Elsevier Editorial System(tm) for Remote Sensing of Environment

Manuscript Draft

Manuscript Number:

Title: EARTHQUAKE DAMAGE MAPPING: AN OVERALL ASSESSMENT OF GROUND SURVEYS AND VHR IMAGE CHANGE DETECTION AFTER L'AQUILA 2009 EARTHQUAKE

Article Type: Research Paper

Keywords: earthquake damage mapping, change detection, classification

Corresponding Author: Professor Nazzareno Pierdicca, M.D.

Corresponding Author's Institution: Sapienza University of Rome

First Author: Nazzareno Pierdicca, M.D.

Order of Authors: Nazzareno Pierdicca, M.D.; Roberta Anniballe, Ph.D; Fabrizio Noto, M.D.; Tanya Scalia; Christian Bignami, Ph.D; Salvatore Stramondo; Marco Chini

Highlights (for review)

- Earthquake damage maps at the scale of single building from VHR optical images.
- Validated against an extensive ground survey in a densely urbanized historical town
- Textural and color features extracted at object scale most effective.
- Support Vector Machine supervised classification outperforms Bayesian classifier
- Satellite provides an effective prompt and cheap support to disaster management

EARTHQUAKE DAMAGE MAPPING: AN OVERALL ASSESSMENT OF GROUND SURVEYS AND VHR IMAGE CHANGE DETECTION AFTER L'AQUILA 2009 EARTHQUAKE

Nazzareno Pierdicca[a], Roberta Anniballe[a], Fabrizio Noto[b], Tanya Scalia[a], Christian Bignami[c], Salvatore Stramondo[c], Marco Chini[d]

[a] *Department of Electronic Engineering, Sapienza University of Rome. Via Eudossiana 18 - 00184 Rome, Italy*

[b] *METIS s.r.l., Via Emanuele Filiberto 207 - 00185 Rome, Italy*

[c] *Istituto Nazionale di Geofisica e Vulcanologia, Via di Vigna Murata 605 - 00143 Rome, Italy*

[d] *Luxembourg Institute of Science and Technology (LIST), Environmental Research and Innovation Department (ERIN), Belvaux, Luxembourg*

ABSTRACT

Earth Observation (EO) data, from both Synthetic Aperture Radar (SAR) and optical sensors, are generally used to map mostly affected urban areas after an earthquake using change detection techniques applied at pixel scale. However, Civil Protection Services require damage assessment of each building according to a well-established scale to manage rescue operations and to estimate the economic losses.

Considering the earthquake that hit L'Aquila city (Italy) on April 6, 2009, this work assess the feasibility of producing damage maps at the scale of single building from Very High Resolution (VHR) optical images collected before and after the seismic event. We

considered the European Macroseismic Scale 1998 (EMS-98) and assessed the possibility to discriminate between collapsed or heavy damaged buildings (damage grade *DG* equal to 5 in the EMS-98 scale) and less damaged or undamaged buildings (*DG* < 5 in the EMS-98).

The proposed approach relies on a pre-existing urban map to identify image objects corresponding to buildings where the change analysis is performed. The latter is carried out to many different parameters with the objective of assessing their effectiveness in singling out changes associated to the building collapse. Features describing texture and color changes, as well statistical similarity and correlation descriptors, such as the Kullbach Leibler divergence and the Mutual Information, were included in our analysis. Two supervised classification approaches, respectively, based on the use of the Bayesian Maximum A Posteriori (MAP) criterion and on Support Vector Machines (SVM), were compared. In our experiment, we considered the whole L'Aquila historical centre comparing classification results with the ground survey performed by the Istituto Nazionale di Geofisica e Vulcanologia (INGV).

The work represents one of the first attempt to detect damage at the scale of single building, validated against an extensive ground survey. It addresses methodological aspects, highlighting the potential of textural features computed at object scale and SVMs, and discuss potential and limitations of EO in this field compared to ground surveys.

Keywords: earthquake damage mapping, change detection, classification

1. Introduction

The use of Earth Observation (EO) in the domain of natural hazards and disaster management is becoming increasingly popular. This is due partially to the increased awareness of environmental issues, the need to face those issues on a global scale perspective, but also to the improvement of satellite technologies and the ability to deliver high quality imagery through fast telecommunication links, different media and disseminate them through the internet. As technology is enhanced, demand and expectations increase for near-real-time image delivery to emergency services in the event of a natural disaster (Joyce et al. 1999). During a seismic event, in particular, it is fundamental to obtain a fast and reliable map of the damage of urban areas to manage civil protection interventions. Moreover, the identification of the destruction caused by earthquakes provides seismology and earthquake engineers with informative and valuable data, experiences and lessons in the long term, and raise some important scientific problems (see the International Association of Seismology and Physics of the Earth's Interior, IASPEI website at http://iaspei.org/commissions). An accurate survey of damage is also important to assess the economic losses, and manage and share the resources to be allocated during the reconstruction phase.

Satellite remote sensing data can provide valuable pieces of information on this regard, thanks to their capability of delivering an instantaneous synoptic view of the scene, especially if the seismic event is located in remote regions, or the main communication systems are damaged. Many works exist in the literature on this topic, considering both optical data and radar data, which however put in evidence some limitations of the nadir looking view, of the achievable level of details and response time, and the criticality of image radiometric and geometric corrections (Dell'Acqua and Gamba, 2012). The visual interpretation of the images is the approach followed in many cases, especially for an operational and rapid release of the damage

extension map (Yamazaki et al., 2005). Many papers evaluated change detection approaches to estimate damage within large areas (e.g., city blocks). They quantified not only the extension of the affected area, but also the level of damage, for instance correlating the collapse ratio (percentage of collapsed buildings in an area) measured on ground with some change parameters derived from two images, taken before and after the earthquake (Stramondo et al., 2006). The case of radar imagery presents additional challenges, as for instance discussed in (Brunner et al., 2010). Nowadays, remotely sensed images at Very High Resolution (VHR) may in principle enable production of earthquake damage maps at single-building scale as done for instance in Chini et al. (2009). The complexity of the image forming mechanisms within urban settlements, especially of radar images, makes the interpretation and analysis of VHR images still a challenging task. Discrimination of lower grade of damage is in particular extremely difficult using nadir looking sensors.

Automatic algorithms to detect the damage are being developed although, as matter of fact, these works focus very often on specific test cases and sort of canonical situations (Marin et al., 2015). In order to make the delivered product suitable for the user community (e.g., civil protection) it is important to assess its quality on a large area and in different and challenging situations. Moreover, the assessment shall be directly compared to the type of data the final user adopts when carrying out his/her operational tasks. This kind of assessment can be hardly found in the literature, especially when the focus is on the development of sophisticated and advanced algorithms.

In the APhoRISM (Advanced Procedures for volcanic and Seismic Monitoring) project (Devanthéry et al., 2016), funded by the European Union under the EC-FP7 call, methodologies to exploit remote sensing for earthquake risk management, and in particular for delivering a

prompt and accurate damage assessment map, were developed. The project aimed at integrating remote sensed observations with other direct conventional observations (e.g., accelerometer networks and derived shake maps) and with prior information on building vulnerability and soil instability to provide a better and reliable damage assessment product to the final user.

In this paper, we present the work performed using the remotely sensed images alone, and in particular VHR optical images, to classify the damage grade of individual buildings. To encounter the need of the final user, especially civil protection authorities, the experiment was carried out on an entire town comparing the results with the ground data collected by the same user and another institution involved in the seismic risk management, thus providing a well-known and reliable reference, as well as a common definition of the product taxonomy. Specifically, the experiment was carried out considering the earthquake that hit the city of L'Aquila, Italy, in 2009. The image classification results were compared to the ground surveys carried out by the Institute of Geophysics and Volcanology and the Italian Department of Civil Protection. The work provides several hints into the problem, on the way to process the images, to evaluate the classification performances and understand advantages and limitations of remote sensing in this field. Moreover, it points out the possible ambiguities related to the ground surveys, the protocols followed for their collection and interpretation, which raises some difficulties when the surveys are used as reference for assessing image classifications accuracy.

The paper is organized as follows. Section 2 describes the test case and the dataset used for the study, both remotely sensed and collected on ground, whilst section 3 provides details on the exploited methodologies regarding feature extraction, and classification methods. Section 4 presents the results of the feature selection and the validation of the resulting damage map. Finally, section 5 draws the main conclusions.

2. The test case and available data

2.1 L'Aquila 2009 earthquake

On April 6, 2009 at 1:32 GMT, an earthquake hit L'Aquila city, in Central Italy. The mainshock was rated 6.3 on the moment magnitude (Mw) scale; the epicentre was located near L'Aquila, at a depth of about 9 km, and was followed in the next week by seven aftershocks with Mw > 5. L'Aquila and the surrounding villages suffered the highest damage. The earthquake was felt throughout central Italy; 309 people died, making this event the deadliest earthquake that hit Italy since the 1980 Irpinia's one. The earthquake caused damage to thousands of buildings in the medieval city centre of L'Aquila. Some buildings fully collapsed, thus leading to a sparse damage distribution within a high-density urban area, a very challenging situation for satellite damage assessment.

2.2 Ground truth data

The ground truth data available for the selected test site comes from the survey performed by Istituto Nazionale di Geofisica e Vulcanologia (INGV) macroiseismic team (QUEST - QUick Earthquake Survey Team, http://quest.ingv.it).

After the catastrophic event, several groups of INGV researchers, in one week of fieldwork, collected information related to the type of building and its vulnerability class, and the suffered damage, according to the European Macroseismic Scale 1998 (EMS-98) (Grünthal, 1998). The damage grade ranges from 0 to 5, i.e., from no damage to a completely collapsed building. It is worth noticing that the inventory data were collected by a visual inspection, looking from outside the buildings, because the INGV teams were not allowed to enter the edifices for safety reasons. More than 1600 buildings were surveyed in the central area of L'Aquila, and a georeferenced vector file was produced to map the collected data in a Geographic Information System (GIS)

(Tertulliani et al., 2011, Tertulliani et al., 2012). The resulting map is shown in Figure 1. The damage distribution (see Figure 2) is mostly concentrated on damage grade D=3, while collapsed buildings (*D*=5) represent 4.4% of the surveyed buildings.

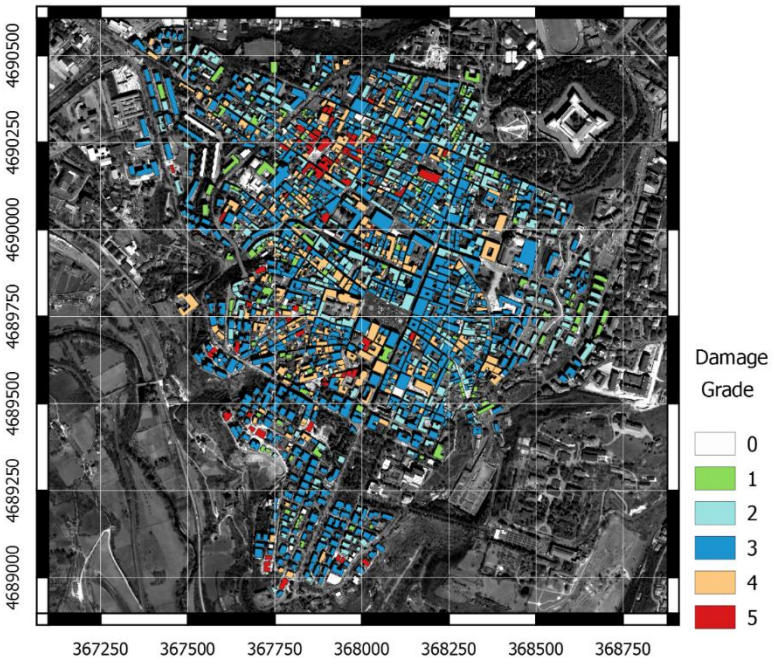


Figure 1. Damage distribution of L'Aquila city centre. The polygons of surveyed buildings, with color representing the damage grade, are superimposed on a very high resolution panchromatic image acquired by the QuickBird satellite.

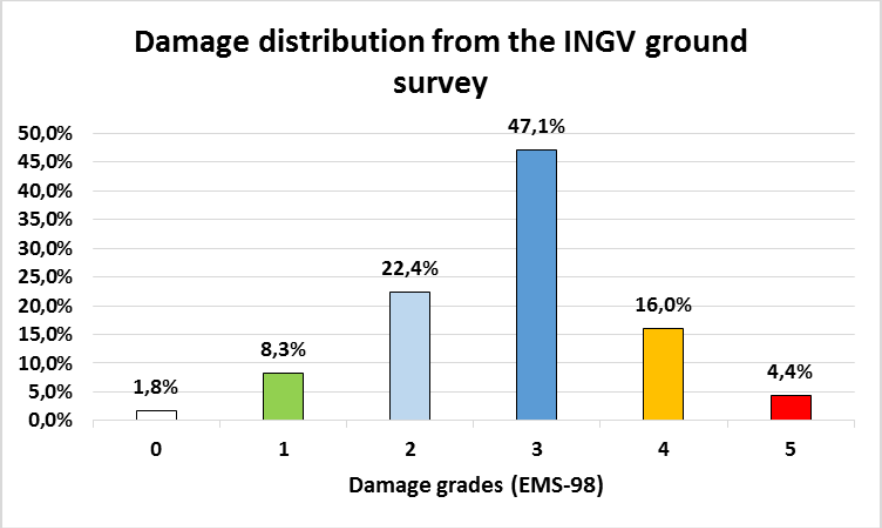


Figure 2: Frequency distribution of damage grades for L'Aquila city centre according to the INGV survey.

2.3 Earth Observation data

We carried out the present work using two VHR optical images taken from the QuickBird satellite. The image before the earthquake is dated September 4, 2006, while the post-seismic image was collected on April 8, 2009, only 2 days after the catastrophic event. Each acquisition is composed of a panchromatic (PAN) and a multispectral (MS) image. The latter is collected in four bands in the blue, green, red and near-infrared wavelength regions. Nominally, at nadir, the spatial resolution of the PAN image is 0.6 m, while the MS image has a 2.44 m resolution (DigitalGlobe 2014). In Table 1 the main characteristics of the images are reported.

VHR Optical Dataset			
Date of Acquisition	**Acquisition Mode**	**Looking Angle**	**Resolution**
04/09/2006	PAN + MS	-3.7° in- track -10.3° cross-track 10.9° off-nadir	0.6m PAN 2.44m MS
08/04/2009	PAN + MS	2.8° in- track 3.9° cross-track 4.8° off-nadir	0.6m PAN 2.44m MS

155 **Table** 1**. Summary of the optical VHR dataset from the QuickBird satellite**

156 ## 3. Background and methodology

157 To assess the feasibility of a damage product at the scale of single building from a pair of VHR
158 optical data, we implemented a change detection approach that works at object scale. In the
159 contest of the earthquake damage assessment, segmenting the pre-event image into objects
160 corresponding to a building allows the change analysis to be focused on the objects of interest,
161 avoiding false alarms due for example to vegetation changes and temporary objects (e.g.,
162 vehicles). In this work, the segmentation of the scene was carried out using a pre-existing
163 building map provided as GIS layer, namely the "Carta Tecnica Regionale" (CTR) of the
164 Abruzzo Region. This approach can be easily reproduced in other cases, as maps of urban areas
165 are available in most towns, otherwise objects corresponding to buildings can be identified from
166 pre-seismic images exploiting segmentation algorithms, as described for example in Chini et al.
167 (2009). Within each building footprint (i.e., an image object) a number of features, potentially
168 capable of detecting changes associated with the building collapse, were computed from pre- and
169 post-seismic images, both panchromatic and multispectral. Different classification approaches
170 were tested and compared.

171 ### *3.1 Image pre-processing steps*

172 Panchromatic images were orthorectified exploiting the Rational Polynomial Coefficients (RPC)
173 provided with the data and a Digital Elevation Model (DEM) of the terrain obtained from a
174 LiDAR overflight of the study area. Pan-sharpened pre- and post-event images were produced
175 from PAN and MS imagery using the Gram-Schmidt Pan Sharpening method (Brower and
176 Laben, 2000) implemented into the IDL/ENVI© software. Furthermore, the latter data were

177 orthorectified exploiting the RPCs of PAN images and the LiDAR DEM according to the same
178 procedure used for PAN data.
179 A further registration between orthorectified images was required in order to achieve a better
180 alignment of the images at street level and with the CTR map. Namely, a rigid shift of 0.6 meters
181 in the North-South direction and of -4.8 meters in West-East direction was applied to the pre-
182 event image. As for the post-event image, a shift of -2.4 meters in the North-South direction and
183 a shift of 7.8 meters toward the East were necessary.
184 As the orthorectification was done with respect to the surface level (i.e. street level), in order to
185 better superimpose the building footprints reported in the GIS layer to the actual image pixels
186 associated to the roof we compensated the parallax error, i.e. the apparent displacement of
187 building roofs with respect to their bases. The parallax displacement occurs along the sensor
188 Line Of Sight, over a distance that depends on building height h and off-nadir view angle θ,
189 according to the equation:

$$\Delta s = h \tan\theta$$

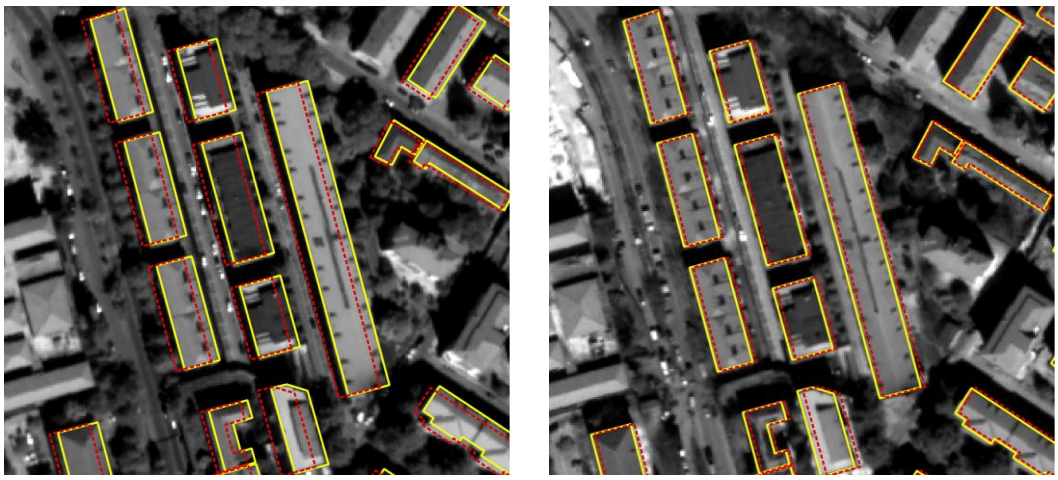
190 Considering a Cartesian reference frame where the *x* axis is oriented towards the geographic
191 East, the *y* axis towards the geographic North and the *z* axis along the vertical upward direction,
192 the parallax displacement components along the *x* and *y* axes can be found as follows:

$$\Delta s_X = \Delta s \sin\varphi$$

$$\Delta s_Y = \Delta s \cos\varphi,$$

193 where φ is the sensor azimuth measured clockwise from the North. Based on the building height
194 reported in the CTR of Regione Abruzzo, building polygons were shifted, generating two layers
195 matching pre-and post-event images, respectively (note that the off-nadir angle of the two images
196 are different).

197 The results of this fine registration step can be appreciated in Figure 3. Figure 3(a) and Figure
198 3(b) show respectively pre- and post-event PAN images orthorectified and co-registered.
199 Building footprints shifted for compensating the parallax error are reported in yellow, while red
200 dashed polygons represent buildings footprints in their original position. The improvement of the
201 position of the polygons with respect to the roofs is particularly evident in the pre-event image
202 (Figure 3(a)) that was collected with a greater off-nadir angle (see Table 1). This fine registration
203 plays a significant role for a better computation of building features and detection of changes, as
204 detailed in what follows.
205 Before extracting change detection features, a histogram matching was performed to
206 radiometrically compensate for the different season and illumination conditions of the pre- and
207 post-seismic acquisitions. A linear transformation was applied to each band of the post-event
208 image in order to get a data distribution with the same median and interquartile range of the
209 corresponding pre-event data. We carried out this step considering only urban pixels that has
210 been identified using the GIS polygons shifted for compensating parallax error, as described
211 before.



212

213 **(a)** **(b)**

Figure 3: Pre- (a) and post-event (b) PAN images with superimposed building footprints in their original position (red dashed polygons) and shifted for compensating the parallax error (yellow polygons).

3.2 Change Detection Features

Many different change parameters have been introduced in the literature to detect urban damage from optical data (Dong and Shan, 2013; Stramondo et al., 2006). In this work, we performed an extensive investigation with the objective of identifying the most suitable set for this application. In particular, we have considered a set of 13 features that can be grouped in four categories: standard change detection metrics, change indicators from information theory, features describing texture and colour changes (see Table 2). In the following, we briefly introduce each of them.

Note that in a change detection procedure the features can be computed within a moving window (usually a square window with odd number of pixels) scanning the entire image, or within predefined regions recognized to be homogenous and pertaining to a given target (or object). The latter is generally referred to as the Object Based Image Analysis (OBIA) approach (Chen et al., 2012; Chini et al. 2009; Chesnel et al., 2007; Rastiveis et al, 2013, Gusella et al., 2005). In this work, we adopted the second approach, with image segmentation based on the CRT map from Regione Abruzzo, as previously described. Table 2 provides the complete list of change detection features considered in the work, whilst their definition is illustrated in what follows.

Group	Name	Description
Information theory features	MI_{pan}	Mutual Information from PAN images
	KLD_{pan}	Kullback-Leibler Divergence from PAN images
	MI_{psh}	Mutual Information from PSH images
	KLD_{psh}	Kullback-Leibler Divergence from PSH images
features (from PAN imag	Δcon	Change in Contrast ($\Delta con = con_{post} - con_{pre}$)
	Δcor	Change in Correlation ($\Delta cor = cor_{post} - cor_{pre}$)

	Δene	Change in Energy ($\Delta ene = ene_{post} - ene_{pre}$)
	Δhom	Change in Homogeneity ($\Delta hom = hom_{post} - hom_{pre}$)
	Δent	Change in Entropy ($\Delta ene = ent_{post} - ent_{pre}$)
Colour features (from PSH images)	Δhue	Change in Hue ($\Delta hue = hue_{post} - hue_{pre}$)
	Δsat	Change in saturation ($\Delta sat = sat_{post} - sat_{pre}$)
	Δlum	Value component difference ($\Delta lum = lum_{post} - lum_{pre}$)
standard change detection features	Δint_pan	Change in the intensity from PAN images ($\Delta int\_pan = int\_pan_{post} - int\_pan_{pre}$)

Table 2. List of the change parameters investigated in this study

Image Differencing

The simplest and more common methods for detecting changes from multi-temporal satellite imagery are Image Differencing and Image Rationing. We considered the former, i.e. the difference of the two panchromatic images (pre- and post-event) computed within each object (i.e. the building footprints) after histogram matching.

Kullback-Leibler Divergence (KL Divergence)

The KL Divergence (Kullback and Leibler, 1951), also called Relative Entropy, is a measure of the difference between two probability density functions (pdf) and then it can be used for assessing the similarity of pixel intensity distributions of two objects.

Given two n-dimensional random vectors X and Y with pdfs $f_X(\mathbf{x})$ and $f_Y(\mathbf{x})$, respectively, the KL Divergence between the two distributions or, equivalently, the relative entropy of X with respect to Y, is given by:

$$KL(\boldsymbol{X}, \boldsymbol{Y}) = \int f_{\boldsymbol{X}}(\mathbf{x}) \log\left(\frac{f_{\boldsymbol{X}}(\mathbf{x})}{f_{\boldsymbol{Y}}(\mathbf{x})}\right) d\mathbf{x} \quad (1)$$

The KL Divergence is always non-negative, and equals zero only if $f_X(\mathbf{x}) = f_Y(\mathbf{x})$, but it is not a true metric because it is non-symmetric. A symmetric version, called KL Distance (KLD), can be defined as:

$$KLD(\boldsymbol{X},\boldsymbol{Y}) = KL(\boldsymbol{X},\boldsymbol{Y}) + KL(\boldsymbol{Y},\boldsymbol{X}) \tag{2}$$

For two multivariate normal distributions, with mean vectors $\boldsymbol{\mu}_X$, $\boldsymbol{\mu}_Y \in \mathbb{R}^n$ and covariance matrices $\mathbf{C}_X$, $\mathbf{C}_Y \in \mathbb{R}^{nxn}$, the KL-Distance, KLD($\mathbf{X}$, $\mathbf{Y}$), is proven to be:

$$\begin{aligned} KLD(\boldsymbol{X},\boldsymbol{Y}) = \frac{1}{2}[tr(\boldsymbol{C}_X^{-1}\boldsymbol{C}_Y) + tr(\boldsymbol{C}_Y^{-1}\boldsymbol{C}_X)] \\ + \frac{1}{2}[(\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y)^T\boldsymbol{C}_X^{-1}(\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y) + (\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y)^T\boldsymbol{C}_Y^{-1}(\boldsymbol{\mu}_X - \boldsymbol{\mu}_Y)] - n \end{aligned} \tag{3}$$

which, in the case of two univariate Gaussian distributions with means μx, μ_Y and variances σ_X^2, σ_Y^2, simplifies to:

$$KLD(X,Y) = \frac{1}{2}\left(\frac{\sigma_X^2}{\sigma_Y^2} + \frac{\sigma_Y^2}{\sigma_X^2}\right) + \frac{1}{2}\left[\frac{(\mu_X - \mu_Y)^2}{\sigma_X^2} - \frac{(\mu_X - \mu_Y)^2}{\sigma_Y^2}\right] - 1 \tag{4}$$

We exploited equations (3) and (4) for extracting the KL Divergence from PSH and PAN images, respectively.

Mutual information (MI)

The Mutual Information is a commutative measure of the difference between the joint probability distribution $f_{X,Y}(\mathbf{x},\mathbf{y})$ and the marginal probability distribution $f_X(\mathbf{x})$ and $f_Y(\mathbf{x})$, of the random variables X and Y. It measures how much knowing one of the two variables reduces the uncertainty about the other and it is defined as (Erten et al., 2012):

$$MI(\boldsymbol{X},\boldsymbol{Y}) = \int_{Y}\int_{X} \log\left(\frac{f_{X,Y}(\mathbf{x},\mathbf{y})}{f_X(\mathbf{x})f_Y(\mathbf{x})}\right) f_{X,Y}(\mathbf{x},\mathbf{y})d\mathbf{x}d\mathbf{y} \quad (5)$$

The MI between two multivariate Gaussian distributions with mean vectors $\boldsymbol{\mu}_X$, $\boldsymbol{\mu}_Y \in \mathbb{R}^n$, covariance matrices $\mathbf{C}_X$, $\mathbf{C}_Y \in \mathbb{R}^{nxn}$ and cross-covariance matrix $\mathbf{C}$, is given by:

$$MI(\boldsymbol{X},\boldsymbol{Y}) = -\frac{1}{2}\log\left(\frac{det(\boldsymbol{C})}{det(\boldsymbol{C}_X)det(\boldsymbol{C}_Y)}\right) \quad (6)$$

When two univariate Gaussian distributions with variances σ_X^2, σ_Y^2, and covariance, σ_{XY}^2 are considered, the following expression is obtained for the MI:

$$MI(X,Y) = -\frac{1}{2}\log\left(\frac{\sigma_X^2\sigma_Y^2 - \sigma_{XY}^2}{\sigma_X^2\sigma_Y^2}\right) = -\frac{1}{2}\log(1-\rho_{XY}^2) \quad (7)$$

where ρ_{XY} is the correlation coefficient. Equation (6) and Equation (7) were used to evaluate the MI from PSH and PAN images, respectively. It can be noted that, in the univariate case (**Error! Reference source not found.**(7)), the MI reduces to a function of the correlation coefficient.

Change detection metrics based on textural characteristics

The exploitation of the changes in the textural properties of the image of the buildings was also investigated. Indeed, a difference in the spatial arrangement of pixel intensity is expected as a consequence of a collapse. In particular, features based on the second order statistics were considered, following the approach proposed by Haralick (Haralick et al., 1973). Namely, the grey level co-occurrence matrix (GLCM) was computed first. The GLCM described in the original Haralick paper is a symmetrical matrix whose element $G(i,j)$ counts how many times two pixels with grey levels i and j occur in the image separated by a given distance d along a given direction θ. Here the GLCM was computed at object scale (Bignami et al., 2011), i.e., considering pixels within each building footprint. Considering four angular directions (0°, 45°,

277 90°, 135°) and a fixed distance d=1, four distinct GLCMs were computed and afterward summed
278 for obtaining a rotational invariant GLCM. From the resulting GLCM, five textural features *Tf*
279 were derived, that is Contrast, Correlation, Energy, Homogeneity, and Entropy (Haralick et al.,
280 1973).
281 Each *Tf* was extracted from pre- and post-event images, obtaining Tf_{pre} and Tf_{post}, from which the
282 change indicator $\Delta Tf = Tf_{post} - Tf_{pre}$ was computed.

283 *Metrics describing changes in the colour space*
284 As additional source of information for discriminating between damaged and undamaged
285 buildings, changes in the Hue (H), Saturation (S), Value (V) colour space were taken into
286 account, considering for instance that rubbles are expected to have low saturation as opposed to
287 changes due to building restoration (e.g., painting or tile roof restoration). The RGB value of a
288 pixel was first transformed into the HSV space using a method suggested in (Smith, 1978). For
289 each HSV colour channel, the mean value within the building footprint was calculated. The
290 difference between the mean values of each HSV component computed respectively from the
291 post- and pre-event image were then evaluated in order to obtain the change metric.

292 *3.3 Damage classification approaches*
293 In order to generate a damage classification product to be released to the final user, it is
294 important to rely on a damage scale that is recognized as a standard. To this aim, we considered
295 the EMS-98 scale (Grünthal, 1998) and assessed the potential to discriminate between collapsed
296 or heavy damaged (D =5 in the EMS-98 scale) buildings and less damaged or undamaged
297 buildings (D < 5 in the EMS98 scale). The investigation was carried out considering two
298 supervised learning algorithms, namely, the Bayesian Maximum A Posteriori (MAP) criterion
299 and the Support Vector Machine (SVM). Ours experiments were carried out in Matlab©

environment, using the LibSVM package (Chang and Lin, 2011) to implement the SVM algorithm. Training and test sets were provided by the INGV ground survey. Since the dataset is very unbalanced (the number of instances for the damaged buildings class is much lower than the number of samples belonging to the class of buildings with damage grade less than 5), it was not possible to split data into training and test samples. Thus, we exploited a k-fold cross-validation (CV) procedure to assess the classification performances. The dataset was split into k=10 disjoint subsets of approximately equal size, preserving the original class proportions: in turn, each of the k subsets was used for testing the classifier trained on the remaining k-1 subsets. For each iteration, we derived the confusion matrix (CM) referred to the current test set and finally, by summing up all the k CMs, we generated a global CM to assess the classification performances on the whole dataset. Classification accuracy measures have to be considered with care in case of unbalanced sets and just one parameter could be not enough for assessing the classification results. Here, several parameters were derived: overall accuracy, sensitivity, precision, and Cohen's Kappa (Congalton and Green, 1999).

Non parametric MAP classification

Given the vector of the observed change features $\mathbf{x} = \{x_1, x_2, ..., x_d\}$, following the implementation of the Bayesian MAP criterion, we assign the corresponding building to the class y_c (c=1,2) with the highest a-posteriori probability $P(y_c \mid \mathbf{x})$.

We estimate the posterior probabilities assuming that both classes are equally likely a-priori. As for the class-conditional pdfs, $p(\mathbf{x}|y_c)$, also known as class likelihood functions, they are estimated from the training set through the non-parametric approach known as Parzen window method, assuming, moreover, the class-conditional independence of all the features, i.e., the Naïve hypothesis. In detail, given a set of n_c training samples from class c, $\mathbf{X}^{(c)} = \{ \mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{nc}\}$,

where $\mathbf{x}_i \in \mathbb{R}^d$, $i = 1,\ldots,n_c$, and considering a Gaussian kernel, the Parzen window method estimates the distribution $p(\mathbf{x}|y_c)$ by the following (Parzen, 1962):

$$p(\mathbf{x}|y_c) = \prod_{j=1}^{d} p(\mathrm{x}_j|y_c) = \prod_{j=1}^{d} \frac{1}{n_c\sqrt{2\pi}h_j} \sum_{i=1}^{n_c} exp\left(-\frac{1}{2}\left(\frac{\mathrm{x}_j - X_{ij}^{(c)}}{h_j}\right)^2\right) \quad (8)$$

where $X_{ij}^{(c)}$ is the i-th observation of the j-th feature from the c-th class, and h is the so called bandwidth parameter or kernel width. Equation (8) shows that, for each feature, the pdf of the data given the class, $p(x_j|y_c)$, is estimated as sum of Gaussian kernel functions placed on each training data point.

Support Vector Machines

A Support Vector Machine or SVM (Cortes and Vapnik, 1995; Vapnik, 1999) is a machine learning algorithm that, conceptually, solves a binary classification problem mapping input data in a higher dimensional feature space. In that space input data are labeled based on which side of an optimal separating hyperplane, they fall. The hyperplane is constructed during the training phase.

Supposing that the training set consists of N samples $(\mathbf{x}_1, y_1)$, ..., $(\mathbf{x}_N, y_N)$, where $\mathbf{x}_i \in \mathbb{R}^d$ is the d-dimensional feature vector representing the i-th training sample, and $y_i \in \{-1;1\}$ is the class label associated with $\mathbf{x_i}$. Given a vector function Φ: $\mathbb{R}^d \rightarrow \mathbb{R}^H$ which maps the d-dimensional input vector $\mathbf{x}$ into an H-dimensional (with H>d) feature space, a SVM constructs the best separating hyperplane, $\mathbf{w}^* \cdot \Phi(\mathbf{x}) + b^* = 0$, solving the following constrained optimization problem:

$$\min_{w,b,\xi} \frac{1}{2}\mathbf{w}\cdot\mathbf{w} + C\sum_{i=1}^{N}\xi_i \quad (9)$$

$$subject\ to: \begin{cases} y_i(\mathbf{w}\cdot\Phi(\mathbf{x}_i)+ \mathrm{b}) \geq 1-\xi_i,\ \forall\, i = 1,..N & (a) \\ \xi_i \geq 0,\ \forall\, i = 1,..N & (b) \end{cases} \quad (10)$$

where the slack variables ξ_i and the regularization parameter *C* are introduced to deal with non-linearly separable training data.

When the training samples are linearly separable in the feature space, $\xi_i = 0\ \forall\ i = 1,..,N$. In this case the SVM algorithm constructs the so called maximum margin hyperplane which is the hyperplane that separates the training data without errors (i.e. the one satisfying the constraints y_i ($\mathbf{w}\ \Phi(\mathbf{x}_i)+ \mathrm{b}) \geq 1,\ \forall\ i = 1,..,N$) and maximizes the distance between itself and the closest training vectors of each class. Such a distance, called margin, is equals to $2/(\mathbf{w}\cdot\mathbf{w})$ and it is maximized by minimizing $(\mathbf{w}\cdot\mathbf{w})/2$. The greater is the margin, the greater is the SVM generalization ability. If training samples cannot be separated without error, the parameters ($\mathbf{w}$, b) that minimize the functional in the (9) under the constraints defined by (10), determine the hyperplane that minimize the training error, measured through the sum of the slack variables ξ, and separates the rest of the elements with the maximum margin. For each training sample lying on the wrong side of the decision boundary (i.e. those with associated a slack variable $\xi_i > 1$) but also for the training samples falling within the margin on the correct side of the decision boundary (i.e the samples for which $0 < \xi_i \leq 1$) a penalty is introduced, increasing the objective function by $C\xi_i$. A too small *C* value, determines many misclassifications. Conversely, a too large *C* value may lead to overfitting problems. The user parameter *C* controls the tradeoff between the two goals of the optimization problem.

In the standard SVM algorithm introduced so far, the regularization parameter *C* equally penalizes misclassification of positive and negative samples. When faced with an unbalanced dataset where the number of negative instances is significantly higher than that of positive ones, SVMs that follow the formulation of equations (9) - (10) tend to produce a decision boundary

severely skewed towards the minority class (Akbani et al., 2004; Batuwita, and Palade, 2013). To cope with this issue, Veropoulos et al. (1999) proposed an approach that uses different misclassification costs, C^+ and C^-, for positive and negative classes, with $C^+ > C^-$. In this method, the SVM optimization problem defined by equations (9) - (10) is modified by replacing the penalty term, $C\sum_{i=1}^{N}\xi_i$, with the sum of two terms, one for each class:

$$\min_{w,b,\xi} \frac{1}{2}\mathbf{w}\cdot\mathbf{w} + C^+ \sum_{i \mid y_i=1} \xi_i + C^- \sum_{i \mid y_i=-1} \xi_i \tag{11}$$

$$subject\ to: \begin{cases} y_i(\mathbf{w}\cdot\Phi(\mathbf{x}_i)+ \mathrm{b}) \geq 1-\xi_i, \ \forall\, i = 1,..N & (a) \\ \xi_i \geq 0, \ \forall\, i = 1,..N & (b) \end{cases} \tag{12}$$

The constrained minimization problem defined by the equations (11) and (12), is usually solved considering its dual-problem counterpart:

$$\max_{\alpha} \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} \alpha_i\, \alpha_j\, y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \tag{13}$$

$$subject\ to: \begin{cases} 0 \leq \alpha_i \leq C^+, & \forall\, i \mid y_i = 1 (a) \\ 0 \leq \alpha_i \leq C^-, & \forall\, i \mid y_i = -1 \ (b) \\ \sum_{i=1}^{N} \alpha_i\, y_i = 0, & \forall\, i = 1,..N \ \ (c) \end{cases} \tag{14}$$

The variables α_i represent the Lagrange multipliers corresponding to the constraints defined by (12). Training vectors for which $\alpha_i > 0$ are called support vectors. They are the only ones that determine the final decision function. In other words, removing a non-support vector from the training set does not change the solution found by the SVM algorithm. $K(\mathbf{x_i}, \mathbf{x_j})$ is a kernel function, K: $\mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$, returning the inner product $K(\mathbf{x_i}, \mathbf{x_j}) = \Phi(\mathbf{x_i})\cdot\Phi(\mathbf{x_j})$ between the images of two data points $\mathbf{x_i}$, $\mathbf{x_j}$ in the high-dimensional feature space induced by Φ. It allows comparing two patterns $\Phi(\mathbf{x_i})$, $\Phi(\mathbf{x_j})$ in the high dimensional feature space, without explicitly mapping the

data to that space (Scholkopf, 2001). Valid kernel functions are those satisfying the Mercer's theorem (Mercer, 1909). For this work we chose the Radial Basis Function (RBF) kernel $K(\mathbf{x}_i, \mathbf{x}_j) = exp\left(-\gamma \|\mathbf{x_i} \cdot \mathbf{x_j}\|^2\right)$, were γ is a non-negative parameter that define the kernel width. Given the solution α* of the dual problem, the class label of a test sample $\mathbf{x}_t$ is obtained according the following decision function:

$$f(\mathbf{x_t}) = sign(\mathbf{w}^* \cdot \Phi(\mathbf{x_t}) + b^*) = sign\left(\sum_{i=1}^{N_s} \alpha_i^* y_i K(\mathbf{x}_i, \mathbf{x_t}) + b^*\right) \quad (15)$$

where it can be noted that the summation is limited to the N_s support vector.

Tuning of the classifier parameters

Both the selected classification algorithms have parameters that need to be tuned. In order to estimate the class-conditional pdfs, the kernel widths of the Parzen window method have to be specified. In this study, we used a single bandwidth *h* for each of the *d* elements of the feature vector **x** (h_j=*h,* for all *j*). To make input data more spherical and justify the use of a common bandwidth for each dimension, data were preliminarily standardized to have mean of zero and standard deviation of one (Ghosh et al., 2006). Many different techniques are available in the literature for choosing the optimal bandwidth from the data. A review of these methods can be found for example in Chiu (1996). One of the most popular methods for selecting the kernel width is the Silverman's rule (Zhang et al., 2006), which assumes that the true distribution is normal. In this work, we used such rule for initializing our algorithm that empirically search for the optimal kernel width, as detailed in the following.

For the SVM algorithm, the hyperparameters to be tuned are the regularization parameters of the positive and negative classes (C^+ and C^-, respectively) and the parameter γ of the RBF kernel

396 which defines its width. We expressed the misclassification cost C^+ of the positive class, which
397 is the minority class we are interesting in, as a function of the misclassification cost C^- of the
398 negative class, through the relation $C^+=wC^-$, were w is a weight greater than 1. Then, the
399 parameters we tuned were C^-, w and γ.
400 In order to optimize the parameters $\boldsymbol{\varphi}$ of the classifiers, i.e. the kernel width h for the MAP
401 classifier ($\boldsymbol{\varphi}_{MAP} = \{h\}$) and the parameters C^-, w and γ for SVM ($\boldsymbol{\varphi}_{SVM} = \{C^-, w, \gamma\}$) we ran a
402 repeated k-fold CV, with k = 10, each repetition resulting in a different Cohen's Kappa of the
403 classifier. Namely, for a fixed classifier parameter vector $\boldsymbol{\varphi}$, the training and the validation are
404 performed k times until all data are used. At the end of the kth iteration of the CV, the Cohen's
405 Kappa, obtained with the current parameters vector $\boldsymbol{\varphi}$, is derived from the global CM. It is
406 therefore a function of $\boldsymbol{\varphi}$, i.e., K= -f($\boldsymbol{\varphi}$), which is considered as an objective function to be
407 minimized. Then, the classifier parameters are updated iteratively, i.e., for each repetition of the
408 CV we test a different value of the classifier parameters selected by a simulated annealing
409 algorithm, implemented through a built-in Matlab© function which aimed to find the minimum
410 value of f($\boldsymbol{\varphi}$) (i.e., maximum Cohen's Kappa).
411 In more details, in order to initialise the kernel width of the Parzen window method, we set a
412 value h_0 computed according the Silveman rule (Zhang et al., 2006). Then we searched for its
413 optimal value in the range [h_0/10, 10h_0]. The hyperparameter γ of SVM was initialized to the
414 LibSVM default value of $\gamma_0 = 1/d$, where d is the dimension of the feature vector used as input
415 (Chang and Lin, 2011), and it was tuned in the range [10^{-3}, 2]. The regularization parameter C^-
416 and the weight w were varied in the ranges [1, 10] and [2, 50], respectively, starting from the
417 initial settings $C_0^- = 1$ and w_0 = n−/ n+, where n+ and n- are the cardinalities of the positive and
418 negative classes. The initial setting chosen for w is a heuristic value commonly used to balance

419 *C+* and *C−*, so that the potential total cost of the false positives equals the potential total cost of
420 the false negatives (Morik et al., 1999).
421 Note that the additive inverse of Cohen's Kappa f(**φ**), computed after the k=10 interactions of the
422 k-folded CV, was chosen as objective function of the optimization algorithm. We preferred the
423 Cohen's Kappa to other classification performance measures, such as the overall accuracy,
424 because the latter is inappropriate when dealing with an unbalanced dataset. If the dataset is
425 highly unbalanced, the classification algorithm can achieve high overall accuracy simply
426 assigning all the samples to the majority class. Consequently, searching for the classifier
427 parameters giving the best cross-validated performances in terms of overall accuracy can lead to
428 a classifier strongly biased towards the majority class. The Cohen's Kappa considers also the off-
429 diagonal elements of the CM, thus providing a more reliable estimate of the classifier
430 performances.

431 **4. Results and discussion**

432 *4.1 Visual interpretation benchmark*

433 In order to have a reference when assessing the performance of an automatic algorithm in
434 detecting damage, a photointerpretation of pre- and post- event Quickbird images was performed
435 building by building. The comparison between photointerpretation results and INGV ground
436 truth is reported in Table 3, which provides a sort of benchmark against which automatic
437 algorithms can be judged.

		Ground Truth data (INGV)				
		D = 5	D = 4	D = 3	D <=2	total
Visual	damaged	45	46	4	0	95

inspection	undamaged	29	221	781	541	1572
	total	74	267	785	541	1667

Table 3: Confusion matrix of the photointerpretation of QuickBird data with respect to the INGV ground survey.

A missed detection error of 39.2% and 82.8% was found for buildings that, according the INGV ground survey, sustained a damage of grade 5 and 4, respectively. Only 4 out of 785 buildings classified as damage grade 3 were recognized as damaged by visual inspection. This result demonstrates that only the detection of the damage grade 5 is feasible using remote sensing data.

A few samples of the appearance of buildings and the problems that can be encountered when an automatic algorithm is used for damage detection are shown in Figure 4. This figure compares pre- and post- event QuickBird images for some buildings classified as damage grade 5 by the INGV ground survey. Assuming that the ground survey is telling the truth, the building severe damage in Figure 4(a) and Figure 4(b) are easily detectable by visual inspection, whereas the same is not true for the building in Figure 4(c) that appears unchanged. We will discuss possible reasons for that later on in the paper.

Pre-event

Figure 4: Comparison of the pre- (upper panels) and post- event (lower panels) QuickBird images for some buildings classified as damage grade 5 by the INGV ground survey

4.2 Change detection feature selection and classification results

Before attempting the final damage classification, a careful selection of the considered features was carried out. This is important to avoid using useless features (noisy or correlated with others) in the classification, but also to understand the physical mechanisms that make damage detectable by an optical sensor. We searched for the most relevant subsets of features according to a wrapper approach, which consists in ranking subset of features based on the classification results they achieve (Guyon and Elisseeff, 2003).

We evaluated the effect on the classifier performances by varying the number and the combination of the features used as input. For each set of d candidate features chosen in an exhaustive way among the 13 features introduced in section 3.2, we considered the classification results obtained after tuning the classifier parameters according to the method sketched in the previous section, and then we selected the combination that provided the best results in terms of the Cohen's Kappa. In particular, due to the uncertainty in the Cohen's Kappa estimation from a finite size sample set, for each group of d features, we selected those combinations giving a Cohen's Kappa within the range [maxKappa(d), maxKappa(d) - 0.02], where maxKappa(d) is

the maximum Cohen's Kappa achieved with *d* features. We performed this analysis independently for MAP and SVM classifiers. Results for groups of features of size from 1 to 5 are summarized in Table 4 and Table 5, which refer respectively to MAP and SVM classifiers. These tables consist of five columns indicating the number of features, and 13 rows, one for each of the considered change parameters. Coloured cells in each sub-column represent selected features. The first sub-column of each group reports the combination of features giving the best performance, i.e. the combination of *d* features for which the Cohen's Kappa was equal to maxKappa(*d*).

When a single feature was considered as input of the change detection algorithm, change in the contrast provided the best results, either with SVM or MAP classifier. By looking at Table 4Table 4, it can be noticed that the best results of the MAP classifier were all achieved with feature sets containing change in the contrast, which is the most selected feature also in the case of SVM (see Table 4). Indeed, when a building collapses, the contrast generally increases due to presence within the building footprint of pixels with very different grey levels close each other. Consistent performances were also achieved including changes in Homogeneity and Entropy, which are frequently included in the optimal subsets. Then, changes in textural properties show themselves as the most powerful features for detecting changes due to the building damage. Being inversely but strongly correlated, Homogeneity and Entropy are interchangeable. In fact, either using SVM or MAP classifier, these two features are selected reciprocally.

By comparing Table 4 and Table 5, it can be seen that there are some features that demonstrated their usefulness depending on the classification approach. The Mutual Information, for example, being very sensitive to residual registration errors between the two images, generally determines a lot of false alarms when used as input to the MAP classifier, while it provides satisfactory

results if exploited, in combination with other features, as input to SVMs. In combination with other features, high values of such parameter can help avoiding false positive due to changes not attributable to the earthquake but associated, for example, to a building restoration intervened in the timeframe between the two image acquisitions. It can be noted that, in the case of the MAP classifier, feature combinations providing the best results in terms of Cohen's Kappa often include change in the colour saturation, which conversely does not appear within the features subset for which SVM provided the best results. This is probably because the Mutual Information from PSH images (i.e., multispectral data), which is one of the most frequently selected features in the case of SVM, brings colour signature as well.

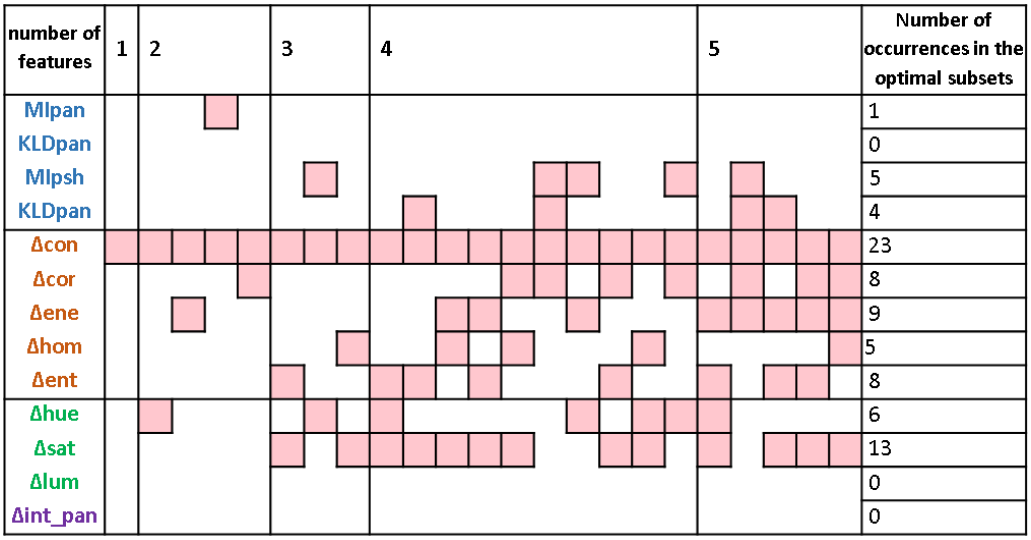
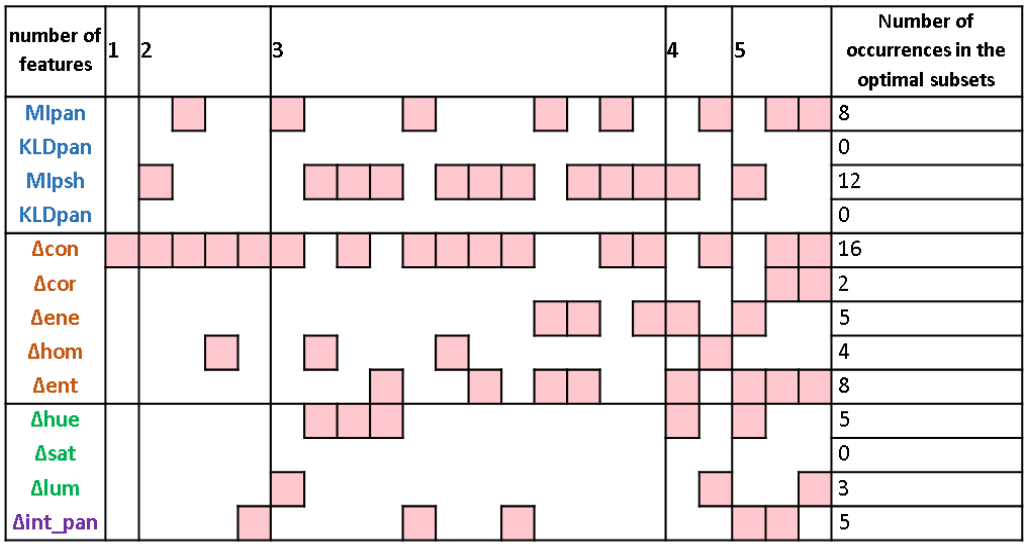


Table 4: Exhaustive feature selection based on Cohen's Kappa for different feature vector size using MAP classifier.



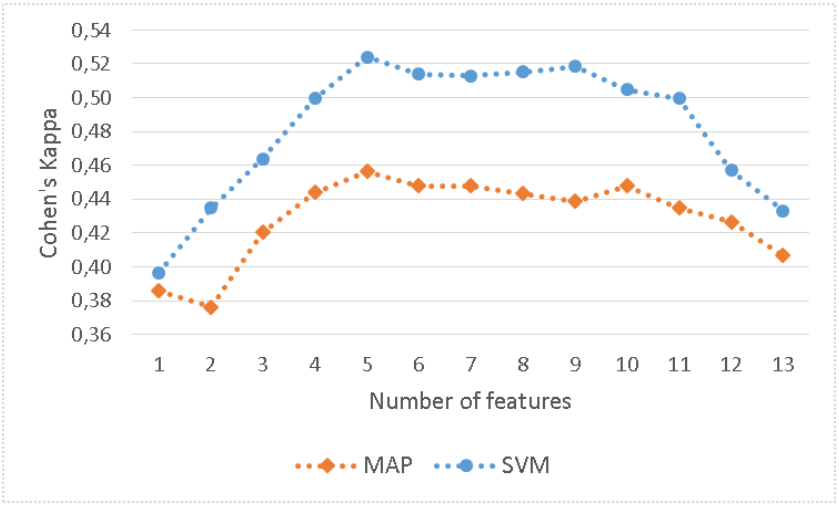
503

504 **Table 5: Exhaustive feature selection based on Cohen's Kappa for different feature vector size using SVM.**

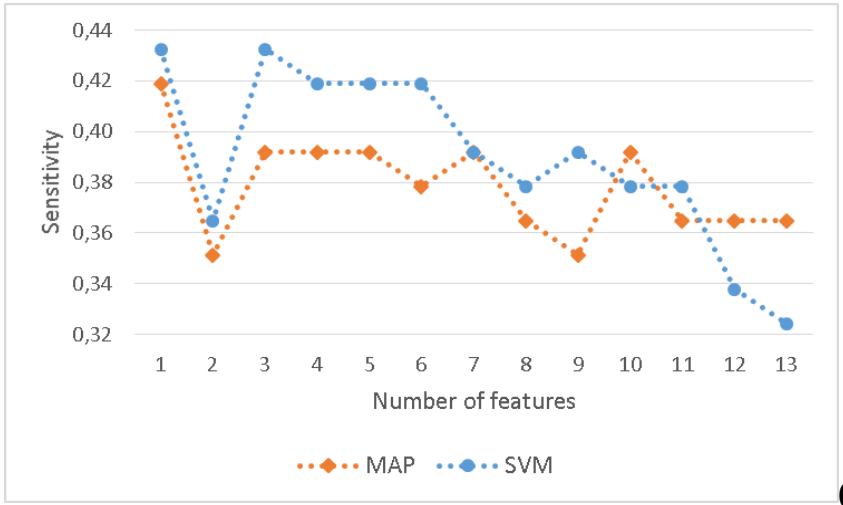
505 Figure 5(a) reports the Cohen's Kappa as a function of the features number for both SVM and
506 MAP classifier, considering the best subset for each group of d features. In Figure 5(b) and
507 Figure 5(c), classification approaches are compared in term of sensitivity and precision,
508 respectively. In the context of damage detection, the sensitivity represents the percentage of
509 building to which a damage grade 5 was assigned from both ground survey and classification
510 algorithm. As for the precision, it represents the percentage of buildings classified as damage
511 grade 5 from satellite data that have actually sustained a damage of grade 5, according to the
512 ground survey.

513 Looking at Figure 5(a), it can be seen that increasing the number of features improves the
514 classification Cohen's Kappa, until reaching the highest value with a subset of 5 features, for
515 both the considered classification approaches. Then, the performances in term of Cohen's Kappa
516 tend to decrease as the input space dimensionality increases. From Figure 5(b) and Figure 5(c), it
517 can be noted that increasing the number of features from 1 to 5 does not improve the sensitivity
518 of the algorithms to the damage detection, but it increases its precision, i.e. the rate of buildings

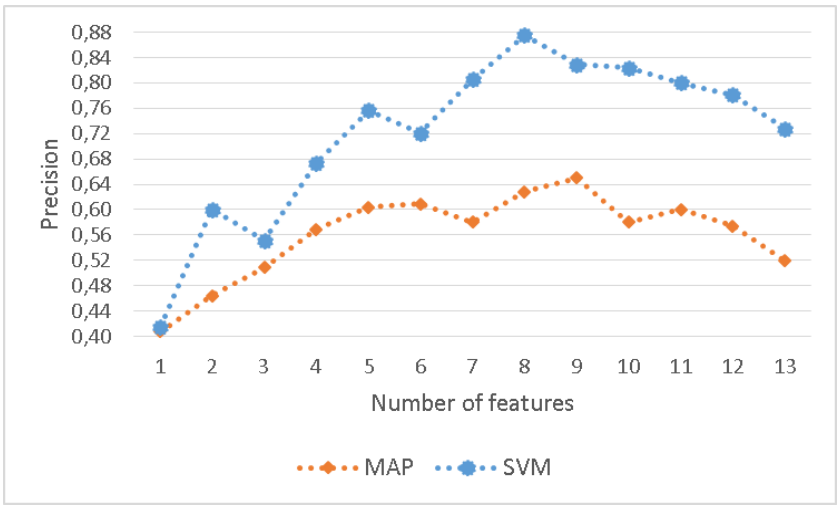
519 classified as damage grade 5 that are real damages (i.e., decrease of false alarms).



(a)



(b)



(c)

520

Figure 5: Classification performances of SVM and MAP classifier as a function of the number of features used as input: (a) Cohen's Kappa (b); Sensitivity; (c) Precision.

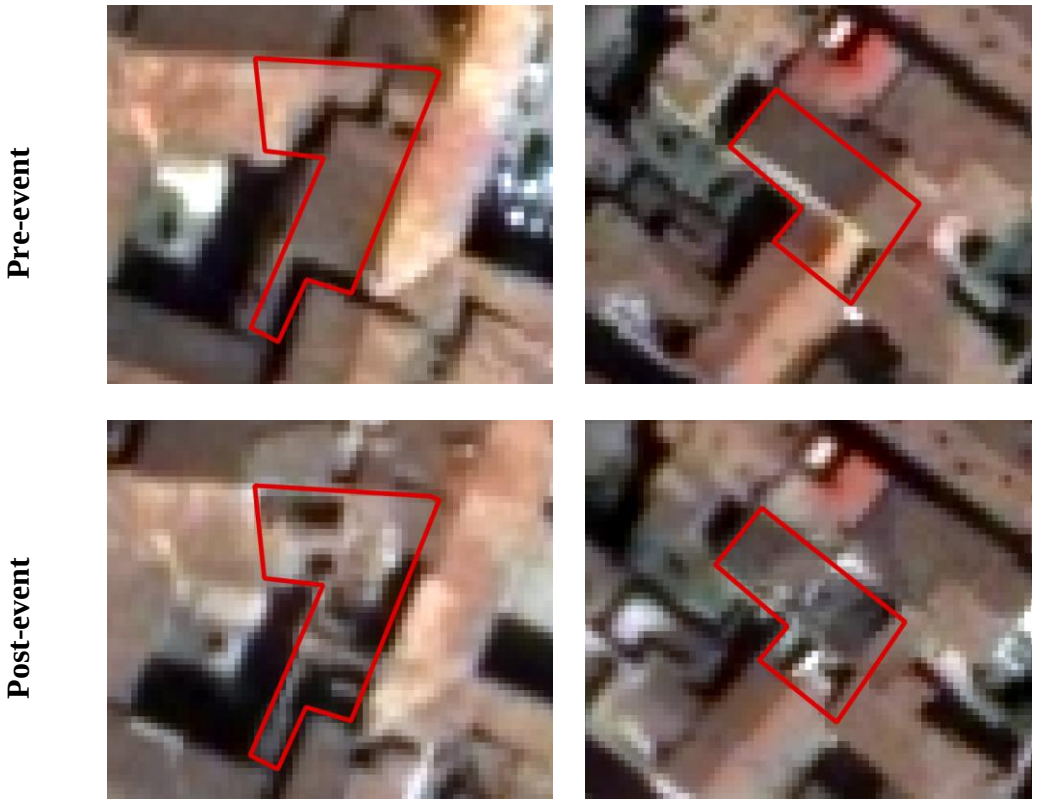
Figure 6 shows that the results we obtained using SVM outperform those achieved by the MAP classifier. With respect to the MAP classification approach, SVMs are able to provide classifications characterized by a lower rate of false alarms. This could be due to the complexity of the input data that a SVM is more able to recover since it works in a higher dimensional feature space. Besides providing better performance in term of precision, in most cases SVM correctly recognizes a greater number of building with damage grade 5. However, performances in term of sensitivity are not so markedly distinct and both algorithms often fail in detecting a building with damage grade 5 according to the ground survey, mainly because, as pointed out in section 4.1, about 40% of these damages are not detectable even by visual inspection.

The confusion matrix shown in Table 6 summarizes the results of the MAP classification approach with the subset of five features that achieved the highest Cohen's Kappa. Using changes in Contrast, Energy, Entropy, Hue and Saturation the MAP classification algorithm exhibits an overall accuracy of 96.2% and a Cohen's Kappa of 45.6%.

		INGV (D=5)	(D<5)			
classifier	(D=5)	**29**	19	48	60.4%	***precision***
	(D<5)	45	**1574**	1619	97.2%	***Negative Predictive Value***
		74	1593	1667	***overall accuracy***	***Cohen's Kappa***
		sensitivity	***specificity***			
		39.2%	98.8%		96.2%	45.6%

Table 6: Confusion Matrix obtained using the MAP classifier with the optimal subset of features (changes in Contrast, Energy, Entropy, Hue and Saturation).

We correctly detect 39.2% of buildings with damage grade D=5, generating 19 false positives. The result in terms of false alarms can be considered satisfactory, especially if it is considered that 5 false positives are determined by buildings that suffered a damage D<5 according to the INGV ground survey, but with an apparent heavy damage according to image photointerpretation, as shown in Figure 6(a) and Figure 6(b). In fact, the figures show two buildings classified respectively as damage grade 4 and 3 in the INGV survey, which were detected as heavily damaged by the MAP classifier. They actually look very differently in the post-seismic image, so that they were likely misclassified during the survey. This poses some questions about the definition of damage grade D=5, the capability of correctly evaluating it during the field work, and the reliability of using a ground survey as an absolute reference for classification accuracy assessment.



551 **Figure 6: Examples of false positives generated by the MAP classifier. Upper panels: pre-event Quickbird**
552 **images. Lower panels: post-event Quickbird images. Polygons delineate buildings that have sustained a**
553 **damage of grade 4 (figure (a)) and 3 (figure (b)) according the INGV ground truth.**

554 As far as the low sensitivity to the damage is concerned (around 40% according to Figure 5b), it
555 is worth pointing out that our algorithm correctly recognized 60% of the buildings with damage
556 level 5 that we were able to identify by visual inspection of the Quickbird images. In other word,
557 most of missed detection of damage 5 according to INGV were not detectable even by visual
558 interpretation of the images.
559 As discussed in section 4.1, the class of the buildings to which a damage grade 5 was assigned is
560 very heterogeneous. In some cases, the damage sustained by the building roof, although
561 detectable by a human image interpreter, was not so extensive but rather it was comparable with
562 lower damage grades, thus preventing its automatic identification without generating a large
563 amount of false alarms. An example is reported in Figure 7(a), where the roof is almost intact,
564 except for the portion within the yellow ellipse. There are, in addition, some buildings suffering a
565 first floor collapse while retaining an intact roof (i.e., suffering a pancake type of collapse), such
566 as that shown in Figure 7(b). In these cases, buildings were wrongly classified as belonging to
567 the class of damage D<5 and the only change in the image is the presence of dust and debris
568 around the building footprint.

Figure 7: Examples of false negatives generated by the MAP classifier. Upper panels: pre-event Quickbird images. Lower panels: post-event Quickbird images. (a) The roof is almost completely intact, except for the portion within the yellow ellipse. (b) Due to the pancake type of collapse, the only change is the presence of dust and debris around the building footprint.

Using SVM we obtained better results, as shown in the confusion matrix reported in Table 7, which refers to the SVM outcomes when Mutual Information from PSH images, changes in Energy, Entropy, Hue and Intensity from PAN images were used as input. Sensitivity and precision increase to 41.9% and 75.6%, respectively. We correctly classify 31 out of 74 buildings with damage grade 5, identifying about 68% of those detected by visual inspection. Even in this case, some of the false alarms (3 out of 10) are due to buildings recognized as damaged by photointerpretation.

		INGV (D=5)	INGV (D<5)			
classifier	(D=5)	**31**	10	41	75.6%	*precision*
	(D<5)	43	**1583**	1619	97.4%	*Negative Predictive Value*
		74	1593	1667	*overall accuracy*	*Cohen's Kappa*
		sensitivity	*specificity*			
		41.9%	99.4%		96,8%	52.4%

Table 7: Confusion Matrix obtained using SVM with the optimal subset of features (Mutual Information from PSH images, changes in Energy, Entropy and Hue, change in Intensity from PAN images).

The classification performances against the INGV ground survey demonstrates the challenge of detecting earthquake damage at single building scale from satellite data. However, it is very important to notice that the classification results assessed against the INGV ground truth (overall accuracy=96.2%, Cohen's Kappa=45.6% using the MAP classifier; overall accuracy=96.8%, Cohen's Kappa=52.4% using SVM) are quite similar and even better than the direct comparison between the INGV surveys and the one made available from the Italian Department of Civil Protection (DPC). The DPC survey was carried out during the six months following the seismic event with the objective of quantifying the grade of damage and assessing the building usability. Information about typological, damage and usability characteristics of residential buildings were collected using the AeDES forms (Agibilità e Danno nell'Emergenza Sismica) (Baggio et al., 2007). Building were inspected internally (where possible) and externally and a damage grade according to EMS-98 was associated to each structural element, namely, vertical structures, floor, stairs, roof, infills and partitions. Monumental and historical edifices were surveyed in the frame of the Safeguard of Cultural Heritage from Natural Disasters action, independently on the acquisitions of the AeDES forms, collecting similar but not identical information. Starting from these data the Istituto per le Tecnologie della Costruzione (ITC) of the Italian National Research Council (CNR) developed a georeferenced database reporting, for each surveyed building, an overall damage indicator and a vulnerability class that were calculated following the EMS-98 scale.

The comparison we performed between the damage maps resulting from the INGV and DPC ground surveys (Pierdicca et al., under review) showed and overall agreement of 92.7% and a Cohen's Kappa of 33.6%. Such result suggest that the damage map obtained using satellite data

over the entire historical town of L'Aquila shows a mismatch with respect to the INGV ground truth comparable to the mismatch of the two ground surveys. It demonstrates the undoubted value of remote sensing, at least for a rapid response, when compared to the highly costly and time-consuming ground surveys.

5. Conclusions

In this paper we have investigated the capability of mapping urban damage due to an earthquake at the scale of individual buildings using a pair (i.e., pre- and post-event) of Very High Resolution satellite optical images. The case study considered in this research was the earthquake that hit L'Aquila historical town in 1996. The automatic detection of collapsed buildings was compared to the ground survey carried out by the *Istituto Nazionale di Geofisica e Vulcanologia* just after the seism, referring to the standard European Macroseismic Scale 1998 (EMS-98) and considering the whole historical town. Thus, the experiment represents an almost unique example of an extensive validation of building collapse mapping from space in quite challenging conditions (i.e., dense urbanization, historical town). Among different features extracted from the image at object scale (i.e., within the footprint of each building), the change of textural features and features identifying change in colour were found as the most effective. The Support Vector Machine supervised classification revealed itself to outperform the Bayesian classifier based on a nonparametric Parzen window approach. In particular, SVM resulted in a fewer number of false alarms. The overall classification performances where not bad (overall accuracy=96.8%, Cohen's Kappa=52.4%). However, a limited sensitivity was observed, in many cases due to the absence of appreciable changes in the post-seismic image, not detectable even by a photointerpreter. False alarms were in some cases related to apparent changes in the post-seismic

images that surprisedly did not correspond to a *collapse* label in the ground survey. The work demonstrates the challenge of damage assessment after an earthquake, a task where satellite remote sensing revealed itself not exhaustive, but anyhow capable of providing an effective prompt and cheap support to the disaster management institutions.

6. Acknowledgement

The work has been funded by the EC-FP7 APhoRISM project (Research, Technological Development and Demonstration Activities, grant agreement n. 606738). The INGV ground survey has been carried out by the QUEST - QUick Earthquake Survey Team (http://quest.ingv.it). We thank the Italian Department of Civil Protection for providing their survey data and the Istituto per le Tecnologie della Costruzione (ITC) of the Italian National Research Council (CNR) for the georeferenced version. The L'Aquila urban maps have been provided by the Regione Abruzzo.

7. References

Akbani, R., Kwek, S., Japkowicz, N. Applying support vector machines to imbalanced datasets. In Proceedings of the 15th European Conference on Machine Learning, pp. 39–50, 2004.

Baggio C, Bernardini A, Colozza R, Di Pasquale G, Dolce M, Goretti A, Martinelli A, Orsini G, Papa F, Zuccaro G, Pinto AV, Taucer F. Field Manual for post-earthquake damage and safety assessment and short term countermeasures (AeDES). EUR 22868 EN, Joint Research Center, ISPRA, Italy, 2007.

649 Batuwita, R., Palade, V. Class imbalance learning methods for support vector machines. In
650 Imbalanced learning: Foundations, algorithms, and applications, (pp. 83–99). John Wiley &
651 Sons, Inc., 2013

652 Bignami, C., Chini, M., Stramondo, S., Emery, W. J., Pierdicca, N. Objects textural features
653 sensitivity for earthquake damage mapping. In 2011 Joint Urban Remote Sensing Event, 333-
654 336, IEEE, April 2011.

655 Bouleimen, K., Lecocq, H. A new efficient simulated annealing algorithm for the resource-
656 constrained project scheduling problem and its multiple mode version. European Journal of
657 Operational Research, 149.2, 268-281, 2003.

658 Brower, B., Laben, C. Process for enhancing the spatial resolution of multispectral imagery using
659 pan-sharpening. US Patent 6011875, 2000.

660 Brunner, D., Lemoine, G., Bruzzone, L. Earthquake damage assessment of buildings using VHR
661 optical and SAR imagery. IEEE Transactions on Geoscience and Remote Sensing, 48(5), 2403-
662 2420, 2010.

663 Chang, C. C., Lin, C. J. LIBSVM: a library for support vector machines. ACM transactions on
664 intelligent systems and technology (TIST), 2(3), 27, 2011.

665 Chen, G., Hay, G. J., Carvalho, L. M., Wulder, M. A. Object-based change detection.
666 International Journal of Remote Sensing. 33(14), 4434-4457, 2012.

667 Chesnel, A. L., Binet, R., Wald, L. Object oriented assessment of damage due to natural disaster
668 using very high resolution images. In IEEE Geoscience and Remote Sensing Symposium, 2007.

669 IGARSS 2007, 3736-3739, 2007

670 Chini, M., Pierdicca, N., Emery, W.J. Exploiting SAR and VHR optical images to quantify
671 damage caused by the 2003 Bam earthquake. IEEE Transactions on Geoscience and Remote
672 Sensing ,47 (1), 145–152, 2009.

673 Chiu, S. T. A comparative review of bandwidth selection for kernel density estimation. Statistica
674 Sinica. 6(1), 129-145, 1996.
675 Congalton, R.G., Green, K. Assessing the Accuracy of Remotely Sensed Data. Boca Raton, FL:
676 Lewis, 1999.

677 Cortes, C., Vapnik, V. Support-vector networks. Machine learning, 20(3), 273-297, 1995.

678 Dell'Acqua, F., Gamba, P. Remote sensing and earthquake damage assessment: Experiences,
679 limits, and perspectives. Proceedings of the IEEE, 100(10), 2876-2890, 2012.

680 Dekkers, A., Aarts, E. Global optimization and simulated annealing. Mathematical
681 programming, 50.1-3, 367-393, 1991.

682 DigitalGlobe Inc., Core Imagery Product Guide V2.0, December 2014, available on line:
683 http://global.digitalglobe.com/resources/technical-information.

684 Devanthéry, N., Luzi, G., Stramondo, S., Bignami, C., Pierdicca, N., Wegmüller, U.,
685 Romaniello, V., Anniballe, R., Piscini, A., Albano, M., Moro, M., Crosetto, M. The validation
686 activities of the Aphorism EC 7FP project, aimed at post seismic damage mapping, through a
687 combined use of EOS and ground data. In Proceedings of the Living Planet Symposium, Prague,
688 Czech Republic, May 9-13, 2016., ESA-SP Volume 740, ISBN: 978-92-9221-305-3, p.76., 2016.

689 Dong, L., Shan, J. A comprehensive review of earthquake-induced building damage detection
690 with remote sensing techniques. ISPRS Journal of Photogrammetry and Remote Sensing, 84, 85-
691 99, 2013.

692 Erten, E., Reigber, A., Ferro-Famil, L., Hellwich, O. A new coherent similarity measure for
693 temporal multichannel scene characterization. IEEE Transactions on Geoscience and Remote
694 Sensing, 50(7), 2839-2851, 2012.

695 Ghosh, A. K., Chaudhuri, P., Sengupta, D. Classification Using Kernel Density Estimates.
696 Technometrics, 48(1), 2006.

697 Grünthal G. (ed.): European Macroseismic Scale 1998 (EMS-98). Cahiers du Centre Européen
698 de Géodynamique et de Séismologie 15, Centre Européen de Géodynamique et de Séismologie,
699 Luxembourg, 99 pp., 1998.

700 Gusella, L., Adams, B. J., Bitelli, G., Huyck, C. K., & Mognol, A. Object-oriented image
701 understanding and post-earthquake damage assessment for the 2003 Bam, Iran, Earthquake.
702 Earthquake Spectra, 21(S1), 225–238, 2005.

703 Guyon, I., Elisseeff, A. An introduction to variable and feature selection. Journal of machine
704 learning research, 3(Mar), 1157-1182, 2003.

705 Haralick, R. M., Shanmugam, K., Dinstein, I. H. Textural features for image classification. IEEE
706 Transactions on System, Man, and Cybernetics, 610-621, 1973.

Joyce, K. E., Belliss, S. E., Samsonov, S. V., McNeill, S. J., Glassey, P. J. A review of the status of satellite remote sensing and image processing techniques for mapping natural hazards and disasters. Progress in Physical Geography, 33(2), 183-207, 2009

Kullback, S., Leibler, R. A. On information and sufficiency. The annals of mathematical statistics, 79-86, 1951.

Marin, C., Bovolo, F., Bruzzone, L. Building change detection in multitemporal Very High Resolution SAR images. IEEE Transactions on Geoscience and Remote Sensing, 53(5), 2664-2682, 2015.

Morik K, Brockhausen P, Joachims T. Combining Statistical Learning with a Knowledge-Based Approach - A Case Study in Intensive Care Monitoring. Proceedings of the Sixteenth International Conference on Machine Learning, pp. 268–277, 1999.

Mercer, J. Functions of positive and negative type, and their connection with the theory of integral equations. Philosophical transactions of the royal society of London. Series A, containing papers of a mathematical or physical character, 415-446, 1909.

Parzen, E. On estimation of a probability density function and mode. The annals of mathematical statistics. 33(3), 1065-1076, 1962.

Pierdicca N., Anniballe R., Noto F., Bignami C., Chini M., Martinelli A., Mannella A. Triple collocation to assess classification accuracy without a ground truth in case of earthquake damage assessment. IEEE Transactions on Geoscience and Remote Sensing, under review, 2017.

726 Rastiveis, H., Samadzadegan, F., Reinartz, P. A fuzzy decision making system for building
727 damage map creation using high resolution satellite imagery. Natural Hazards and Earth System
728 Sciences, 13.2, 455, 2013.

729 Scholkopf, B. The kernel trick for distances. Advances in neural information processing systems,
730 301-307, 2001.

731 Smith, A. R. Color gamut transform pairs. ACM Siggraph Computer Graphics. Vol. 12. No. 3.
732 ACM, 1978.

733 Stramondo, S. Bignami, C., Chini, M., Pierdicca, N., Tertulliani, A. Satellite radar and optical
734 remote sensing for earthquake damage detection: results from different case studies. International
735 Journal of Remote Sensing, 27, 4433 - 4447, 2006.

736 Tertulliani A., Arcoraci L., Berardi M., Bernardini F., Camassi R., Castellano C., Del Mese S.,
737 Ercolani E., Graziani L., Leschiutta I., Rossi A., Vecchi M. An application of EMS98 in a
738 medium-sized city: The case of L'Aquila (Central Italy) after the April 6, 2009 Mw 6.3
739 earthquake. Bulletin of Earthquake Engineering, 9, 67–80, 2011.

740 Tertulliani A., Leschiutta I., Bordoni P., Milana G. Damage Distribution in L'Aquila City
741 (Central Italy) during the 6 April 2009 Earthquake. Bulletin of the Seismological Society of
742 America 102, 1543–1553. 2012.

743 Vapnik, V. N. An overview of statistical learning theory. IEEE Transactions on Neural
744 Networks, 10(5), 988-999, 1999.

745 Veropoulos, K., Campbell, C., Cristianini, N. Controlling the sensitivity of support vector
746 machines. In Proceedings of the International Joint Conference on Artificial Intelligence, pp. 55-
747 60, 1999.

748 Yamazaki, F., Yano, Y., Matsuoka, M. Visual damage interpretation of buildings in bam city
749 using quickbird images following the 2003 Bam, Iran, earthquake. Earthquake Spectra, 21(S1),
750 329-336, 2005.

751 Zhang, X., King, M. L., Hyndman, R. J. A Bayesian approach to bandwidth selection for
752 multivariate kernel density estimation. Computational Statistics & Data Analysis, 50(11), 3009-
753 3031, 2006.
754