This is an excerpt from the thesis "*Estimation of seismic building structural types using remote sensing and machine learning*".

Please contact Patrick Aravena Pelizari for a full version of the thesis.

Extended works on this topic are documented in: Geiß, C., Aravena Pelizari, P., Marconcini, M., Sengara, W., Edwards, M., Lakes, T., and Taubenböck, H. (2015): Estimation of seismic buildings structural types using multi-sensor remote sensing and machine learning techniques. ISPRS Journal of Photogrammetry and Remote Sensing, 104, 175–188.

LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
Fakultät für Geowissenschaften
Department für Geographie

Estimation of seismic building structural types using remote sensing and machine learning

Abschätzung seismischer Gebäudestrukturtypen mittels Fernerkundung und Maschinellem Lernen

MASTERTHESIS
zur Erlangung des akademischen Grades
Master of Science

Studiengang:
Umweltsysteme und Nachhaltigkeit
Monitoring, Modellierung, Management

Eingereicht von:
Patrick Manuel Aravena Pelizari

Gutachter:
Prof. Dr. Ralf Ludwig

München, April 2013

This master thesis was elaborated at the German Remote Sensing Data Center (DFD)
of the German Aerospace Center (DLR)
Oberpfaffenhofen

Department of Geo-Risks and Civil Security
Head of Department
Prof. Dr.-Ing. Günter Strunz

Contents

Abstract

The current trend of urbanization leads to an increase of seismic vulnerability in earthquake prone regions. There is great demand for methods contributing to a comprehensive analysis of seismic vulnerability to face the urgent challenges of mitigation and catastrophe management. Remote sensing has high potential to contribute to an area-wide and up-to-date assessment of seismic vulnerability. For an estimation of building stock damage the built-inventory is generally categorized into different seismic building structural types, representing a construction's seismical behavior.

This study reveals indirect correlations between remotely sensed data and seismic building structural types, which enable a supervised classification. Site of research is the City of Padang, Indonesia, whose urban environment is characterized by 145 features calculated by means of high resolution optical imagery, height information from a normalized digital surface model and multi-temporal medium resolution optical data. In-situ building information is given through survey data collected after the earthquake event of September 2009. Using Machine Learning techniques a work flow is presented to classify seismic building structural types. A feature selection analysis is carried out, and the features most explanatory for the determination of seismic building structural types are identified. Coping with large amounts of features and in-situ data scarcity, plausible classification results are achieved and dependencies between remotely sensed data and building stability are verified.

Zusammenfassung

Die fortschreitende Urbanisierung führt zu einer erhöhten Vulnerabilität erdbebengefährdeter Regionen. Um den dringlichen Herausforderungen von Mitigation und Katastrophenmanagement begegnen zu können, benötigt man Methoden, die eine aktuelle und flächendeckende Bewertung der Erdbebenvulnerabilität ermöglichen. Die Fernerkundung verfügt über großes Potential, dies zu gewährleisten. Zur Abschätzung von Gebäudeschäden werden Gebäudebestände in der Regel in unterschiedliche "seismische Gebäudestrukturtypen" eingeteilt, die jeweils das Verhalten eines Gebäudes bei seismischer Erschütterung wiedergeben.

In der vorliegenden Arbeit werden indirekte Korrelationen zwischen Fernerkundungsdaten und seismischen Gebäudestrukturtypen aufgezeigt, die eine überwachte Klassifikation ermöglichen. Untersuchungsort ist die Stadt Padang in Indonesien. Die Gebäude und ihr jeweiliges urbanes Umfeld werden durch 145 Attribute charakterisiert, die aus hoch aufgelösten optischen Daten, Höheninformationen aus einem digitalen Oberflächenmodell sowie aus multitemporalen optischen Daten mittlerer Aufösung berechnet wurden. Die verwendeten in situ Gebäudeinformationen stammen aus Gutachterdaten, die nach dem Erdbeben vom September 2009 erhoben wurden. Im Rahmen der vorliegenden Studie wurde eine Prozesskette zur Klassifikation seismischer Gebäudestrukturtypen entwickelt, in der Techniken aus dem Maschinellen Lernen zur Anwendung kommen. Es wird eine Selektionsanalyse der Attribute vorgenommen, wobei die Attribute mit dem höchsten Erklärungsgehalt identifiziert werden. Unter Berücksichtigung von Problemen, die sich aus der sehr hohen Anzahl von Attributen sowie einer sehr geringen Anzahl an in situ Daten ergeben können, werden Zusammenhänge zwischen Fernerkundungsdaten und Gebäudestabilität verifiziert und plausible Klassifikationsergebnisse erzielt.

List of Figures

List of Tables

List of Abbreviations

AIFDR	Australia-Indonesia Facility for Disaster Reduction
C-SVC	C-Support Vector Classifier
CART	Classification and Regression Tree
CFS	Correlation-based Feature Selection
CM	Confined Masonry
COR	Pearson's Correlation Coefficient
CV	Cross-Validation
DSM	Digital Surface Model
DTM	Digital Terrain Model
EQ	Earthquake
GR	Gain Ratio
IG	Information Gain
ITB	Institut Teknologi Bandung
LOO-CV	Leave-One-Out Cross-Validation
M_w	Moment Magnitude Scale
MBR	Minimum Bounding Rectangle
ML	Machine Learning
MMI	Modified Mercalli Intensity Scale
nDSM	Normalized Digital Surface Model
OA	Overall Accuracy
OBIA	Object Based Image Analysis
OC-SVM	One-Class Support Vector Machine
OOB	Out-Of-Bag
PA	Producer's Accuracy

PDF	Probability Density Function
PGA	Peak Ground Acceleration
PGV	Peak Ground Velocity
RBF	Radial Basis Function
RC high	Reinforced Concrete high
RC low	Reinforced Concrete low
RelF	Relief-F
RF	Random Forest
SAR	Synthetic Aperture Radar
SBST	Seismic Building Structural Type
SF	Steel Frame
SMOTE	Synthetic Minority Over-sampling Technique
SV	Suport Vector
SVDD	Support Vector Data Description
SVM	Support Vector Machine
TF non-res	Tmber Frame Non-Residential
TF res	Timber Frame Residential
UA	User's Accuracy
URM	Unreinforced Masonry
USGS	U.S. Geological Survey

1 Introduction

1.1 Motivation

"Vulnerability and exposure to disasters is increasing as more people locate in areas of high risk" (UNISDR/WMO, 2012). From 1970 to 2011 the world's population has grown by 86% (UNDESA, 2012). Strongly connected to the phenomenon of rural to urban migration, urban growth has absorbed most of this increase (Bilham, 2009). Since 2008 for the first time in history there are more people who live in urban areas than in rural areas. More than half of the large cities, with populations ranging from 2 to 15 million, are situated in areas of high seismic risk (UNISDR/WMO, 2012) and yield the high vulnerability of modern societies and technologies. These developments strongly contributed to the dramatic increase in losses caused by natural catastrophes, as observed in the last few decades (Calvi et al., 2006). Until 2050 the world population is expected to increase by further 70% attended by accelerated urban sprawl (UNDESA, 2012; 2011). Accordingly, in the foreseeable future urban populations in earthquake prone regions are anticipated to suffer greater measures of fatalities, injuries, structural damage and economic losses resulting from earthquakes than in the documented past (Bilham, 2009).

Particularly in developing countries urbanisation is very dynamic. Thereby, non-sustainable, unplanned use of the territory and inadequate construction practises are often resulting in highly vulnerable settlements with an inherent variability over short time scales (Oliveira et al., 2006; Wieland et al., 2012). The biggest risk in case of an earthquake emanates from buildings. Therefore the identification and localisation of structures and their subsequent evaluation regarding their physical vulnerability is the basis for the planning and implementation of mitigation strategies to reduce vulnerability (Taubenböck et al., 2009b). A comprehensive and realistic evaluation of earthquake vulnerability needs up-to date and area-wide exposure information. In-situ information gathering alone is not able to fulfill these demands. Remote sensing has a large potential to contribute to such an assessment.

1.2 Research Framework & Objective

Seismic risk is the potential or probability of a loss due to the occurrence of an earthquake (EQ) and can be seen as a combination of three main elements (Fig. 1.1): the EQ hazard, the assets at risk (i.e, the value that is threatened by the EQ) and the vulnerability of the assets to the effect of the earthquake (Scawthorn, 2008).

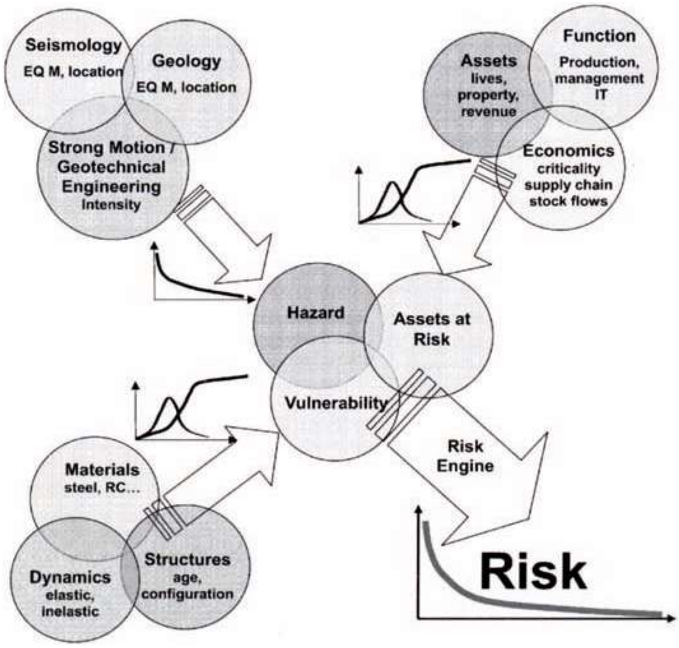

Figure 1.1: Elements of seismic risk (Scawthorn, 2008)

Hazard refers to the physical effects of the EQ (e.g. faulting, ground shaking, liquefaction landslides, tsunamis etc.) on the natural environment. Assets at risk involve all values threatened by the earthquake hazard, which could either be of physical (e.g. human lives, property) or of non-physical value (e.g. education, social cohesiveness, functional downtime). Vulnerability is the susceptibility of the assets to the impact of the hazard, and can be defined as their degree of loss resulting from a given level of hazard (Scawthorn, 2008). Mathematically the causal relationship between the mentioned concepts can be expressed as:

$$R_{ij} = H_j \times V_{ij} \tag{1.1}$$

where for an element at risk i: R_{ij} is the risk, the probability or average rate of loss to element i due to earthquake ground motion severity j. H_j is the hazard, the probability or average expected rate of experiencing earthquake ground motion (or an other earthquake related damaging event) of severity j. V_{ij} is the vulnerability, the level of loss that would be caused to element i of experiencing earthquake hazard of severity j (Coburn and Spence, 2002). EQs can neither be predicted nor prevented. Damage usually occurs rapidly during the few seconds of shaking, so that people cannot be evacuated. This rapidity of occurrence and the lack of warning leads to human casualties. The biggest risk of being killed or injured and suffering losses related to an earthquake event emanates from building structures, which accordingly constitute a major determinant factor of the physical EQ vulnerability at a certain site. Hence, in earthquake affected regions population, risk mitigation and emergency planners as well as insurance and reinsurance industries have a strong demand in the evaluation of the vulnerability of their building structures (Douglas, 2007; Coburn and Spence, 2002). Such an assessment requires detailed structural engineering knowledge and is commonly addressed to the seismic engineering domain.

The evaluation of the vulnerability of structures in a certain region, at first demands an in-situ building stock taking and analysis. If the investigation is related to more than just a few buildings, the building stock is categorized into a typology of distinct *seismic building structural types* (SBSTs) whose seismic behavior is likely to be similar. The primary load-bearing structure is the most affecting factor for earthquake damage and thus usually the first property considered to categorize a building. Further parameters which may reflect the seismic performance of a building and hence are frequently used for categorization are e.g. the number of storeys, the period of construction or the presence of structural irregularities (Coburn and Spence, 2002). Based on empirical, analytical, hybrid approaches and/or expert knowledge for each SBST under investigation, a function is determined that correlates the magnitude, e.g. the Modified Mercalli Intensity (MMI) or peak ground acceleration (PGA), of the seismic hazard at site to the physical damage probably suffered by the structural built system (Calvi et al., 2006). By means of such a fragility function the probable damage distribution of the building stock at risk, given at a certain level of hazard, can be predicted (Douglas, 2007). These vulnerability models are often integrated into a more holistic modeling frame work, comprising further elements of seismic risk for instance HAZUS (FEMA, 2012), OpenQuake (GEM, 2013) or RiskScape (Schmidt et al., 2011). In this manner seismic engineering enables an accurate stability assessment of individual structures by an extensive house-by-house inspection and the analysis of construction plans (Mück et al., 2012). Following this way, however, a detailed area-wide and regularly updated assessment, which is able to account for the rapidly changing spatio-temporal conditions in many present day cities, is prohibitive owing to time and cost (Wieland et al., 2012). Very few countries may maintain registries which may partially allow to infer SBSTs and their spacial distribution, in general however – if at all – broad and detailed structural information on buildings is only available for lifeline inventory and critical facilities. That applies in particular for developing or threshold countries subjected to high dynamics of urban sprawl (Oliveira et al., 2006).

Due to recent developments of different Earth observation platforms with increased spatial, spectral and temporal resolution remote sensing provides more detailed information on land cover and the environmental state than ever before. Different sensor systems operate in different wavelengths, ranging from visible to microwave, supplying different but complementary information (Waske et al., 2009). Taking advantage of object-based image analysis (OBIA) methods (Blaschke, 2010), in addition to the spectral information from individual pixels, geometric and contextual information of multi-source data can be exploited. This delivers a high variety of different information, which has shown high potential to describe urban environment and to be suitable to contribute the evaluation of seismic risk (Geiß and Taubenböck, 2012). Remote sensing could make it possible to maintain a thorough and up to date inventory of exposed built-up structures in order to contribute to a comprehensive and effective seismic vulnerability assessment. Spatial distribution of potential losses can greatly improve the organization of rapid response or mitigation actions (Pittore and Wieland, 2012).

The aim of this study is to classify SBSTs, as described above, area-wide and in a supervised manner, utilizing the whole range of available remote sensing data. The research area for this task is the Indonesian city Padang, which possesses a densely populated and dynamic urban system in a region of extreme high probability of severe earthquakes (Rusnardi et al., 2012). The key in-situ data represent building inventory data affiliated with SBSTs, that was collected in the context of the M_w 7.6 earthquake event in September 2009. The remotely sensed data base comprises n=145 features. These features were calculated from multi-sensor imagery by means of state of the art information extraction techniques and characterize the buildings and their urban environment. To perform a supervised classification of SBSTs a work flow utilizing several methods from the Machine Learning[1] domain is applied; herein a feature selection analysis which allows to identify the features, most suitable to distinguish between different SBSTs. Figure 1.2 coarsely outlines the methodological framework of this study.

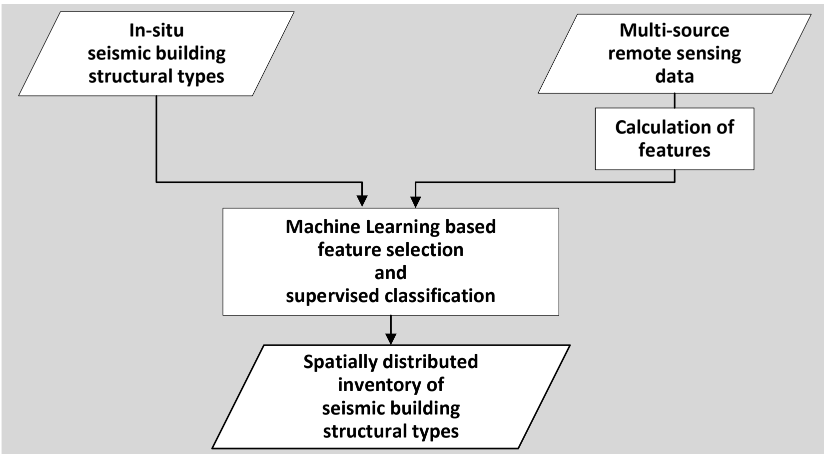

Figure 1.2: Methodological framework for the classification of seismic building structural types

The aim of this methodology is to tackle the following research questions:

1. Do SBSTs correlate with remote sensing data?
2. Which parameters that can be derived by remote sensing data are most explanatory to describe the SBST of buildings?
3. How suitable are these parameters to estimate the SBST of a building?
4. Does a spatially distributed estimation of SBSTs show reasonable results?

[1]"Machine Learning is an area of artificial intelligence and generally refers to the development of methods that optimize their performance iteratively by *learning from the data*" (Waske et al., 2009)

This is an excerpt from the thesis "*Estimation of seismic building structural types using remote sensing and machine learning*".

Please contact Patrick Aravena Pelizari for a full version of the thesis.

Extended works on this topic are documented in: Geiß, C., Aravena Pelizari, P., Marconcini, M., Sengara, W., Edwards, M., Lakes, T., and Taubenböck, H. (2015): Estimation of seismic buildings structural types using multi-sensor remote sensing and machine learning techniques. ISPRS Journal of Photogrammetry and Remote Sensing, 104, 175–188.

Bibliography

Abe, S. (2003). Analysis of multi-class support vector machines. *Proc. International Conference on Computational Intelligence for Modelling Control and Automation*, pages 385–396. Url: http://www.lib.kobe-u.ac.jp/repository/90000226.pdf, accessed 22.03.2013.

Abe, S. (2005). *Support vector machines for pattern classification (advances in pattern recognition)*. Springer, New York.

Banerjee, A., Burlina, P., and Diehl, C. (2006). A support vector method for anomaly detection in hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing,*, 44(8):2282–2291.

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

Bishop, C. M. (2006). *Pattern recognition and machine learning (information science and statistics)*. Springer, New York.

Blaschke, T. (2010). Object based image analysis for remote sensing. *ISPRS Journal of Photogrammetry and Remote Sensing*, 65(1):2 – 16.

BNPB (2009). Badan Nasional Penanggulangan Bencana (National Disaster Management Agency, Indonesia). Laporan Harian Pusdalops BNPB Minggu, 01 November 2009 (Daily BNPB report, Saturday 01 November 2009). Url: http://www.bnpb.go.id, accessed 17.11.2012.

Bottou, L., Cortes, C., Denker, J. S., Drucker, H., Guyon, I., Jackel, L. D., LeCun, Y., Muller, U. A., Sackinger, E., Simard, P., and Vapnik, V. (1994). Comparison of classifier methods: a case study in handwritten digit recognition. In *International Conference on Pattern Recognition*, pages 77–87.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24:123–140.

Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.

Breiman, L. (2002). Manual on setting up, using, and understanding random forests v3.1. Department of Statistics, University of California, Berkeley. Url: http://oz.berkeley.edu/users/breiman/Using_random_forests_V3.1.pdf, accessed 07.02.2013.

Breiman, L., Friedman, J. H., Olshen, R. A., and Stone, C. J. (1984). *Classification and Regression Trees*. Wadsworth.

Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, 2(2):121–167.

Calvi, G. M., Pinho, R., Magenes, G., Bommer, J. J., Restrepo-Vélez, L. F., and Crowley, H. (2006). Development of seismic vulnerability assessment methodologies over the past 30 years. *ISET Journal of Earthquake Technology*, 43(3):75–104.

Camps-Valls, G. and Bruzzone, L., editors (2009). *Kernel methods for remote sensing data analysis*. John Wiley & Sons.

Chang, C.-C. and Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27. Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm, accessed 15.03.2013.

Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. *J. Artif. Intell. Res. (JAIR)*, 16:321–357.

Chlieh, M., Avouac, J. P., Sieh, K., Natawidjaja, D. H., and Galetzka, J. (2008). Heterogeneous coupling of the Sumatran megathrust constrained by geodetic and paleogeodetic measurements. *J. Geophys. Res.*, 113(B5):B05305.

Cieslak, D. A., Chawla, N. V., and Striegel, A. (2006). Combating imbalance in network intrusion datasets. In *2006 IEEE International Conference on Granular Computing*, pages 732–737.

Coburn, A. and Spence, R. (2002). *Earthquake protection*. John Wiley & Sons.

Colaninno, N., Roca, J., and Pfeffer, K. (2011). An automatic classification of urban texture: form and compactness of morphological homogeneous structures in Barcelona. In *51st European Congress of the Regional Science Association International*, pages 1–20.

Congalton, R. and Green, K. (2002). *Assessing the accuracy of remotely sensed data: principles and practices*. Mapping Science Series. Taylor & Francis.

Cortes, C. and Vapnik, V. (1995). Support vector networks. *Machine Learning*, 20(3):273–297.

Cristianini, N. and Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge University Press, 1 edition.

Cutler, D., Edwards Jr, T., Beard, K., Cutler, A., Hess, K., Gibson, J., and Lawler, J. (2007). Random forests for classification in ecology. *Ecology*, 88(11):2783–92.

Dash, M. and Liu, H. (1997). Feature selection for classification. *Intelligent Data Analysis*, 1:131–156.

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

Douglas, J. (2007). Physical vulnerability modelling in natural hazard risk assessment. *Natural Hazards and Earth System Science*, 7(2):283–288.

Duch, W. (2006). Filter methods. In Guyon, I., Gunn, S., Nikravesh, M., and Zadeh, L. A., editors, *Feature Extraction: Foundations and Applications (Studies in Fuzziness and Soft Computing)*, chapter 3, pages 90–117. Springer, New York.

Duchêne, C., Bard, S., Barillot, X., Ruas, A., Trevisan, J., and Holzapfel, F. (2003). Quantitative an qualitative description of building orientation. In *fifth workshop on progress in automated map generalisation*, Paris, France.

Duda, R., Hart, P., and Stork, D. (2001). *Pattern classification*. Pattern Classification and Scene Analysis: Pattern Classification. John Wiley & Sons.

EERI (2009). Earthquake Engeneering Research Institute. Learnig from Earthquakes – The M_w 7.6 Western Sumatra Earthquake of September 30, 2009. *EERI Special Earthquake Report*, Dec. 2009. USA.

Fayyad, U. M. and Irani, K. B. (1993). Multi-interval discretization of continuous-valued attributes for classification learning. In *IJCAI*, pages 1022–1029.

FEMA (2012). HAZUS MH-2.1 multi-hazard loss estimation methodology – earthquake model. *Technical Manual*. Department of Homeland Security, Federal Emergency Management Agency, Mitigation Division. Washington, D.C.

Folleco, A., Khoshgoftaar, T., Van Hulse, J., and Bullard, L. (2008). Identifying learners robust to low quality data. In *IEEE International Conference on Information Reuse and Integration, 2008. IRI 2008.*, pages 190–195.

Foody, G. and Mathur, A. (2004). A relative evaluation of multiclass image classification by support vector machines. *IEEE Transactions on Geoscience and Remote Sensing,*, 42(6):1335–1343.

Foody, G. M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, 80(1):185–201.

Geiß, C. and Taubenböck, H. (2012). Remote sensing contributing to assess earthquake risk: from a literature review towards a roadmap. *Natural Hazards*. Doi: 10.1007/s11069-012-0322-2.

Geiß, C., Taubenböck, H., Tyagunov, S., Tisch, A., Post, J., and Lakes, T. (2013). Assessment of seismic building vulnerability from space. *Earthquake Spectra*, Accepted for publication.

GEM (2013). Global earthquake model – openquake. Url: http://www.globalquakemodel.org/openquake/about/, accessed 03.04.2013,.

Genuer, R., Poggi, J.-M., and Tuleau, C. (2008). Random forests: some methodological insights. Rapport de recherche RR-6729, INRIA.

Gislason, P. O., Benediktsson, J. A., and Sveinsson, J. R. (2006). Random forests for land cover classification. *Pattern Recognition Letters*, 27(4):294 – 300. Pattern Recognition in Remote Sensing (PRRS 2004).

Gualtieri, J. A. (2009). The support vector machine (svm) algorithm for supervised classification of hyperspectral remote sensing data. In *Kernel Methods for Remote Sensing Data Analysis*, pages 49–83. John Wiley & Sons.

Gualtieri, J. A. and Cromp, R. F. (1998). Support vector machines for hyperspectral remote sensing classification. In *Proceedings of the SPIE*, volume 3584, pages 221–232.

Guo, L., Chehata, N., Mallet, C., and Boukir, S. (2011). Relevance of airborne lidar and multispectral image data for urban scene classification using random forests. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(1):56 – 66.

Guyon, I. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3:1157–1182.

Guyon, I., Gunn, S., Nikravesh, M., and Zadeh, L. A., editors (2006). *Feature extraction: foundations and applications (studies in fuzziness and soft computing)*. Springer, New York, Secaucus, NJ, USA.

Hall, M., Eibe, F., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. (2009). The weka data mining software: An update. *SIGKDD Explorations*, 11(1). Weka 3 (2013). Software available at http://www.cs.waikato.ac.nz/ml/weka, accessed 23.03.2013.

Hall, M. and Holmes, G. (2003). Benchmarking attribute selection techniques for discrete class data mining. *IEEE Transactions on Knowledge and Data Engineering,*, 15(6):1437–1447.

Hall, M. A. (1999). *Correlation-based Feature Selection for Machine Learning*. PhD thesis, Department of Computer Science, The University of Waikato, Hamilton, New Zealand.

Hall, M. A. (2000). Correlation-based feature selection for discrete and numeric class machine learning. In *ICML '00: Proceedings of the Seventeenth International Conference on Machine Learning*, pages 359–366, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Han, J. and Kamber, M. (2006). *Data mining: concepts and techniques*. Elsevier, Morgan Kaufmann, 2nd edition.

Haralick, R., Shanmugam, K., and Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics,*, SMC-3(6):610–621.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning: data mining, inference and prediction*. Springer, 2nd edition.

He, H. and Garcia, E. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9):1263–1284.

Hirt, C., Filmer, M. S., and Featherstone, W. E. (2010). Comparison and validation of the recent freely available aster-gdem ver1, srtm ver4.1 and geodata dem-9s ver3 digital elevation models over australia. *Australian Journal of Earth Sciences*, 57(3):337–347.

Hsu, C.-W., Chang, C.-C., and Lin, C.-J. (2010). A practical guide to support vector classification. Technical report, Department of Computer Science, National Taiwan University.

Hsu, C.-W. and Lin, C.-J. (2002). A comparison of methods for multi-class support vector machines. *IEEE Transactions on Neural Networks*, 13(2):415–425.

Huang, C., Davis, L. S., and Townshend, J. R. G. (2002). An assessment of support vector machines for land cover classification. *International Journal of Remote Sensing*, 23(4):725–749.

Hughes, G. (1968). On the mean accuracy of statistical pattern recognizers. *IEEE Transactions on Information Theory*, 14(1):55–63.

Irish, R. (2008). Landsat 7 science data user's handbook. *Landsat Project Science Office*, 21:186.

Irsyam, M., Dangkua, D., Hendriyawan, Hoedajanto, D., Hutapea, B., Kertapati, E., Boen, T., and Petersen, M. (2008). Proposed seismic hazard maps of sumatra and java islands and microzonation study of jakarta city, indonesia. *Journal of Earth System Science*, 117:865–878.

Japkowicz, N. and Stephen, S. (2002). The class imbalance problem: A systematic study. *Intell. Data Anal.*, 6(5):429–449.

Kira, K. and Rendell, L. A. (1992). A practical approach to feature selection. In *Proceedings of the ninth international workshop on Machine learning*, ML92, pages 249–256, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.

Knerr, S., Personnaz, L., and Dreyfuss, G. (1990). Single-layer learning revisited: a stepwise procedure for building and training a neural network. In *Neurocomputing: Algorithm, Architectures and Applications*. Springer.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1137–1143. Morgan Kaufmann.

Kohavi, R. and John, G. H. (1997). Wrappers for feature subset selection. *Artificial intelligence*, 97(1):273–324.

Kononenko, I. (1994). Estimating attributes: Analysis and extensions of relief. In Bergadano, F. and Raedt, L. D., editors, *European Conference on Machine Learning*, pages 171–182. Springer.

Lawrence, R. L., Wood, S. D., and Sheley, R. L. (2006). Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications (randomForest). *Remote Sensing of Environment*, 100:356–362.

Liaw, A. and Wiener, M. (2002). Classification and regression by randomForest. *R News: The Newsletter of the R Project*, 2(3):18–22.

Liu, H., Li, J., and Wong, L. (2002). A comparative study on feature selection and classification methods using gene expression profiles and proteomic patterns. In Lathrop, R., Nakai, K., Miyano, S., Takagi, T., and Kanehisa, M., editors, *Genome Informatics 2002*, volume 13, pages 51–60+. Universal Academy Press, Tokyo, 2002.

Liu, H. and Yu, L. (2005). Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on Knowledge and Data Engineering,*, 17(4):491–502.

MacEachren, A. M. (1985). Compactness of geographic shape: Comparison and evaluation of measure. *Geografiska Annaler.*, 67(1):53–67.

McCaffrey, R. (2009). The tectonic framework of the Sumatran subduction zone. *Annual Review of Earth and Planetary Sciences*, 37:345–366.

McCloskey, J., Antonioli, A., Piatanesi, A., Sieh, K., Steacy, S., Nalbant, S., Cocco, M., Giunchi, C., Huang, J., and Dunlop, P. (2008). Tsunami threat in the Indian Ocean from a future megathrust earthquake west of Sumatra. *Earth and Planetary Science Letters*, 265(1–2):61–81.

McCloskey, J., Lange, D., Tilmann, F., Nalbant, S. S., Bell, A. F., Natawidjaja, D. H., and Rietbrock, A. (2010). The September 2009 Padang earthquake. *Nature Geoscience*, 3(2):70–71.

Melgani, F. and Bruzzone, L. (2004). Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing,*, 42(8):1778–1790.

Michel, G. W., Yu, Y. Q., Zhu, S. Y., Reigber, C., Becker, M., Reinhart, E., Simons, W., Ambrosius, B., Vigny, C., Chamot-Rooke, N., Pichon, X. L., Morgan, P., and Matheussen, S. (2001). Crustal motion and block behaviour in se-asia from gps measurements. *Earth and Planetary Science Letters*, 187(3–4):239–244.

Mountrakis, G., Im, J., and Ogole, C. (2011). Support vector machines in remote sensing: A review. *Isprs Journal of Photogrammetry and Remote Sensing*, 66:247–259.

Mück, M., Taubenböck, H., Post, J., Wegscheider, S., Strunz, G., Sumaryono, S., and Ismail, F. (2012). Assessing building vulnerability to earthquake and tsunami hazard using remotely sensed data. *Natural Hazards*. Doi: 10.1007/s11069-012-0481-1.

Muñoz-Marí, J., Bovolo, F., Gomez-Chova, L., Bruzzone, L., and Camp-Valls, G. (2010). Semisupervised one-class support vector machines for classification of remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing,*, 48(8):3188–3197.

Nalbant, S. S., Steacy, S., Sieh, K., Natawidjaja, D., and McCloskey, J. (2005). Seismology: Earthquake risk on the Sunda trench. *Nature*, 435(7043):756–757.

Nguyen, G. H., Bouzerdoum, A., and Phung, S. L. (2009). Learning pattern classification tasks with imbalanced data sets. In Yin, P.-Y., editor, *Pattern Recognition*. InTech. Url: http://cdn.intechopen.com/pdfs/9154/InTech-Learning_pattern_classification_tasks_with_imbalanced_data_sets.pdf, accessed 25.03.2013.

Novack, T., Esch, T., Kux, H., and Stilla, U. (2011). Machine learning comparison between worldview-2 and quickbird-2-simulated imagery regarding object-based urban land cover classification. *Remote Sensing*, 3(10):2263–2282.

Oliveira, C. S., Roca, A., and Goula, X. (2006). Assessing and managing earthquake risk. In Oliveira, C. S., Roca, A., and Goula, X., editors, *Assessing and managing earthquake risk*, chapter 1, pages 1–14. Springer.

Pacifici, F., Chini, M., and Emery, W. J. (2009). A neural network approach using multi-scale textural metrics from very high-resolution panchromatic imagery for urban land-use classification. *Remote Sensing of Environment*, 113(6):1276 – 1292.

Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26(1):217–222.

Pal, M. and Mather, P. M. (2005). Support vector machines for classification in remote sensing. *International Journal of Remote Sensing*, 26(5):1007–1011.

Pemerintah Kota Padang (2012). Padang local government. Url: http://www.padang.go.id, accessed 19.11.2012.

Petersen, M. D., Dewey, J., Hartzell, S., Mueller, C., Harmsen, S., Frankel, A. D., and Rukstales, K. (2004). Probabilistic seismic hazard analysis for Sumatra, Indonesia and across the Southern Malaysian Peninsula. *Tectonophysics*, 390(1–4):141–158.

Pittore, M. and Wieland, M. (2012). Toward a rapid probabilistic seismic vulnerability assessment using satellite and ground-based remote sensing. *Natural Hazards*. Doi: 10.1007/s11069-012-0475-z.

Platt, J. C., Cristianini, N., and Shawe-Taylor, J. (2000). Large margin DAGs for multiclass classification. In *Analysis*, volume 12, pages 547–553. MIT Press.

R Core Team (2013). *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Refaeilzadeh, P., Tang, L., and Liu, H. (2009). Cross-validation. In Liu, L. and Özsu, M. T., editors, *Encyclopedia of Database Systems*, pages 532–538. Springer.

Richards, J. and Jia, X. (2006). *Remote sensing digital image analysis: an introduction*. Springer.

Richter, R. (1996). A spatially adaptive fast atmospheric correction algorithm. *International Journal of Remote Sensing*, 17:1201–1214.

Robnik-Šikonja, M. and Kononenko, I. (2003). Theoretical and empirical analysis of relieff and rrelieff. *Machine Learning*, 53(1-2):23–69.

Rodriguez-Galiano, V., Chica-Olmo, M., Abarca-Hernandez, F., Atkinson, P., and Jeganathan, C. (2012a). Random forest classification of mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. *Remote Sensing of Environment*, 121(0):93 – 107.

Rodriguez-Galiano, V., Ghimire, B., Rogan, J., Chica-Olmo, M., and Rigol-Sanchez, J. (2012b). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67(0):93 – 104.

Rusnardi, P. R., Kiyono, J., Ono, Y., and Parajuli, H. R. (2012). Seismic hazard analysis for Indonesia. *Journal of Natural Disaster science*, 33(2):59–70.

Sarabandi, P. and Kiremidjian, A. S. (2007). Development of algorithms for building inventory compilation through remote sensing and statistical inferencing. Technical Report 158, The John A. Blume Earthquake Engeneering Center. Stanford University.

Scawthorn, C. (2008). A brief history of seismic risk assessment. In Bostrom, A., French, S., and Gottlieb, S., editors, *Risk assessment, modeling and decision support: strategic decisions*, Risk, Governance and Society, 14. Springer.

Schmidt, J., Matcham, I., Reese, S., King, A., Bell, R., Henderson, R., Smart, G., Cousins, J., Smith, W., and Heron, D. (2011). Quantitative multi-risk analysis for natural hazards: a framework for multi-risk modelling. *Natural Hazards*, 58(3):1169–1192.

Schölkopf, B., Platt, J. C., Shawe-Taylor, J. C., Smola, A. J., and Williamson, R. C. (2001). Estimating the support of a high-dimensional distribution. *Neural Comput.*, 13(7):1443–1471.

Schölkopf, B. and Smola, A. J. (2001). *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT Press, Cambridge, MA, USA.

Schölkopf, B., Smola, A. J., Williamson, R. C., and Bartlett, P. L. (2000a). New support vector algorithms. *Neural Comput.*, 12(5):1207–1245.

Schölkopf, B., Williamson, R. C., Smola, A. J., Shawe-Taylor, J., and Platt, J. (2000b). Support vector method for novelty detection. *Advances in neural information processing systems*, 12(3):582–588.

Sengara, I., Suarjana, M., Beetham, D., Corby, N., Edwards, M., Griffith, M., Wehner, M., and Weller, R. (2010). The 30th september 2009 west sumatra earthquake: Padang region damage survey. *Geoscience Australia*, Record 2010/44:201 pp.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell system technical journal*, 27.

Sieh, K., Natawidjaja, D. H., Meltzner, A. J., Shen, C.-C., Cheng, H., Li, K.-S., Suwargadi, B. W., Galetzka, J., Philibosian, B., and Edwards, R. L. (2008). Earthquake supercycles inferred from sea-level changes recorded in the corals of West Sumatra. *Science*, 322(5908):1674–1678.

Sirmacek, B., Taubenböck, H., Reinartz, P., and Ehlers, M. (2012). Performance evaluation for 3-d city model generation of six different dsms from air- and spaceborne sensors. *Selected Topics in Applied Earth Observations and Remote Sensing, IEEE Journal of*, 5(1):59–70.

Steiniger, S., Lange, T., Burghardt, D., and Weibel, R. (2008). An approach for the classification of urban building structures based on discriminant analysis techniques. *Transactions in GIS*, 12(1):31–59.

Stumpf, A. and Kerle, N. (2011). Object-oriented mapping of landslides using random forests. *Remote Sensing of Environment*, 115(10):2564 – 2577.

Svetnik, V., Liaw, A., Tong, C., and Culberson, J. C. (2003). Random forest: A classification and regression tool for compound classification and QSAR modeling. *Journal of Chemical Information and Computer Sciences*, 43(6):1947–1958.

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. (2009a). "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(4):1509–1528.

Taubenböck, H., Post, J., Kiefl, R., Roth, A., Ismail, F., Strunz, G., and Dech, S. (2008a). Risk and vulnerability assessment to tsunami hazard using very high resolution satellite data – The case study of Padang, Indonesia. In Carsten, J., editor, *Proceedings of the EARSeL Joint Workshop: Remote Sensing: New challanges of high resolution*.

Taubenböck, H., Post, J., Roth, A., G, S., Kiefl, R., Dech, S., and Ismail, F. (2008b). Multi-scale assessment of population distribution utilizing remotely sensed data – The case study Padang, West Sumatra, Indonesia. In *International Conference on Tsunami Warning (ICTW)*. Bali, Indonesia.

Taubenböck, H., Roth, A., Dech, S., Mehl, H., Münich, J., Stempniewski, L., and Zschau, J. (2009b). Assessing building vulnerability using synergistically remote sensing and civil engineering. In Kreck, A., Rumor, M., Zlatnova, S., and Fendel, E., editors, *Urban and Regional Data Management*, pages 287–300. Taylor & Francis Group, London.

Taubenböck, H., Wurm, M., Setiadi, N., Gebert, N., Roth, A., Strunz, G., Birkmann, J., and Dech, S. (2009c). Integrating remote sensing and social science - the correlation of urban morphology with socioeconomic parameters. In *Urban Remote Sensing Join Event*. Shanghai, China.

Tax, D. M. J. (2001). *One-class classification – concept-learning in the absence of counter-examples*. PhD thesis, Technische Universiteit Delft. Netherlands.

Tax, D. M. J. and Duin, R. P. W. (1999). Support vector domain description. *Pattern Recognition Letters*, 20:1191–1199.

Tzotsos, A. (2006). A support vector machine approach for object based image analysis. In *ISPRS Archives – Volume XXXVI-4/C42, 2006 – Bridging Remote Sensing and GIS 1st International Conference on Object-based Image Analysis (OBIA 2006). Salzburg.*

UNDESA (2011). United Nations Department for Economic and Social Affairs. Population Division. World Population Prospects – The 2010 Revision. Highlights and Advance Tables. ESA/P/WP.220, New York. Url: http://esa.un.org/unpd/wpp/Documentation/pdf/WPP2010_Highlights.pdf, accessed 17.11.2012.

UNDESA (2012). United Nations Department for Economic and Social Affairs. Population Division. World urbanization prospects – The 2011 revision. Final Report with Annex Tables, New York. Url: http://esa.un.org/unpd/wup/pdf/FINAL-FINAL_REPORT%20WUP2011_Annextables_01Aug2012_Final.pdf, accessed 17.11.2012.

UNISDR/WMO (2012). United Nations Office for Disaster Risk Reduktion / World Meteorological Association: UN System Task Team on the Post-2015 UN development agenda: Disaster risk and resilience – Thematic Think Piece. Url: http://www.un.org/millenniumgoals/pdf/Think%20Pieces/3_disaster_risk_resilience.pdf, accessed 17.11.2012.

USGS (2009). U.S. Geological Survey. ShakeMap Southern Sumatra, Indonesia. Url: http://earthquake.usgs.gov/earthquakes/shakemap/global/shake/2009mebz accessed 17.11.2012.

USGS (2011). U.S. Geological Survey. ShakeMap Scientific Background. Url: http://earthquake.usgs.gov/earthquakes/shakemap/background.php, accessed 17.11.2012.

USGS (2012). U.S. Geological Survey. Earthquake Report: Magnitude 7.6 - Southern Sumatra, Indonesia. Url: http://earthquake.usgs.gov/earthquakes/eqinthenews/2009/us2009mebz, accessed 17.11.2012.

Van Hulse, J., Khoshgoftaar, T. M., Napolitano, A., and Wald, R. (2009). Feature selection with high-dimensional imbalanced data. In *Proceedings of the 2009 IEEE International Conference on Data Mining Workshops*, ICDMW '09, pages 507–514, Washington, DC, USA. IEEE Computer Society.

Vapnik, V. N. (1995). *The nature of statistical learning theory*. Springer.

Vapnik, V. N. (1998). *Statistical learning theory*. Wiley, New York, 1 edition.

Waske, B. and Braun, M. (2009). Classifier ensembles for land cover mapping using multitemporal sar imagery. *ISPRS Journal of Photogrammetry and Remote Sensing*, 64(5):450 – 457.

Waske, B., Fauvel, M., Benediktsson, J. A., and Chanussot, J. (2009). Machine learning techniques in remote sensing data analysis. In Camps-Valls, G. and Bruzzone, L., editors, *Kernel Methods for Remote Sensing Data Analysis*, chapter 1, pages 1–24. John Wiley & Sons.

Weszka, J. S., Dyer, C. R., and Rosenfeld, A. (1976). A comparative study of texture measures for terrain classification. *IEEE Transactions on Systems, Man and Cybernetics,*, SMC-6(4):269–285.

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.

Wilkinson, S. M., Alarcon, J. E., Mulyani, R., Whittle, J., and Chian, S. C. (2012). Observations of damage to buildings from Mw 7.6 Padang earthquake of 30 September 2009. *Natural Hazards*, 63:521–547.

Witten, I. H., Frank, E., and Hall, M. A. (2011). *Data mining: practical machine learning tools and techniques*. Elsevier, Morgan Kaufmann, Amsterdam, 3 edition.