .001$  (between earthquake year and distance zone);  $F(1.737, 67) = .934$ ,  $p = .384$ , partial  $\eta^2 = .014$  (between earthquake year and period). There was a significant simple main effect of period,  $F(2.968, 116.355) = 9.927$ ,  $p < .001$ , partial  $\eta^2 = .129$  which was present regardless of location or earthquake occurrence. In particular, the 2-month post-earthquake period had the highest average anomaly density (.0095±.0073). The lowest mean anomaly density (.0048±.0080) was found in the co-seismic period.

Average anomaly density from all earthquake cases is shown in Figure A3.3, grouped per period, zone and earthquake year.

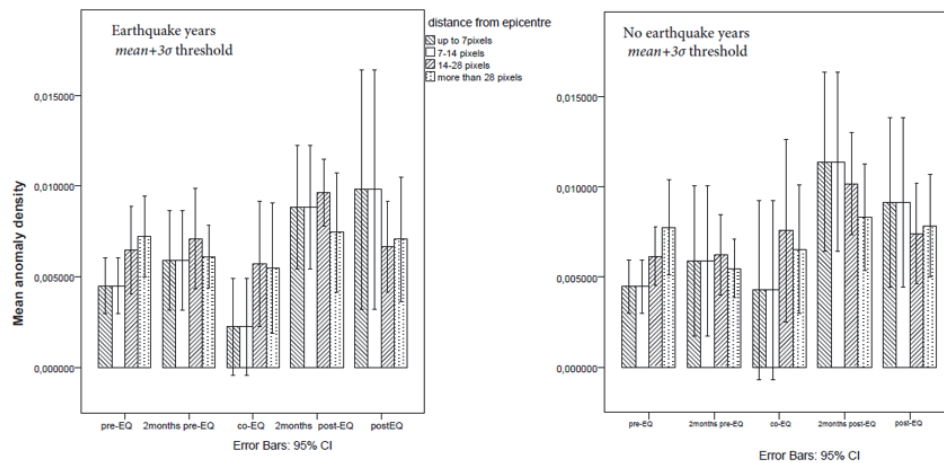


Figure A3.3. Anomaly density using a  $\mu+3\sigma$  threshold. Results from all earthquake cases grouped per time period, spatial zone and earthquake year.

#### Statistical analysis of anomaly density data, $\mu+3\sigma$ threshold, with adjustment of the co-seismic period

In the data there were 13 outliers, as assessed by examination of studentized residuals for values greater than  $\pm 3$ . The outliers are traced back in anomaly density of Myanmar (2012 and 2011) and Baja California (June 2010) and were spread across distance zones and periods. They were kept in the data because they did not materially affect the results as assessed by a comparison of the results with and without the outliers. Anomaly density was not normally distributed in all cases, as assessed by a Shapiro-Wilk's test. Transformation of the data did not establish normality in all cases so analysis was performed on the original values. There was homogeneity of variances, as assessed by Levene's test for equality of variances ( $p > .05$ ) except for the case of the pre-earthquake period of non-earthquake years ( $p = .013$ ). Mauchly's test of sphericity indicated that the assumption of sphericity was violated for the two-

way interaction,  $\chi^2(9) = 49.772$ ,  $p < .001$ . Thus, the Greenhouse-Geisser correction was applied (following Maxwell and Delaney [2004], with  $\varepsilon < .75$ ).

No significant interaction was found between earthquake occurrence, period and distance from the earthquake,  $F(8.507, 155.958) = .422$ ,  $p = .915$ , partial  $\eta^2 = .022$ . There was no statistically significant two-way interaction between period and distance zone,  $F(7.604, 139.413) = .788$ ,  $p = .608$ , partial  $\eta^2 = .041$ ; neither between earthquake year and distance zone  $F(3,55) = .626$ ,  $p = .601$ , partial  $\eta^2 = .033$ . There was a significant two-way interaction between earthquake year and period,  $F(2.836, 55) = 6.485$ ,  $p < .001$ , partial  $\eta^2 = .105$ . There was a significant simple main effect of period,  $F(2.535, 139.413) = 6.631$ ,  $p < .001$ , partial  $\eta^2 = .108$  which was present regardless of location or earthquake occurrence; and a simple main effect of earthquake year,  $F(1, 55) = 14.914$ ,  $p < .001$ , partial  $\eta^2 = .213$ . In particular, the post-earthquake period in no-earthquake years had the highest average anomaly density ( $.0097 \pm .0062$ ). The lowest mean anomaly density ( $.0036 \pm .0045$ ) was found in the co-seismic period of no-earthquake years.

Average anomaly density from all earthquake cases is shown in Figure A3.4, grouped per period, zone and earthquake year.

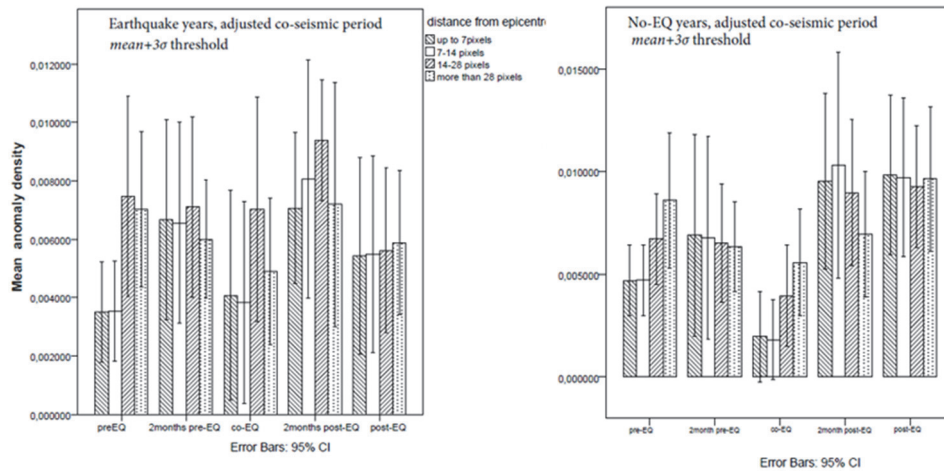


Figure A3.4. Anomaly density using a  $\mu+3\sigma$  threshold after a co-seismic period adjustment for long seismic sequences. Results from all earthquake cases grouped per time period, spatial zone and earthquake year.



*Netherlands Research School for the Socio-  
Economic and Natural Sciences of the Environment*

# D I P L O M A

## *For specialised PhD training*

The Netherlands Research School for the Socio-Economic  
and Natural Sciences of the Environment (SENSE) declares  
that

***Efthymia Pavlidou***

born on 27 February 1980 in Alexandroupolis, Greece

has successfully fulfilled all requirements of the Educational  
Programme of SENSE.

Enschede, 27 September 2018

the Chairman of the SENSE board

the SENSE Director of Education

Prof. dr. Huub Rijnaarts

Dr. Ad van Dommelen

*The SENSE Research School has been accredited by the Royal Netherlands Academy of Arts and  
Sciences (KNAW)*



K O N I N K L I J K E N E D E R L A N D S E  
A K A D E M I E V A N W E T E N S C H A P P E N





The SENSE Research School declares that **Efthymia Pavlidou** has successfully fulfilled all requirements of the Educational PhD Programme of SENSE with a work load of 62.7 EC, including the following activities:

#### SENSE PhD Courses

- o Spatiotemporal analysis and big data processing using free and open source software (2014)
- o Bayesian statistics (2016)
- o Uncertainty propagation in modeling (2016)
- o Environmental research in context (2017)
- o Research in context activity: 'Organizer of the first SENSE Alumni meetup' (2017)

#### Selection of other PhD Courses

- o Academic Leadership, University of Twente (2013)
- o Negotiating, University of Twente (2014)
- o Earth observation summer school, European Space Agency/ESRIN, Italy (2014)
- o Summer school on data mining and machine learning, INIT/AERFAI, Spain (2015)
- o Time Management, University of Twente (2015)

#### External workshops and seminars

- o Land surface temperature applications Workshop, EUMETSAT, United Kingdom (2015)
- o Quantifying and Communicating Uncertainty, ECMWF, United Kingdom (2015)
- o Workshops on Statistical Seismology, StatSei9, Germany (2015) and StatSei10, New Zealand (2017)

### Selection of Management and Didactic Skills Training

- o Teaching in the MSc courses 'GIS Science and Earth Observation', 'Research Skills', 'Geotechnical modelling' and 'Modelling Natural Resource degradation' (2015-2016)
- o Workshop Transferable Skills, University of Twente (2016)
- o Board member of Promovendi Network University of Twente (P-Nut) and Twente representative to Promovendi Network Netherlands, PNN (2015-2016)
- o University council representative, University of Twente (2016-2018)
- o UTQ (BKO) Teaching Qualification, University of Twente (2017)

### International Oral Presentation

- o *Time series of gridded CMSAF TET*. EUMETSAT workshop, 20-24 March 2017, Krakow, Poland

SENSE Coordinator PhD Education



Dr. Peter Vermeulen