SECOND EUROPEAN CONFERENCE ON EARTHQUAKE ENGINEERING AND SEISMOLOGY, ISTANBUL AUG. 25-29, 2014

REMOTE SENSING FOR SEISMIC BUILDING VULNERABILITY ASSESSMENT

Christian GEIß[1], Martin KLOTZ[1] and Hannes TAUBENBÖCK[1]

ABSTRACT

We quantitatively evaluate the suitability of multi-sensor remote sensing to assess the seismic vulnerability of buildings. Therefore, features are derived from remote sensing data to characterize the urban environment. The derived features are combined with *in situ* observations. Machine learning approaches are deployed subsequently to identify meaningful sets of features that are suitable to predict seismic vulnerability levels of buildings. Supervised regression techniques are used to estimate the actual vulnerability level. Empirical evidence with respect to the viability of the approach is revealed for the city of Padang (Indonesia). When assessing the vulnerability level according to a scoring method, the overall mean absolute percentage error is 10.6%, if using a supervised Support Vector Regression approach. This study uncovers potential for a rapid screening assessment of large areas which should be further explored in the future.

INTRODUCTION

The impact of natural disasters such as earthquakes on mankind has increased dramatically over the last decades. Global urbanization processes and increasing spatial concentration of exposed elements such as people, buildings, infrastructure, and economic values in earthquake prone regions induce seismic risk at a uniquely high level. This situation, when left unmitigated, will cause unprecedented death tolls, enormous economic and ecological losses, critical infrastructure and service failures, and poses a significant threat for civil security, and a sustainable development in the future (Bilham 2009*)*. More specifically, a dynamic urban growth is often accompanied by the construction of large shares of unplanned, spontaneous and often highly vulnerable settlements. Simultaneously, these settlements are highly variable over short time scales. In this regard, the continuous assessment and monitoring of the seismic vulnerability of buildings is a challenging task, especially when large-area evaluations are required. Numerous studies emphasize that remote sensing can play a valuable role in supporting the extraction of relevant features for pre-event vulnerability analysis of built-up structures (French and Muthukumar 2006, Mueller et al. 2006, Sarabandi et al. 2007, Taubenböck et al. 2009, Borfecchia et al. 2010, Sahar et al. 2010, Borzi et al. 2011, Deichmann et al. 2011, Wieland et al. 2012). The intrinsic advantage of remote sensing is the ability to offer an overview of building stocks and serve as a screening method for derivation of building vulnerability related features, such as shape characteristics, height, roof material, year of construction, structure type and spatial context (Geiß and Taubenböck 2013).

In this manner, this contribution aims to demonstrate how to derive and identify meaningful features using a combination of complementary remote sensing data and how to quantitatively evaluate their predictive power. By means of a sequential procedure of calculating features from very

[1] German Aerospace Center (DLR), German Remote Sensing Data Center (DFD), Weßling, Germany

high resolution multispectral data, height information from a normalized digital surface model, spatiotemporal analyses based on Landsat data, scare *in situ* knowledge and of applying machine learning algorithms, the seismic vulnerability levels of buildings could be estimated with viable accuracies, spatially distributed and area-wide for the example of the city of Padang (Indonesia) (Geiß et al. 2014).

DATA

The experimental analyses are carried out in Padang (Indonesia) since it is located in a very earthquake prone region and comprehensive *in situ* and remote sensing data was collected and provided for this study. The *in situ* data on seismic vulnerability in Padang was collected in February/March, 2008 within the "*Last-Mile*" project (Taubenböck et al. 2009b). The building inventory database compiled for Padang, based on a ground truth survey, includes information about physical characteristics of 434 buildings in the city. The sampling scheme of the buildings aimed at both the incorporation of all existing housing types of Padang and broad spatial coverage. The database includes information about geometry, material of bearing structures and walls, foundations and local soil conditions, material of the roof, type of building, etc. For most of the inspected buildings the dataset is supplemented by results of physical tests providing information about the reinforcement and quality of concrete of the main bearing structures. Additionally, for every building in the database there is an indication of the damage level due to previous earthquakes. To assess the buildings based on these parameters, a scoring approach was developed. The building parameters are first expressed quantitatively on a normalized scale. High values express characteristics that are considered as favorable, regarding the buildings' seismic vulnerability. Subsequently, individual weights for the respective parameters are assigned based on expert observations made in the study area during previous research. For instance, an established method to assess the stability properties of reinforced concrete is the aforementioned Schmidt rebound hammer test. Completed structural survey results in the study area confirmed the hammer test outcomes to be the most important indicator and are therefore given the highest weight. Finally the weighted values of the respective parameters are summed up. For brevity the reader is referred to Mück et al. (2013) for a detailed description of conceptual and methodological details of this approach. Calculated scoring values lie within an interval of [10.85, 25.6], and have a mean value of 18.4 and a standard deviation of 2.89. Lower scoring values express higher building vulnerability, while higher values represent lower vulnerability.

In addition to the *in situ* data, a set of remote sensing data was acquired. Optical IKONOS imagery (acquisition date: 2005-04-12), which covers a spectral range of 0.445-0.853 μm, with a geometric resolution of 1 m in the panchromatic band and 4 m in the multispectral (blue, green, red, nir) bands, was acquired. Height information was derived from a digital surface model (DSM) and a digital terrain model (DTM). Measurements from the models were based on the return signals received by two radar antennas mounted on an aircraft, and the application of SAR interferometry. Both data sets have a geometric resolution of 5 m and a Root-mean-square-error *(RMSE)* regarding the vertical accuracy of 1 m. In addition, we used data from the LANDSAT sensors Thematic Mapper (acquisition date: 1989-07-25) and Enhanced Thematic Mapper (acquisition date: 2000-07-15). Both sensors have 7 multispectral bands covering a spectral range of 0.45-2.35 μm with a maximum geometric resolution of 30 m (TM) and 15 m (ETM+). The data can be accessed free-of-charge within a public image archive which dates back to 1972.

METHODS

A sequential procedure is introduced, which comprises feature calculation from remote sensing data, feature selection and function estimation. Before we calculated features from the remote sensing data, we first manually digitized the building footprints and derived building blocks from a road network (Taubenböck et al. 2008). In the subsequent steps, features were calculated for both. The building level was used for calculating features that refer to characteristics of the respective building, whereas the block level was used to describe the spatial setting the respective buildings are embedded in.

In particular, the features relate to the two-dimensional extent of buildings as well as the description of their shape characteristics. In addition, statistical values of 1^{st} and 2^{nd} order were extracted from the available IKONOS imagery at building and block level. The first serve as a descriptor of roof surface material and arrangement whereas the latter are intended to describe the composition of distinct urban structures. Mean and standard deviation values of the different image bands as well as band ratios, which are intended to emphasize spectral dissimilarities, were calculated. Additionally, rotation-invariant texture measures for the panchromatic and near-infrared band were computed using both the co-occurrence matrix (GLCM) and grey level difference vector (GLDV). Features explicitly aiming to describe the spatial context are calculated at block level and consist of the area of building blocks and the average size of the buildings located within. Furthermore, spatial metrics such as proportion measures of land cover classes are computed. Additionally, a semantic classification ("Structure Type $_S$"), which is built on physical features that describe the urban morphology, is incorporated. The classification describes the socio-economic status of the population by distinguishing "slums", "suburbs", "low income areas", "medium income areas", and "high income areas". Beyond, the incorporation of height information allows the calculation of 3D features such as building floor number, floor space, ratio of diameter and height, ratio of width and height, as well the average building height within a building block. The mean slope for each building block was calculated to describe topographic location characteristics. By analyzing two Landsat images from 1989 and 2000, the period of construction is approximately described based on a post classification change detection procedure. Overall, each building object is represented by a 132-dimensional feature vector, whereby 73 features are related to individual buildings, and 59 provide block level information.

The selection of features to be used for building regression models is generally a difficult task, especially when dealing with a huge number of features as in this study. These often exhibit redundancy, are highly correlated, and suffer from the "Hughes phenomenon". The latter describes the effect that for a limited amount of samples the predictive power decreases as the dimensionality of the feature vector increases. Therefore, two machine learning based feature selection algorithms were applied on the data set. In this manner, one can discriminate algorithms which evaluate individual features and those which assess subsets of features (Hall & Holmes 2003). As such, the Relief-F (Kononenko 1994) approach was chosen because it enables to rank individual features. Additionally the Correlation-based Feature Selection (CFS) method was chosen, since it enables the scoring of the value of groups of features (Liu et al. 2002).

For estimating the scoring values we compare the merits of multi-linear regression models (Montgomery et al. 2001) and Support Vector Machine (SVM) based regression models which are able to represent non-linear boundaries between classes. Multi-linear regression is based on the assumption that the dependent variable Y and its predictors $X_1, X_2, \ldots, X_n$ are directly related by a linear combination. Since linear regression models predict poorly in the presence of a nonlinear or non-additive relationship, a nonlinear Support Vector Regression (SVR) approach is additionally utilized. SVMs determine a suitable set of parameters that places a decision surface, the so called hyperplane, between the different classes of training samples according to their position in an n-dimensional feature space. The optimal separating hyperplane is identified as the maximized margin between the different classes and the hyperplane. In a modified form SVMs can also be applied for function estimation (see Smola and Schölkopf 2004). Detailed theoretical background of SVMs is given in Vapnik (1998).

RESULTS

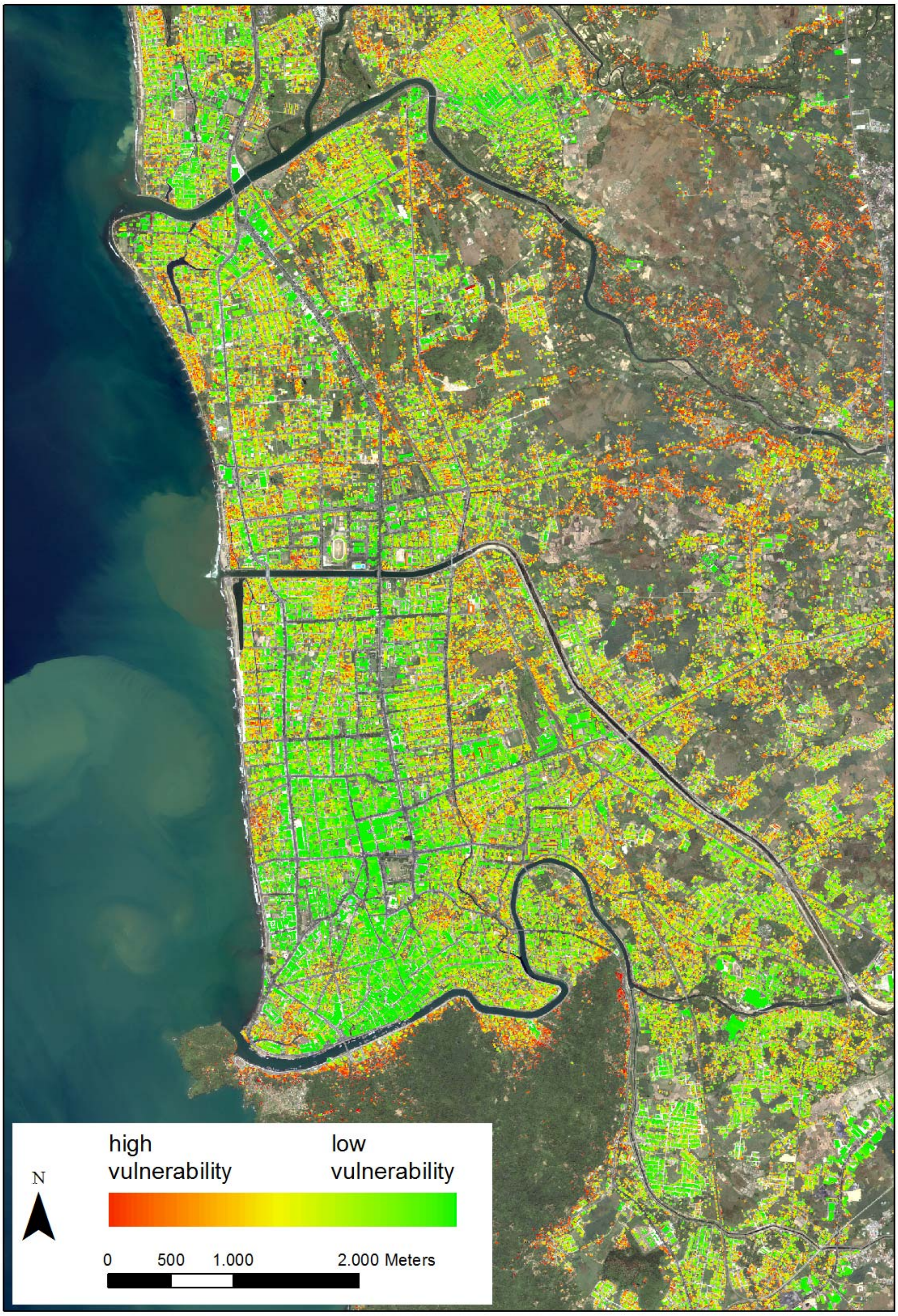
The multi-linear and SVR models are calculated based on (*i*) the originate 132-dimensional feature vector, (*ii*) the 70 features with a positive degree of relevance as evaluated by the Relief-F algorithm, (*iii*) the group of 21 features with the highest merit as revealed by the CFS approach. The evaluated model estimation results are given in table 1. Regarding the comparison of the actual and estimated vulnerability values, we calculated a set of statistical accuracy measures: Mean Error (*ME*), Mean Percentage Error (*MPE*), Mean Absolute Error (*MAE*), Mean Absolute Percentage Error (*MAPE*), and Pearson product-moment correlation coefficient (*R*).

Table 1. Evaluated results of the model predictions of different feature sets compared to reference values) (Geiß et al. 2014).

Multi-linear regression					
Used features	*MAE*	*MAPE*	*ME*	*MPE*	*R*
All features	2.33	13.32	–0.56	–3.02	0.43
Relief-F	2.02	11.84	–0.11	–0.58	0.48
CFS	1.84	10.99	–0.17	–0.91	0.56
Support Vector Regression					
All features	1.88	11.26	0.02	0.13	0.53
Relief-F	1.79	11.07	0.43	2.31	0.57
CFS	1.72	10.61	0.23	1.25	0.59

When using the CFS feature set and the SVR approach the best results are achieved, with lowest *MAE* (1.72) / *MAPE* (10.61) and highest linear correlation (*R* =0.59) of the model estimates. This demonstrates the viability of the approach. Furthermore, the nonlinear approach shows better predictions compared to the linear approach regarding the respective set of features used. In general, the models overestimate low scoring values, buildings with a high seismic vulnerability – and underestimates high scoring values - characterizing buildings with a low seismic vulnerability. The applied regression model to the building inventory is shown in figure 1. It can be seen that especially large, engineered buildings are assigned a high scoring value (green). Such kind of buildings can be found primarily in the central parts of Padang. In contrast, especially small and informal buildings located in remote parts of the town have a low scoring value (orange).

Fig 1. Applied regression model for the building inventory for Padang (Indonesia). Orange color depict buildings that are highly vulnerable and green color depict buildings that have a low vulnerability level.



CONCLUSIONS

This study shows that an indirect correlation between physical information in the (urban) environment, drawn from remote sensing data and seismic vulnerability of buildings, exists. We demonstrated how to derive and identify meaningful features using a combination of remote sensing data and how to quantitatively evaluate their explanatory power. By means of a sequential procedure of calculating features from very high resolution multispectral data, height information and spatiotemporal analyses, and of applying machine learning approaches, the seismic vulnerability of buildings can be estimated with viable accuracies. We conclude that remote sensing data and methods have a high capability to support large area assessments of building vulnerability, indicating the need for systematic application and validation of these findings. Lastly, we want to trigger an open dialogue between the remote sensing and earthquake engineering community, to better understand how remote sensing data can be linked to assessment approaches from engineering science in a robust, standardized and transferable way, to define common scales and enable systematic large-area assessments and monitoring of dynamic earthquake prone urban areas around the globe.

ACKNOWLEDGEMENT

We want to acknowledge the support by the European Commission's Seventh Framework Programme [FP7/2007-2013], under grant agreement no. 312972 "Framework to integrate Space-based and in-situ sENSing for dynamic vUlnerability and recover Monitoring"

REFERENCES

Bilham, R., 2009. The seismic future of cities, *Bulletin of Earthquake Engineering* **7**, 839–887.

French, S.P., and Muthukumar, S., 2006. Advanced technologies for earthquake risk inventories, *Journal of Earthquake Engineering* **10**, 207–236.

Mueller, M., Segl, K., Heiden, U., and Kaufmann, H., 2006. Potential of high-resolution satellite data in the context of vulnerability of buildings, *Natural Hazards* **38**, 247–258.

Sarabandi, P., Kiremidjian, A., 2007. Development of algorithms or building inventory compilation through remote sensing and statistical inferencing. The John A. Blume Earthquake Engineering Center, Department of Civil and Environmental Engineering, Stanford University, Stanford, CA, USA, Report No. 158, 419p.

Taubenböck, H., Roth, A., Dech, S., Mehl, H., Münich, J.C., Stempniewski, L., and Zschau, J., 2009. Assessing building vulnerability using synergistically remote sensing and civil engineering, in *Urban and Regional Data Management*, Kreck, A., Rumor, M., Zlatanova, S., Fendel, E., Eds, Taylor & Francis Group, London, 287–300

Borfecchia, F., Pollino, M., De Cecco, L., Lugari, A., Martini, S., La Porta, L., Ristoratore, E., and Pascale, C., 2010. Active and passive remote sensing for supporting the evaluation of the urban seismic vulnerability, *Italian Journal of Remote Sensing* **42**, 129–141.

Sahar, L., Muthukumar, S., and French, P., 2010. Using aerial imagery and GIS in automated building footprint extraction and shape recognition for earthquake risk assessment of urban inventories, *IEEE Transactions on Geoscience and Remote Sensing* **48**, 3511–3520.

Borzi, B., Dell'Acqua, F., Faravelli, M., Gamba, P., Lisini, G., Onida, M., and Polli, D., 2011. Vulnerability study on a large industrial area using satellite remotely sensed images, *Bulletin of Earthquake Engineering* 9, 675–690.

Deichmann, U., Ehrlich, D., Small, C., and Zeug, G., 2011. Using high resolution satellite data for the identification of urban natural disaster risk. JRC, GFDRR, IPSC, WB, 80pp.

Wieland, M., Pittore, M., Parolai, S., Zschau, J., Moldobekov, B., and Begaliev, U., 2012. Estimating building inventory for rapid seismic vulnerability assessment: Towards an integrated approach based on multi-source imaging, *Soil Dynamics and Earthquake Engineering* **36**, 70-83.

Geiß, C., Taubenböck, H., 2013. Remote sensing contributing to assess earthquake risk: from a literature review towards a roadmap. *Natural Hazards* **68**, 7-48.

Geiß, C., Taubenböck, H., Tyagunov, S., Tisch, A., Post, J., and Lakes, T., 2014. Assessment of seismic building vulnerability from space. *Earthquake Spectra*, in press. Doi: http://dx.doi.org/10.1193/121812EQS350M

Taubenböck, H., Goseberg, N., Setiadi, N., Lämmel, G., Moder, F., Oczipka, M., Klüpfel, H., Wahl, R., Schlurmann, T., Strunz, G., Birkmann, J., Nagel, K., Siegert, F., Lehmann, F., Dech, S., Gress, A., and Klein, R., 2009b. « Last-Mile » preparation for a potential disaster – Interdisciplinary approach towards tsunami early warning and an evacuation information system for the coastal city of Padang, Indonesia. *Natural Hazards and Earth System Science* **9**, 1509–1528.

Mück, M., Taubenböck, H., Post, J., Wegscheider, S., Strunz, G., Sumaryono, S., Ismail F.A., 2013. Assessing building vulnerability to earthquake and tsunami hazard using remotely sensed data. Natural Hazards, 68, 97-114.

Taubenböck, H., Post, J., Kiefl, R., Roth, A., Ismail, F.A., Strunz, G., and Dech, S., 2008. Risk and vulnerability assessment to tsunami hazard using very high resolution satellite data, in *Proceedings of the EARSeL Joint Workshop: Remote Sensing: New Challenges of high resolution*, Carsten, J. (editor), Bochum, Germany

Hall, M., and Holmes, G., 2003. Benchmarking attribute selection techniques for discrete class data mining, *IEEE Transactions on Knowledge and Data Engineering* **15**, 1–16.

Kononenko, I., 1994. Estimating attributes: Analysis and extension of RELIEF, in *Proceedings of the European Conference on Machine Learning*, Springer, Catania, Italy, pp. 171–182

Liu, H., Li, J., and Wong, L., 2002. A comparative study on feature selection and classification methods using gene expression profiles and proteomic patterns, *Genome Informatics* **13**, 51–60.

Montgomery, D.C., Peck, E.A., and Vining, G.G., 2001. Introduction to Linear Regression Analysis, 3rd edition, Wiley, New York, NY, 672 pp.

Smola, A.J., and Schölkopf, B., 2004. A tutorial on support vector regression, *Statistics and Computing* **14**,199–222.

Vapnik, V.N., 1998. Statistical learning theory, *Wiley Series on Adaptive and Learning Systems,* Wiley, New York