A hybrid soil mapping approach using SOTER, SoLIM and Classification trees

Ulrich Schuler[A], Petra Erbe[B], Karl Stahr[C] and Ludger Herrmann[C]

[A]Federal Institute for Geosciences and Natural Resources (BGR), B2.2 Spatial Information Soil and Water, Stilleweg 2, D-30655 Hanover, Germany, Email ulrich.schuler@bgr.de
[B]The Uplands Program, Hohenheim Office, Faculty of Agriculture, Chiang Mai University, Chiang Mai 50200, Thailand, Email petra.erbe@gmail.com
[C]University of Hohenheim, Institute of Soil Science and Land Evaluation (310), D-70593 Stuttgart, Germany, Email kstahr@uni-hohenheim.de, herrmann@uni-hohenheim.de

Abstract
The investigated mapping approach combines different (soil) mapping concepts including SOTER (FAO 2005), SoLIM (Zhu *et al.* 2001), expert knowledge and classification trees (Breiman *et al.* 1984). North-western Thailand was stratified into SOTER terrains using DEM data and geological information. Detailed soil mapping was tested in three pilot areas within different SOTER terrains in north-western Thailand. Therefore, classification trees were calibrated with soil data from transect-based sampling points. Maps of the pilot areas were created by implementing classification rules derived from Classification Tree (CART algorithm) and expert knowledge in SoLIM. Validation was performed using soil reference maps and independent sampling points. The reference soil maps contain information from transects, sampling points along local trails, sampling points for areas of low point density or high soil variability, reference profiles, LANDSAT and SPOT images, and topographic information. The mean sampling distance was around 200 m. The reference soil maps were manually created based on expert knowledge using ArcGIS 9.2 software. The hybrid mapping approach based soil maps showed high correspondence with the respective reference soil map and a very high degree of matching with independent sampling points. This hybrid mapping approach seems to be very useful for reconnaissance soil mapping.

Key Words
Transect mapping, reconnaissance soil mapping, and expert knowledge, northern Thailand.

Introduction
Many developing and emerging countries face unsustainable agricultural land use. The obligatory land use planning requires regional soil information. However, the data density of soil information is commonly rather low for such areas. Intensive soil mapping is very costly and time consuming. Therefore, the development of alternative quick, cheap, but sufficiently accurate mapping methods is indispensable. The main objective of this study was to develop an efficient and transparent soil mapping approach, facilitating the regionalization of soil information detailed enough to be used as baseline information for land use planning and modelling. The approach tested blends different mapping concepts including SOTER (FAO 2005), Classification Tree (CART algorithm) (Breiman *et al.* 1984), and SoLIM (Zhu *et al.* 2001). This mapping approach was tested in three pilot areas within different landscape units, reflecting together the major landscape variability in north-western Thailand.

Methods
Research area
The entire research area is located in north-western Thailand covering the area between the border to Myanmar in the north and west and the basin of Chiang Mai in the west and a line south of Doi Inthanon (Figure 1). Detailed soil mapping was carried out in three different pilot areas representing together the major landscapes of the region. The resulting maps (as well as all other available data) were used as reference for the studied mapping approach. The Mae Sa Mai pilot area (10.5 km^2) represents a SOTER terrain with high gradient mountains mainly consisting of granite and gneiss. The Huay Bong pilot area (6.8 km^2) has high gradient mountains with sandstone as major lithology. The Bor Krai pilot area (8.5 km^2) shows strong karst features and hence limestone as major lithology.

Materials
Digitized topographic and geological maps, aerial photographs as well as LANDSAT and SPOT images were available as baseline information for the field surveys. A topographic map with a scale of 1:50,000 compiled by the Royal Thai Survey Department (1976) provided contour lines with 20 m intervals. The

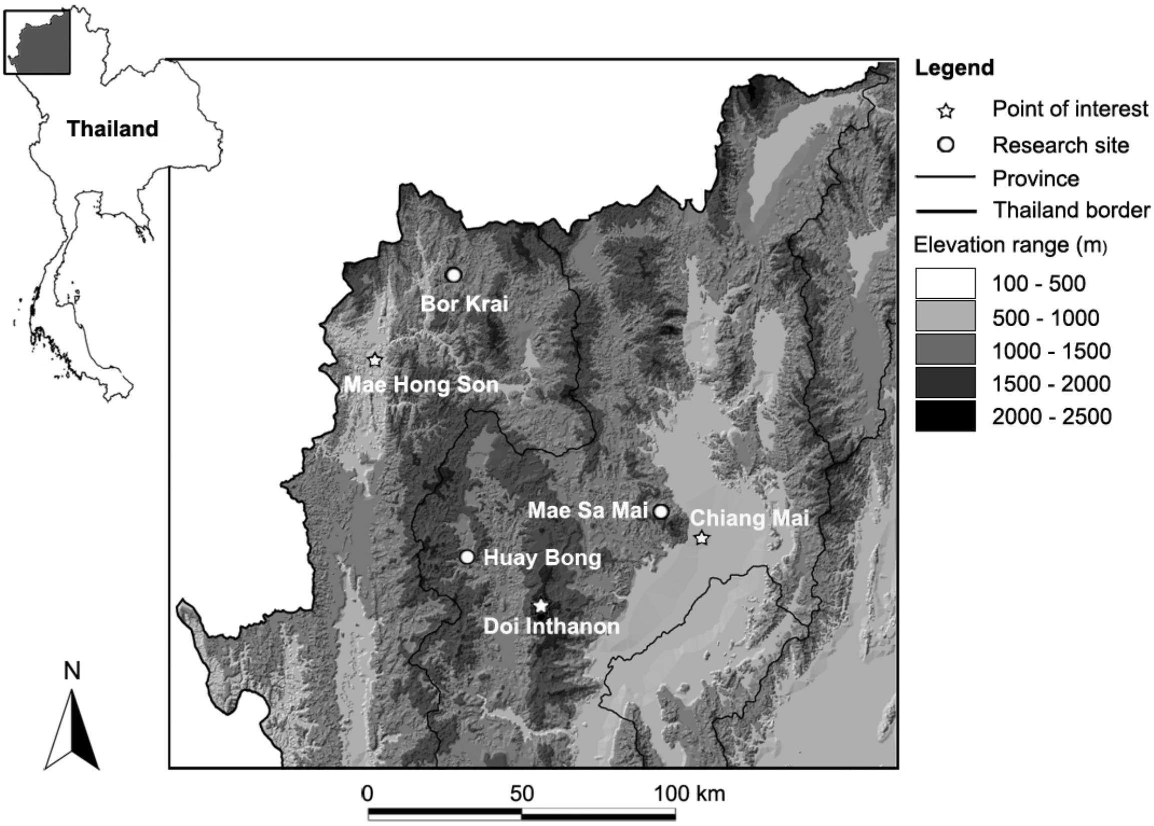

Figure 1. The research area including the three pilot areas in northwestern Thailand.

geological map has a scale of 1:250,000 and was compiled by the German Geological Mission (1979). The aerial photographs (taken in November 1999) for the Mae Sa watershed have a scale of 1:15,000 and were provided by the Royal Thai Survey Department (1999). The LANDSAT 7 ETM+ image was provided by 'Global Land Cover Facility - GLCF' (GLFC 2007). This image was taken on 5 March 2000. The SPOT 5 images were provided by 'Geo-Informatic and Space Technology Agency – GISTDA' (GISTDA 2007). The SPOT 5 image covering the Mae Sa Mai area was taken on 6 November 2006. The SPOT 5 images for Huay Bong and Bor Krai were taken on 22 February 2007, and 1 December 2006, respectively. During the field trips, a hand-held Garmin GPS III (Garmin, USA) was used to obtain coordinates of the observation points. The evaluation of the field and laboratory data was carried out using MS Access 2003 (Microsoft, USA), Past 1.81 (Paleontological Museum Oslo, Norway) and ArcGIS 9.2 (ESRI, USA) software.

Mapping

The tested hybrid mapping approach consists of the following steps:
1.) Delineation of SOTER terrains in order to stratify different landscapes as a baseline for Classification Tree and expert knowledge based soil mapping. The SOTER terrains were delineated blending geological information with stratified terrain classes. These terrain classes were beforehand generated blending classes of slope, relief intensity, dissection and hypsometry (Dobos *et al.* 2005).
2.) Transect based soil mapping in three pilot areas within different SOTER terrains. The transect lines, which cover different geomorphic units, parent materials and land use types, facilitate the detection of rules for local soil distribution (Schlichting *et al.* 1995). The number of sampling points used for transect mapping was 199 in Mae Sa Mai, 170 in Huay Bong and 322 in Bor Krai. Due to the rugged, steep, and partly inaccessible terrain the sampling was mainly performed along local trails.
3.) Using CART algorithm based classification trees (Breiman *et al.* 1984) to derive soil distribution rules for each SOTER terrain. The classification trees were generated using the SPSS 16.0 software package. The used input data were elevation, slope, curvature, aspect, petrography, LANDSAT 7 (bands 1-8), and SPOT 5 (bands 1-4) extracted at the locations of all transect-based training points.
4.) Implementation of soil distribution rules into the SoLIM model.
5.) Modifying and adding soil distribution rules based on expert knowledge, where soil distribution rules with the Classification Tree could not be generated due to insufficient soil information.
6.) Generating a soil map by running the SoLIM model.

Validation
Soil maps based on the hybrid mapping approach were validated in comparison with reference soil maps and with independent sampling points. The reference maps included the maximum amount of information available. All reference soil maps contain information from soil catenas, sampling points along local trails, and for areas of low point density or high soil variability, reference profiles, LANDSAT and SPOT images, and topographic information. The mean sampling distance was around approximately 200 m. Finally, the reference soil maps were manually created based on expert knowledge using ArcGIS 9.2 software. Additionally, for each area 15% of all sample points were randomly selected as validation points. These points were exclusively used for validation. In Mae Sa Mai 36, in Huay Bong 30, and in Bor Krai 55 validation points were used.

Results
For more than 75% of the soils clay illuviation was identified as the major soil forming process. Accordingly, soils were mainly classified as Alisols and Acrisols. Less frequent soil types were Cambisols, Umbrisols and Regosols. The remaining soil types mapped (Anthrosols, Chernozems, Ferralsols, Fluvisols, Gleysols, Leptosols, and Technosols) represent less than 2% of all soils. Reference soil maps showed that the soil cover in the Mae Sa Mai area is dominated by Acrisols (84%) followed by Cambisols (9%), Umbrisols (4%), and Technosols (2%). Anthrosols, Chernozems, Gleysols, Leptosols, and water bodies were present in the remaining area. In contrast to Mae Sa Mai, Alisols prevail in Huay Bong (77%), followed by Cambisols (13%), Regosols (9%), Leptosols (2%), and Fluvisols (0.1%). In Bor Krai the predominant mapping units were Alisols (64%), Acrisols (27%), and limestone outcrops (5%), while Cambisols, Chernozems, Ferralsols, Fluvisols, Gleysols, Leptosols, Luvisols, and Umbrisols comprise less than 1%. In all three areas, the classification tree-based maps corresponded to at least 70 % with the respective reference soil map. The correspondence to the reference map was 79 % at Mae Sa Mai, 70 % at Huay Bong, and 84 % at Bor Krai. The validation with independent sampling points showed matches of 77 % at Mae Sa, 73 % at Huay Bong and 74 % at Bor Krai.

Discussion
The tested hybrid mapping approach combines the advantages of different mapping concepts. The SOTER concept (FAO 1995) provides a guideline to stratify an area into homogenous landscape units or so called SOTER terrains. Each SOTER terrain has its characteristic topography, major lithology, climate, and soil distribution. The rules of soil distribution within the SOTER terrain units can be best explored with classification trees (Breiman *et al.* 1984). Finally, the SoLIM model (Zhu *et al.* 2001) enables the realization of classification rules and the combination of soil information of the different SOTER terrains. The major advantage of the tested hybrid mapping approach in comparison with other digital soil mapping approaches like artificial neural networks (Behrens *et al.* 2005), nominal logistic regression (Debella-Gilo and Etzelmüller 2009) or random forest (Breiman 2001) is its transparency. Each soil distribution rule is well documented and can be modified later. Further it is possible to map an area stepwise or to harmonize different map sheets.

Conclusions
The tested hybrid mapping approach is highly suitable for soil mapping especially in developing and emerging countries. The documentation of soil distribution rules facilitates a continuous improvement of soil maps by different projects or editors. Soil maps based on this hybrid mapping approach can be easily harmonized.

Acknowledgements
This research, which was carried out within the framework of the Special Research Program 564, was funded by the Deutsche Forschungsgemeinschaft (DFG) whose support is gratefully acknowledged. Thankfully acknowledged is Dirk Euler for his expert advice regarding statistical questions. Thanks go to Dr. J.F. Maxwell for critical reviewing of the manuscript and for giving his expert advice. We are very grateful to the villagers of Bor Krai, Huay Bong, and Mae Sa Mai for their support.

References
Behrens T, Förster H, Scholten T, Steinrücken U, Spies E.-D, Goldschmitt M (2005) Digital soil mapping using artificial neural networks. *Plant. Nutr. Soil Sci.* **168,** 21-33.

Breiman L, Friedman JH, Olshen RA, Stone CJ (1984) Classification and regression trees. Wadsworth International group, California, United States of America 358 pp.

Breiman L (2001) Random forests. *Machine Learning* **45,** 5-32.

FAO (1995) 'Global and national soils and terrain digital databases (SOTER)'. World Soil Resources Reports 74. (FAO: Rome).

IUSS Working Group WRB (2006) World reference base for soil resources 2006. World Soil Resources Reports 103, Rome.

German Geological Mission (1979) 'Geological map of northern Thailand 1:250.000'. (Federal Institute for Geosciences and Natural Resources: Germany).

Debella-Gilo M, Etzelmüller B (2009) Spatial prediction of soil classes using digital terrain analysis and multinomial logistic regression modeling integrated in GIS: Examples from Vestfold County, Norway. *Catena* **77,** 8-18.

Dobos E, Daroussin J, Montanarella L (2005) 'An SRTM-based procedure to delineate SOTER Terrain Units on 1:1 and 1:5 million scales'. (EUR 21571 EN, Institute for Environment and Sustainability, Joint Research Centre: Ispra).

GISTDA (2007a) 'SPOT 5 image; SPOT-5 K-J: 255-312, Date: 06-11-2006'. http://www.gistda.or.th/Gistda/HtmlGistda/Html/index2.htm

GISTDA (2007b) 'SPOT 5 image; SPOT-5 K-J: 255-311, Date: 01-12-2006'. http://www.gistda.or.th/Gistda/HtmlGistda/Html/index2.htm

GISTDA (2007c) 'SPOT 5 image; SPOT-5 K-J: 255-313, Date: 22-12-2007'. http://www.gistda.or.th/Gistda/HtmlGistda/Html/index2.htm

GLFC (2007) 'Landsat ETM+, WRS-2, Path 131, Row 047, Date: 2000-03-05, EarthSat, Ortho, GeoCover Myanmar (Burma), Thailand'. http://glcf.umiacs.umd.ed

Herrmann L (Editor) (2005) ‚Das kleine Bodenkochbuch (Version 2005)'. (Institute of Soil Science and Land Evaluation, University of Hohenheim: Stuttgart).

Royal Thai Survey Department (1976) 'Topographic Maps, Scale 1:50.000'. (Bangkok).

Royal Thai Survey Department (1999) 'Aerial photographs of the Mae Sa watershed probably taken in November 1999, Scale 1:15.000'. (Bangkok).

Schlichting E, Blume HP, Stahr K (1995) ‚Bodenkundliches Praktikum'. Pareys Studientexte, Berlin 81, 1-295.

Zhu AX, Hudson B, Burt J, Lubich K, Simonson, D, (2001) Soil mapping using GIS, expert knowledge, and fuzzy logic. *Science Society of America Journal* **65,** 1463-1472.