

IMPACT ASSESSMENT OF SOIL INFORMATION AND LAND COVER CHANGE ON FLASH FLOOD MODELLING ON A WATERSHED SCALE

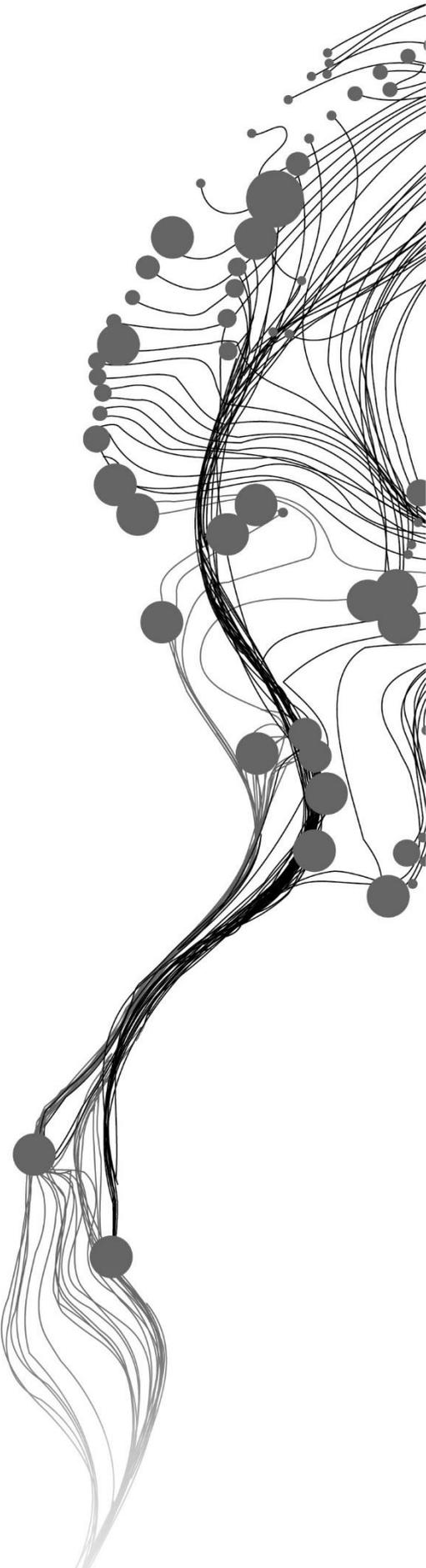
KASIMIR ALEXANDER ORLOWSKI

June, 2019

SUPERVISORS:

Dr. D.B.P. Shrestha

Prof. Dr. V.G. Jetten



IMPACT ASSESMENT OF SOIL INFORMATION AND LAND COVER CHANGE ON FLASH FLOOD MODELLING ON A WATERSHED SCALE

KASIMIR ALEXANDER ORLOWSKI
Enschede, The Netherlands, June, 2019

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.

Specialization: Applied Earth Sciences, with specialization in Natural Hazards, Risk and Engineering

SUPERVISORS:

Dr. D.B.P. Shrestha
Prof. Dr. V.G. Jetten

THESIS ASSESSMENT BOARD:

Prof. Dr. F.D. van der Meer (Chair)
Dr. C. Mannaerts (External Examiner, ITC – WRS Department)

DISCLAIMER

This document describes work undertaken as part of a programme of study at the Faculty of Geo-Information Science and Earth Observation of the University of Twente. All views and opinions expressed therein remain the sole responsibility of the author, and do not necessarily represent those of the Faculty.

ABSTRACT

Flash floods frequencies and magnitudes are increasing, influenced by climate change and land cover alterations. To reduce losses by floods, hazard modelling is crucial. However, soil information which are vital for modelling are often missing or of insufficient quality. Therefore, the publicly available global digital soil database SOILGRIDS250m potentially represents a way to bridge the gap between data availability and data demand. So far, its applicability for local scale hydrological modelling has not been sufficiently investigated. In the light of that, this study focused on the analysis of two soil datasets, (I) detailed field data (FD) and (II) SoilGrids (SG) in order to assess their similarity and to evaluate the sensitivity of flood dynamics to their soil hydraulic properties (SHPs), and to different soil depths when applied in an integrated flood model. Furthermore, the effects of land cover change on runoff generation and flood behaviour were investigated. The first part of this research was dedicated to the soil data analysis. In the course of that, soil properties of FD and SG were examined independently in relation to land cover and the terrain. Subsequently, in a comparative assessment, the similarity between both datasets was quantified. The second part of the study focused on land cover mapping with Google Earth Engine, and foremost on the identification of land cover changes. In the third part, the preceding efforts were joined to build the input for the integrated flood model openLISEM. The model was applied with: (I) SHPs derived in the laboratory (FD); (II) SHPs predicted with pedotransfer functions (SG); and (III) with changing land cover information. Results indicate that the FD and SG do not share many commonalities. FD is highly influenced by land cover, whereas SG variability is limited throughout the watershed. Soil properties such as clay content, bulk density and soil organic matter were overestimated by SG. Therefore, the use of SG led to far-reaching consequences in the hydrology, including a considerable increase in flood extent, depth and duration. Increasing soil depth influenced both datasets similarly by promoting infiltration and reducing surface water. However, using FD, the flood dynamics were more sensitive to changes in soil depth. Changes in land cover were predominately represented by deforestation and subsequent fruit tree cultivation. Changed land cover information affected flood dynamics only minor, but an increase in runoff amounts was apparent. Quality issues of the digital elevation model, including errors in elevation and flow connectivity, impeded the model calibration efforts and led consequently to wrong flood patterns. However, as overall flood quantities are expected to be correct, the conclusions for the objectives set were not invalidated. Future research should continue exploring SG data applicability for hydrological modelling as it represents a valuable source of information. To be able to make a profound statement about SG quality, it is necessary to conduct studies in various regions of the world. This will help to investigate its quality dependency on factors such as terrain, climate and vegetation.

Keywords: Hydrological modelling, openLISEM, Flash flood, SoilGrids, Soil properties, Land cover change, Google Earth Engine, Thailand, Asia

ACKNOWLEDGEMENTS

I would like to express my profound gratitude to all three supervisors who accompanied me during this enriching journey, Dr. Dhruva Shrestha, Prof. Dr. Victor Jetten and Dr. Dinand Alkema. I have enjoyed learning from and with you very much, whether being in Thailand in the jungle, or in your office at ITC.

Further, it is worth to note that I'm very grateful to Bastian Bout who lent me his brain from time to time, and to my friends and companions Fernandes Fernandes Brenner and Felipe Augusto Fonseca Arévalo for their support.

Ms. Nanette Kingma, you always told us: “you are in the driving seat”. However, sometimes it was good that you helped us to navigate.

Furthermore, I express my sincere gratitude to my supervisors at the Asian Disaster Preparedness Center in Bangkok, Dr. Peeranan Towashiraporn and Prof. Dr. Dr. Farrukh Chishtie, who facilitated my graduate research within the SERVIR-Mekong project. Also, a big thanks goes to my dear colleagues and friends Dr. Ate Poortinga and Mr. Biplov Bhandari, you two taught me a lot and increased my fascination to RS once more, thank you!

Last but not least, I also would like to thank the staff of the Naresuan University – Phitsanulok, to provide us with the necessary laboratory equipment and guidance.

Kasimir Alexander Orłowski
Enschede, June 2019



LIST OF ABBREVIATION

ASF	-	Alaska Satellite Facility
BRDF	-	Bidirectional Reflectance Distribution Function
D_b	-	Bulk Density
DEM	-	Digital Elevation Model
DEM _m	-	DEM manipulated
DEM _o	-	DEM original
DEM _v	-	DEM without vegetation
EBBI	-	Enhanced Built-up and Bareness Index
ETM	-	Enhanced Thematic Mapper
EVI	-	Enhanced Vegetation Index
FAO	-	Food and Agriculture Organization of the United Nations
FD	-	Field Data
GEE	-	Google Earth Engine
GLAS	-	Geoscience Laser Altimeter System
IBI	-	Index-based Built-up Index
IDF	-	Intensity Duration Frequency
ISRIC		International Soil Reference and Information Centre
K_s	-	Saturated Hydraulic Conductivity
LAI	-	Leaf Area Index
LCCS	-	Land Cover Classification System
LOI	-	Loss-on-Ignition
MNDWI	-	Modified Normalised Difference Water Index
NBR	-	Normalised Burn Ratio
ND	-	Normalised Differences
NDBI	-	Normalised Difference Built-up Index
NDVI	-	Normalised Difference Vegetation Index
NDWI	-	Normalised Difference Water Index
NIR	-	Near Infrared

OLI	-	Operational Land Imager
OSM	-	OpenStreetMap
PCA	-	Principal Component Analysis
PTFs	-	Pedotransfer Functions
PSD	-	Particle Size Distribution
Psi	-	Matric Suction
RF	-	Random Forest
RLCMS	-	Regional Land Cover Monitoring System
RP	-	Return Period
RR	-	Random Roughness
SAVI	-	Soil-Adjusted Vegetation Index
SCS	-	Sun-Canopy-Sensor
SG	-	SoilGrids
SHP	-	Soil Hydraulic Properties
SL1	-	Soil Layer 1
SL2	-	Soil Layer 2
SOC	-	Soil Organic Carbon
SOM	-	Soil Organic Matter
SRTM	-	Shuttle Radar Topography Mission
SWIR	-	Shortwave Infrared
tcAngle	-	Tassel Cap Angle
tcDist	-	Tassel Cap Distance
TDOM	-	Temporal Dark Outlier Mask
TIRS	-	Thermal Infrared Sensor
USGS	-	U.S. Geological Service
WB-C	-	Walkley and Black Carbon

TABLE OF CONTENTS

1.	INTRODUCTION.....	1
1.1.	Background.....	1
1.2.	Problem statement.....	4
1.3.	Objectives and research questions.....	5
2.	RESEARCH AREAS.....	6
3.	METHODOLOGY.....	8
3.1.	Data.....	8
3.2.	Soil data analysis.....	8
3.2.1.	Pedotransfer functions.....	9
3.2.2.	Independent soil assessment.....	9
3.2.3.	Comparative soil assessment.....	10
3.2.4.	Soil sampling strategy.....	10
3.2.5.	Soil sampling.....	11
3.2.6.	Laboratory analysis.....	12
3.2.7.	Slope unit delineation.....	12
3.3.	Land cover mapping.....	13
3.3.1.	Random forest classifier.....	14
3.3.2.	Land cover typology.....	15
3.3.3.	Image processing.....	16
3.3.4.	Covariate layers.....	18
3.3.5.	Reference data.....	21
3.3.6.	Machine learning.....	21
3.3.7.	Validation data.....	22
3.3.8.	Accuracy assessment.....	22
3.3.9.	Land cover change analysis.....	23
3.4.	OpenLisem flood model.....	24
3.4.1.	Data preparation.....	25
3.4.2.	Model calibration.....	29
4.	RESULTS AND DISCUSSIONS.....	30
4.1.	Soil analysis.....	30
4.1.1.	Slope units.....	30
4.1.2.	Influence of land cover on soil properties.....	31
4.1.3.	Field data related to the terrain.....	35
4.1.4.	SoilGrids variability in the landscape.....	38
4.1.5.	Similarity analysis of SoilGrids and field data.....	40
4.1.6.	Summary of soil analysis.....	42
4.2.	Land cover analysis.....	42
4.2.1.	Land cover change analysis.....	45
4.3.	Flash flood modelling.....	48
4.3.1.	Final model parametrisation.....	48
4.3.2.	Calibration and the effects of different elevation models.....	50
4.3.3.	Sensitivity of flood dynamics to soil information.....	52
4.3.4.	Effects of land cover change on flood dynamics.....	57
5.	CONCLUSION AND RECOMMENDATIONS.....	62
	ANNEXES.....	74

LIST OF FIGURES

Figure 1. Location map of Ban Da Na Kham and Laplae watershed.....	6
Figure 2. Climate diagram of the research area. Based on data from the Thai Meteorological Department for Uttaradit city station (ID: 351002) for the period 2006 to 2018.....	7
Figure 3. Slope positions in a natural landscape; adapted from Schoeneberger et al. (2012).....	11
Figure 4. Common landforms recognisable with Geomorphon (Jasiewicz & Stepinski, 2013).....	12
Figure 5. Workflow of the land cover classification; adapted from Saah et al. (2019).....	13
Figure 6. Random Forest classification; V = random variables and S = random samples.....	14
Figure 7. Land cover classes in the research areas. a) Cropland, b) Orchard, c) Teak Plantation, d) Mixed Forest, e) Urban, f) Water Bodies (Google Earth Pro, 2019).....	16
Figure 8. Locations of historical flood height measurements in Laplae city (left) and flood mark painted on a power pole (right).....	29
Figure 9. Slope position map.....	30
Figure 10. Variability of soil properties per land cover; a) Saturated Hydraulic Conductivity, b) Soil Organic Matter, c) Porosity, d) Bulk Density.....	32
Figure 11. Variability of soil properties per slope unit; a) clay, b) sand, c) silt, d) Saturated Hydraulic Conductivity, e) Porosity, f) Bulk Density, g) Organic Matter Content.....	37
Figure 12. Land cover map Ban Da Na Kham watershed, a) 2005 and b) 2018.....	44
Figure 13. Change map of Ban Da Na Kham watershed.....	46
Figure 14. Main road in 2005 a) and 2018 b) ; Source Google Earth Pro.....	46
Figure 15. Design storm rainfall graph.....	48
Figure 16. Soil depth map for scenario D1.....	49
Figure 17. Flood simulation results in Laplae watershed using different DEMs: a) DEM _o ; b) DEM _v ; c) DEM _m and d) SRTM DEM.....	51
Figure 18. Example of DEM irregularities in cropland areas; a) section of land cover map, b) flood simulation with deep water ponds in parts of the cropland, c) hillshade showing undulating surface of cropland area, d) photo taken with view in north-east direction standing at black dot c)	52
Figure 19. Maximum flood depth (m) for field data (FD) and SoilGrids (SG) for different soil depths: a) FD-D1, b) FD-D2, c) FD-D3, d) FD-D4, e) SG-D1, f) SG-D2, g) SG-D3 and h) SG-D4, D1= min: 0.36 m; max: 2.04 m, D2= min: 1.36 m; max: 3.04 m, D3= min: 2.36 m; max: 4.04 m, D4= min: 3.36 m; max: 5.04 m.....	54
Figure 20. Maximum flood duration (hr) using field data a) and SoilGrids b) with soil depth scenario D1 (min: 0.36 m; max: 2.04 m). Zoomed area shows how flood time changes when using different soil information.....	55
Figure 21. Total infiltration (mm) for field data a) and SoilGrids b) using soil depth scenario D1 = (min: 0.36 m; max: 2.04 m). Zoomed section of a) shows high infiltration values next to impermeable areas. Zoomed section of b) shows artificial border between low and high infiltration values.....	56
Figure 22. Maximum flood depth for 2005 a) and 2018 b) . Zoomed in area indicates an increase in maximum flood depth in the lowland area due to land cover changes.....	60
Figure 23. Flood duration for 2005 a) and 2018 b) . Zoomed in area indicates prolonged floods in the lowland areas due to land cover changes.....	61

LIST OF TABLES

Table 1. Data used and its source and properties.....	8
Table 2. Soil samples per slope position and land cover.....	12
Table 3. Band combinations for normalised differences; adapted from Saah et al. (2019).....	19
Table 4. Number of training data points collected in each class.	21
Table 5. Number of field validation points for each class.	22
Table 6. Validation data points for 2005 and 2018.	22
Table 7. Main input data required for openLISEM.	25
Table 8. Surface parameter for openLISEM.	27
Table 9. Daily rainfall prior, during (red square) and after the event.....	28
Table 10. Confusion matrix of slope unit classification.	31
Table 11. Statistics of soil properties per land cover.	31
Table 12. Correlation (r) of ancillary variables with K_s per land cover.....	33
Table 13. Soil properties per orchard subclass.....	34
Table 14. Soil properties per slope unit.	35
Table 15. Correlation of terrain derivatives with soil properties.....	36
Table 16. Correlation of K_s per slope unit with soil properties.	36
Table 17. Statistics of SoilGrids soil properties per land cover.	39
Table 18. Statistics of SoilGrids soil properties per slope unit.....	39
Table 19. Cosine Similarity per Position.	40
Table 20. Cosine Similarity per Land Cover.....	40
Table 21. Results of the test of normality.	41
Table 22. Output of Wilcoxon Signed-Rank test on SoilGrids derived properties.....	41
Table 23. Land cover area for 2005 and 2018.....	43
Table 24. Accuracy assessment of land cover map 2018.	45
Table 25. Accuracy assessment of the land cover map 2005.....	45
Table 26 Land cover change matrix of Ban Da Na Kham watershed (2005 to 2018).....	47
Table 27. Soil hydraulic properties SL1.....	49
Table 28. Soil hydraulic properties SL2 and texture class per slope position.....	49
Table 29. Water balance for eight model simulations using field data (FD) and SoilGrids (SG) data by using four soil depth scenarios; (SG-FD) represents the difference between SG and FD for each soil depth scenario.....	53
Table 30. Flood depth related to soil data information and total reduction from D1 to D4.....	55
Table 31. Water balance of Ban Da Na Kham watershed for 2005 and 2018.	58
Table 32. Runoff amounts per land cover for 2005 and 2018.	59
Table 33. Flood depths and their corresponding extent due to land cover change.	60

LIST OF ANNEXES

Annex 1. Geological map of Uttaradit (Ministry of Natural Resources and Environment Thailand, 2019)	75
Annex 2. Pedotransfer script by Jetten and Shrestha (2018) based on the equations of Saxton et al. (2006)	76
Annex 2.1. Soil sampling locations Ban Da Na Kham watershed.....	78
Annex 2.2. Morphological properties of the soil for each site (field measurements)	79
Annex 2.3. Physical and chemical properties of the soil for each site (field measurements)	80
Annex 2.4. Description of laboratory measurements for soil property determination	81
Annex 2.5. Physical and chemical properties of SoilGrids for each site	83
Annex 3. Description of Landsat images used for yearly and seasonal composites.....	84
Annex 3.1. Google Earth Engine code for cloud removal by Saah et al. (2019)	86
Annex 3.2. Google Earth Engine code for cloud shadow removal by Saah et al. (2019)	87
Annex 3.3. Google Earth Engine code for BRDF correction by Saah et al. (2019)	88
Annex 3.4. Google Earth Engine code for topographic correction by Saah et al. (2019)	92
Annex 3.5. Google Earth Engine code to generate covariate layer (seasonal) adapted from Saah et al. (2019)	95
Annex 3.6. Google Earth Engine code to generate covariate layer (yearly) adapted from Saah et al. (2019)	99
Annex 3.7. Selected covariates layers for the classifications of 2005 and 2018	103
Annex 4. IDF curves for Uttaradit province (Rittima et al., 2013)	104
Annex 5. Stream dimension measurements, locations and average values for validation	105
Annex 5.1. Channel maps Ban Da Na Kham a) and Laplae b) watershed	106
Annex 6. PCRaster script for soil depth map generation based on Kuriakose et al. (2009)	107
Annex 7. Soil sampling site specific Cosine Similarity results	108
Annex 8. Land cover map Laplae watershed (2005)	109
Annex 9. Gumble plot of daily long-term precipitation measurements (1951-2018). Red circle represents the flash flood event in 2006.	110
Annex 10. Simulated flood height measurements in comparison to measured flood height points	111
Annex 11. Physical property layers of SoilGrids for soil layer 1 (5 cm)	112

1. INTRODUCTION

1.1. Background

Hydrological hazards such as floods and droughts cause the loss of lives and economic damage around the world. According to the World Disaster Report 2016, in the period between 2006 – 2015 flood hazards were reported the deadliest, which also caused the highest economic loss as compared to the effects of other natural hazards such as earthquakes and storms. With almost 700 events, Asia is the continent which was affected most by flooding (IFRC, 2016). The 2011 flood was, for instance, one of the most disastrous events in the recent history of Thailand, with a death toll of 884 and millions left homeless or displaced (Aon Benfield, 2012). Another example is the year 2015 when Myanmar was devastated with flood problems. Exceptional strong monsoon rains triggered landslides, flash floods, and river floods, with 69 casualties affecting approximately 250 thousand people, countrywide. Moreover, more than 520 thousand acres of agriculture land were destroyed (USAID, 2015).

Model-based projections and international long-term trend studies of hydrological processes prognoses an increase of frequency and intensity of rainfall events in many parts of the world in the future due to the effects of climate change (IPCC, 2014). Consequently, urgent attention should be drawn to flash floods to prevent future disastrous events by means of capacity building and awareness raising activities. Understanding the underlying processes and, favourable conditions for formation as well as potentially triggering and influencing factors can provide implications for flood risk management and therefore prevent the loss of human lives and economic assets.

Flash floods can be characterised by their temporal and spatial scale. Bout and Jetten (2017) associate them with local rainfall events with high intensities and short durations, occurring mostly in mountainous upstream watersheds. As such, they are related to short watershed response times with rapid increase and release of discharge. Resulting floods may last several hours, but durations are rarely exceeding one day (Bout & Jetten, 2017; Marchi et al., 2010). Therefore, it can be argued, that flash flood generation and behaviour are among others related to the shape and size of the watershed (e.g., circular or elongated) and to the pattern of the rainfall event (intensity and duration).

Runoff represents the main transport process for flood water during flash floods. Factors influencing runoff generation are manifold. Rainfall represents the most fundamental factor. However, also hydrological pre-conditions (e.g., initial soil moisture), soil physical properties, terrain (e.g., slope gradient), and land use and land cover are decisive for runoff occurrence (Marchi et al., 2010). Physical properties comprise particle size distribution (PSD), porosity, water retention properties and hydraulic conductivity of the soil. PSD provides information about the grain size distribution of soil, hence about the percentage of sand, silt, and clay, which affects soil hydraulic properties (SHPs) that are closely linked to runoff generation.

Decisive SHPs for runoff generation are infiltration, porosity and saturated hydraulic conductivity (K_s) (Marchi et al., 2010). As described by Schaetzl and Anderson (2005), infiltration is the process of water entering into the soil. Its rate depends, for example, on the pore size and initial soil moisture. Thereby, porosity describes the pore space which can be filled by water, with increasing moisture content, the infiltration rate decreases. K_s characterises the ease with which water or other fluids can move through the soil. Fine textures (e.g., silt and clay) are smooth, having finer pores and a substantially greater volume of

open space compared to coarse textures like sandy soils. Therefore, their water holding and retention capacity will be higher. At the same time, sandy soils have a higher K_s . Thus, the water can drain faster, and less surface runoff will occur (Schaeztl et al., 2005).

Steady socio-economic developments can influence the occurrence of extreme flash flood events (Marchi et al., 2010). Man-made land use and land cover changes like deforestation and urbanisation can modify the hydrological processes of watersheds dramatically, and hence influence surface runoff characteristics and flood dynamics (Sajikumar & Remya, 2015). In general, alterations can emerge for instance from compaction (e.g., tillage), surface sealing (e.g., urban structures) and vegetation cover changes (e.g., deforestation and cultivation) (Bronstert et al., 2002). Compaction caused by, for example, heavy agricultural machinery describes the densification of soil particles and the loss of pore spaces. It will lead, for instance, to a decrease in K_s and a reduced infiltration capacity. Surfaces sealed by physical structures such as roads or buildings will make the ground impermeable for water and favour runoff generation. Hence, runoff occurs either when (I) the surface is impermeable, (II) the precipitation intensity exceeds the infiltration rate, or (III) the soil is fully saturated.

A conventional way to foster understanding of complex surface and subsurface processes is hydrological modelling (Raudkivi, 1979). In the literature, three main types of hydrological models are identified, (I) empirical models, (II) conceptual models, and (III) physical based models. Empirical models, on the one hand, are observation-based, data-driven models, conceptual models, on the other hand, consider all components of the hydrological process and work with semi-empirical equations. Whereas physical models are based on mathematical equations to represent real-world processes as realistic and simple as possible (Devia & Ganasri, 2015; Merritt et al., 2003). They can provide a variety of information and can be applied to a wide range of applications such as flood process modelling, vegetation growth modelling, and groundwater modelling.

According to Bout and Jetten (2017), physically based models can be further classified as decoupled models and integrated watershed models. The former separates upstream runoff generation from flooding in the downstream areas. While an upstream model is used to generate discharge values, a downstream model is then responsible for the flood simulation. The latter simulates the hydrology at the watershed scale and generates runoff based on rainfall and infiltration. Bout and Jetten (2017) further pointed out that decoupled models are often not applicable for flash flood modelling as flash floods are not necessarily linked to an overflow of a channel since they are also generated in adjacent sloping terrain.

Widely used integrated catchment models are for instance MIKE-11 (DHI, 2017), the Hydrologic Modelling System HEC-HMS (Scharffenberg, 2016), the Soil and Water Assessment Tool (SWAT) (Shekhar & Xiong, 2008) and the Limburg Soil Erosion Model (openLISEM) (Bout & Jetten, 2018). A distinction can be made considering input data requirements, the processes modelled, or the temporal scale they operate in. Where MIKE-11, SWAT, and HEC-HMS, for example, can work on a temporal scale of years, they incorporate evapotranspiration and groundwater flow. OpenLISEM, on the other hand, is a purely event-based model working with single rainfall events and therefore neglects such processes. Operating in time steps of minutes, openLISEM is tailored to model flash flood processes.

The drawback of physically based models is their immense data demand and the output dependency on the input data quality (Sanchez-Moreno et al., 2014). OpenLISEM requires a minimum of 24 maps, which can be derived from four main sources encompassing rainfall, topography, surface and soil related information

(Bout et al., 2018). Among others, one characteristic of physically based hydrological models is, that soil data is an essential baseline data, and the model output is highly dependent on its quality (Merritt et al., 2003).

Usable soil data for any kind of application are rare and substantial investments have to be taken in new detailed soil measurements (Sanchez et al., 2009). Therefore, it can be assumed that especially developing countries, which are usually most prone to natural hazards and the effects of climate change (IFRC, 2016), exhibit a deficit of appropriate soil data. Existing soil data are often inaccurate, outdated, with coarse spatial resolution, disregard soil properties, as well as site-specific geomorphological processes due to inadequate methods (Arrouays et al., 2014; Grunwald et al., 2011; McBratney et al., 2003). In the light of that, the data are not valid to support efforts of sustainable development, which is gaining importance in the countenance of growing pressure on the planet, caused by climate change, biodiversity loss, land degradation and urbanization (FAO & ITPS, 2015; Montanarella & Vargas, 2012).

Recently, the International Soil Reference and Information Centre (ISRIC) made 'SoilGrids' (SG), a global soil dataset publicly available to bridge the gap between soil data demand and availability (Hengl et al., 2015). SG is based on machine learning and comes in a 250 m grid resolution. According to Hengl et al. (2017), SG spatial predictions are based on approximately 150,000 individual soil profiles spread over the world, which had to be merged. Standardisation methods were applied to translate soil data provided in national classification systems (up to 20 %) to the international classification systems (World Reference Base and USDA). For areas with no existing observation points (e.g., mountain tops, steep slopes and inaccessible tropical rainforest), expert-based pseudo-observation were input. Remote sensing soil-covariates such as MODIS land products (e.g., land cover, surface temperature and Enhanced Vegetation Index) and Shuttle Radar Topography Mission (SRTM) digital elevation model (DEM) derivatives (e.g., slope, profile curvature, and valley depth) provide additional support for the predictions (Hengl et al., 2017). Since it is estimated that detailed soil profiles are available for only one-third of the world (Bonfante & Bouma, 2015), the significance of global digital datasets like SG is increasing to overcome existing data gaps.

So far, there has been no initiative in the published literature to verify SG performance and accuracy in disaster risk research. Nevertheless, SG was deployed in several studies. Bout and Jetten (2017) used it to validate flow approximations in hydrological modelling, Shrestha (2014) for flash flood modelling in the Fella basin in Italy and Chen et al. (2016) in a multi-hazard risk assessment. Further applications can be found in disciplines like crop modelling (Han et al., 2015) and studies concerning carbon stock estimations (Tifafi et al., 2018). However, none of these studies assessed how SG may have influenced the study outputs with its prediction-based data points, and coarse spatial resolution and thereby potentially not reflected spatial variability. None of them discussed the limitations of SG nor conducted an accuracy assessment.

1.2. Problem statement

Soil data is essential baseline information for hazard modelling and therefore, also for flash flood modelling. In many parts of the world, affected by flash flood hazards, detailed hazard assessments are needed to support sustainable development. Though, detailed soil data are still lacking, especially in remote areas. The global digital soil dataset SG, developed by ISRIC, represents a possibility to bridge the gap between soil data demand and availability, but its performance has to be tested. Besides this, it is evident that flash flood frequencies and intensities are increasing, due to climatic alterations and due to socio-economic developments and the accompanied land use and land cover changes. Therefore, in this thesis, an impact assessment of soil information on flash flood modelling was performed, to compare the performance of different soil information sources and to identify potential limitations of SG data. Furthermore, the effects of long-term land cover changes were investigated. For this purpose, the integrated flood model openLISEM was deployed, (I) with SG data, (II) with detailed field data (FD), and (III) with changed land cover information. The respective model outputs provide information regarding SG and the effects of land cover changes. A better understanding of SG limitations is expected to increase both its attractiveness for the research community, governmental bodies and national and international organizations working in the area of disaster risk reduction, and cautiousness when used. In addition, with the knowledge about the influence of land cover changes on flash floods, risk-informed decision making, and sustainable spatial planning can be promoted. In conclusion, this thesis is aligned with the following global goals and agendas, to make a meaningful contribution:

Transforming our World - the 2030 Agenda for Sustainable Development - Particularly the following Sustainable Development Goals (SDGs):

- **SDG 1:** End poverty in all its forms everywhere; Target 1.5: By 2030, build the resilience of the poor and those in vulnerable situations and reduce their exposure and vulnerability to climate-related extreme events and other economic, social and environmental shocks and disasters.
- **SDG 13:** Take urgent action to combat climate change and its impacts; Target 13.1: Strengthen resilience and adaptive capacity to climate-related hazards and natural disasters in all countries.

The Sendai Framework for Disaster Risk Reduction (SFDRR):

- **Priority 1:** Understanding disaster risk
- **Priority 3:** Investing in disaster risk reduction for resilience
- **Priority 4:** “Build Back Better” in recovery, rehabilitation, and reconstruction

1.3. Objectives and research questions

The main objective of this research is to conduct an impact assessment of soil information and land cover change on flash flood modelling on a watershed scale in Thailand. To achieve the main objective, the following specific objectives with corresponding research questions were set:

Objective 1: Comparative analysis of SoilGrids (SG) versus field data (FD) for flash flood modelling;

1. How well do SG and FD correlate?
2. How do the soil properties of (I) FD, and (II) SG relate to the main land cover types?
3. How do the soil properties of (I) FD, and (II) SG relate to the terrain?
4. Is model calibration based on historical flood marks from a nearby watershed possible?
5. What are the quantitative differences of the model output using, (I) FD, and (II) SG in relation to flood dynamics?
6. What is the sensitivity of the flood dynamics to different soil depths using, (I) FD, and (II) SG?

Objective 2: Analysis of the effects of land cover change on flash flood behaviour;

1. Which land cover changes occurred in the study area between 2005 and 2018?
2. What are the possible reasons behind these land cover changes?
3. Which land covers generate the highest average runoff?
4. How do these land cover changes affect runoff generation and flood dynamics?

2. RESEARCH AREAS

The research was carried out in two areas that are located in Uttaradit province in Northern Thailand within the Latitudes 17°37'N – 17°52'N, and Longitudes 99°55'E – 100°09'E (Figure 1). The first is Ban Da Na Kham watershed (86.9 km²), and the second Laplae watershed (156.9 km²). Two watersheds were selected as in Ban Da Na Kham watershed diverse types of land cover changes were observable, whereas Laplae offers the possibility of model calibration as flood marks from a flash flood event in 2006 are available. Both sides are located in the vicinity to each other where the former is located in Mueang district and the latter in Laplae district (Figure 1). In general terms, Ban Da Na Kham watershed is tube-shaped, whereas Laplae is more elongated. However, even having different shapes, Laplae is expected to be suitable for calibration purposes since, geology, land cover and topography are nearly identical in both watersheds.

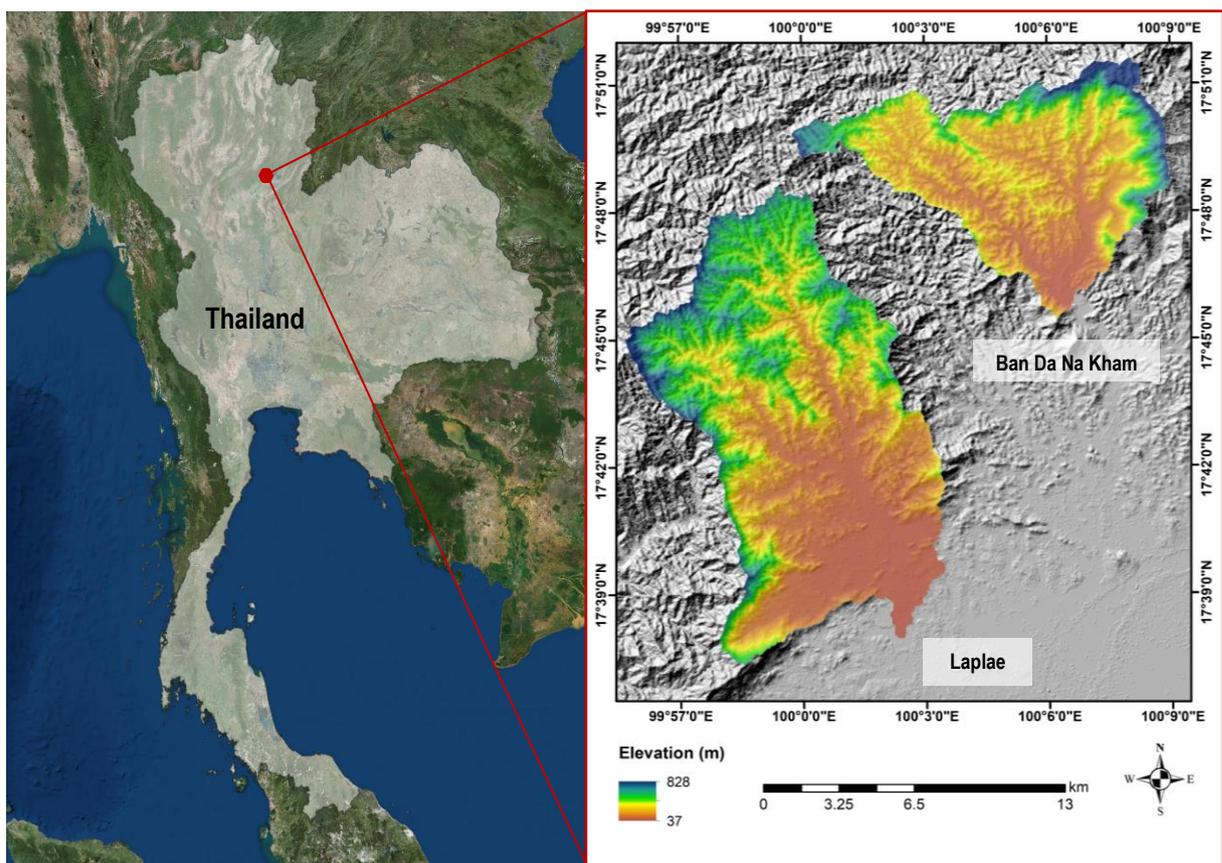


Figure 1. Location map of Ban Da Na Kham and Laplae watershed.

Topography and Land Cover

Most areas in the watersheds are characterised by mountainous terrain crisscrossed by small valleys. Highest areas are in the northern parts of both watersheds with elevations up to 828 m above mean sea level in Laplae and up to 754 m in Ban Da Na Kham. The lowest areas are in the southern parts of the watersheds with a minimum of 37 m in Laplae and 66 m in Ban Da Na Kham.

A variety of land cover types ranging from urban areas over orchards to natural forest are present in both watersheds. Sloping terrain and narrow valleys are cultivated with fruit trees e.g. long kong, banana and durian trees, whereby larger valleys are used for paddy rice cultivation. Mountain summits and steep slopes

are covered by impassably natural forest. Furthermore, teak plantations for commercial timber production can be found in different places within Ban Da Na Kham watershed. In the southern lowland area of Laplae watershed, the city Si Phanommat with a population of 3200 is situated. On the contrary, Ban Da Na Kham is sparsely populated, having small isolated villages lined along the main roads of the watershed.

Geology

From a geological point of view, both the watersheds have three distinct geological formations. In the higher altitudes in the northern parts, they consist of the Khao Ploung formation comprising shale or mudstone interbedded with greywacke sandstone. In the south, the Lab Lae formation takes over occupying the majority of both watersheds with sandstone (greywacke) interbedded with shale. Alluvial deposits containing gravel, sand, silt and clay can be found in the southern lowland areas (Annex 1).

Climate

Uttaradit province has a humid tropical climate influenced by the north-eastern and south-western Monsoon which determines the three seasons, namely cold (November – February), hot (February – May) and rainy (May – October). Annual precipitation ranges from approximately 830 to 2.100 mm, with August being usually the wettest month of the year. The warmest month is April with an average temperature of 30 °C and the coldest average temperatures with 24 °C is in January (Figure 2).

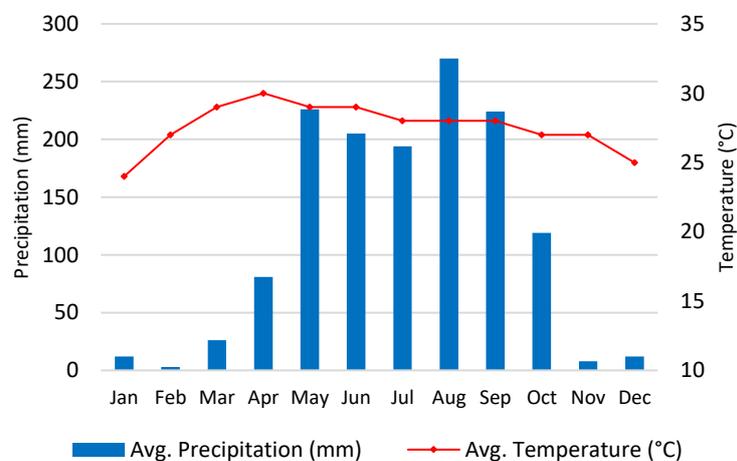


Figure 2. Climate diagram of the research area. Based on data from the Thai Meteorological Department for Uttaradit city station (ID: 351002) for the period 2006 to 2018.

Historic flash flood event

On 22nd of May 2006 flash floods, landslides, and debris flows were triggered by prolonged heavy rainfall in several provinces in lower Northern Thailand. Among others, Uttaradit province was affected by this unusual event. Flood heights of several meters were reported among different communities in the province. In the wake of the disaster, 87 people died, 700 houses were totally damaged, and almost 4000 partially. Additionally, more than 370 hectares of mountainous mixed-fruit tree orchards were completely destroyed due to landslides, and some hundred hectares of lowland orchard and cropland were flooded and buried by mud (Boonyanuphap, 2013). These consequences are due to three days of consecutive rainfall with a total of approximately 400 mm, having its peak on May 22, 2006, with a total rainfall amount of 263.7 mm recorded at the meteorological station in Uttaradit city.

In general, the northern part of Thailand is frequently affected by flash floods and landslides (CFE-DM, 2018). However, the event in 2006 marks a historical event in Thailand's history because it took the lives of many people and caused tremendous economic damage. Thus, this event was selected to serve as the basis for the modelling efforts in this research.

3. METHODOLOGY

This chapter gives first an overview of the data used in this research project, and second describes the methodologies applied. The methodology used to satisfy the research objectives is divided into three parts. The first part consists of the methodological steps undertaken to do the soil data analysis and to further conduct a comparative assessment between FD and SG. The second part has a strong remote sensing component, focusing on Google Earth Engine (GEE) and machine learning algorithms in order to conduct land cover mapping and land cover change analysis. In the third part, the soil data and land cover information are combined and used as an input for the hydrological modelling.

3.1. Data

The data used for this research study is summarized with its sources and properties in Table 1.

Table 1. Data used and its source and properties.

Type	Method	Spatial Resolution	Source
Alos Palsar Digital DEM	Satellite (Radar)	12.5 m	Alaska Satellite Facility, 2006-2011 (www.asf.alaska.edu)
SRTM DEM	Shuttle Radar	30 m	U.S. Geological Survey (www.usgs.gov)
Digital Soil Data	Random Forest classification	250 m	Hengl et al. (2017) (soilgrids.org)
Soil Data	Undisturbed and disturbed sampling	Point	Fieldwork
Rainfall	Rain gauge (daily measurements)	Point	Thai Meteorological Department (1952-2018)
Land Cover Maps	Landsat 5 and 8 Random Forest classification	30 m	-
Road Network	-	Vector	OpenStreetMap (www.openstreetmap.org)
Historical flood height measurements	Field measurements	Point	Fieldwork

3.2. Soil data analysis

In this sub-section, the methodology to answer the following research questions is outlined:

- How well do SG and FD correlate?
- How do the soil properties of (I) FD, and (II) SG relate to the main land cover types?
- How do the soil properties of (I) FD, and (II) SG relate to the terrain?

The first part comprises statistical methods for the assessment of (I) the FD, and (II) the global digital soil dataset SG (250 m) in relation to land cover types and the terrain. The second part is dedicated to the

methodology of the comparative analysis between the two datasets. In the third part, the soil sampling strategy, the sampling process and the methods used for the laboratory analysis of the soil data are described. Lastly, the generation of a slope unit map using the tool geomorphon (Stepinski & Jasiewicz, 2011) is explained.

3.2.1. Pedotransfer functions

Pedotransfer Functions (PTFs) are regression equations established mostly based on large soil datasets and are used to predict SHPs (Looy et al., 2017). For predictions, soil texture (sand, silt and clay contents) and bulk density (D_b) are commonly required (Williams et al., 1992). Some PTFs make also use of soil organic matter (SOM) for example, see Wösten et al. (1999), Nemes et al. (2005) and Saxton & Rawls (2006). PTFs generally reflect the interactions of soil properties from the soils they were constructed of, and hence, their prediction capabilities are coupled to the underlying soil database (Nemes et al., 2005). PTFs for the humid tropics and especially for Asia are limited in number according to a review article by Botula et al. (2014). Often, they are built based on small or even unknown sample sizes, which potentially introduce a large uncertainty. For this study, the PTFs of Saxton & Rawls (2006) were used to estimate SHPs of SG data. The PTFs are based on a combination of nonlinear multiple regression equations with texture and SOM, combined with the hydraulic conductivity and water retention equations of Brooks and Corey (1964). Those PTFs were originally created from approximately three thousand USA soil samples and represent, according to Gijssman et al. (2003), the most accurate PTFs compared to other common methods. The conversion from SG physical and chemical properties to SHPs was done using the PCRaster script from Jetten and Shrestha (2018) (Annex 2).

3.2.2. Independent soil assessment

For the two soil datasets used in this study (FD and SG), an independent assessment was conducted. This assessment included the calculation of descriptive statistics such as the mean and standard deviation for statistical data summarization and measure of data dispersion. Furthermore, Box and Whisker charts were plotted for visual investigation of the data variability within the landscape.

To foster an understanding of the soil-landscape relationship, the soil data were grouped based on the land cover class and based on their slope position (explained below). Moreover, selected pairwise correlations of soil properties were conducted using Pearson's coefficient r , which is a measure of linear interdependency of two variables (Eq. 1). This was done to investigate statistically dependencies. The coefficient ranges from -1 to +1, with -1 being an indication of a negative dependency, 0 describes independency, and +1 a positive dependency. The computations were done with the software package IBM SPSS statistics 25. Equation 1 describes Pearson's r .

$$r = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum(X - \bar{X})^2} \sqrt{\sum(Y - \bar{Y})^2}} \quad (1)$$

Where X and Y are the two variables of interest and \bar{X} and \bar{Y} are the respective mean values of the variables.

3.2.3. Comparative soil assessment

Comparative soil assessment was conducted by using Cosine Similarity and Wilcoxon Signed-Ranked test. The Cosine Similarity is usually applied as a similarity measure for text documents by creating frequency vectors of keywords (Han et al., 2013). In the frame of this research, the frequency vectors are substituted by soil property vectors. Computed is the cosine of the angle between two vectors. According to Han et al. (2013) a cosine of 0 does mean that the two vectors have a 90 degrees angle, and therefore are orthogonal to each other, meaning no match between the two vectors. The closer the cosine value is to 1, the more similar the vectors are. The advantage of cosine similarity is that the vectors can accommodate as many variables as needed. Therefore, different combinations of soil properties can be tested. Interpretation of the results is comparable to Pearson's r . Equation 2 defines the cosine similarity.

$$\cos(X, Y) = \frac{X \cdot Y}{\|X\| \|Y\|} = \frac{\sum_{i=1}^n X_i Y_i}{\sqrt{\sum_{i=1}^n X_i^2} \sqrt{\sum_{i=1}^n Y_i^2}} \quad (2)$$

Where X and Y are two vectors of comparison, $\|X\|$ and $\|Y\|$ are the Euclidean norms, or the length of the vectors. As soon as the soil property vectors are composed of soil properties with different units, e.g. D_b (g cm^{-3}) and K_s (mm h^{-1}), a comparison becomes difficult. Therefore, the normalisation of the measurements is a necessary data preparation step. Data normalisation was done with min-max normalisation using equation 3 (Han et al., 2013).

$$\frac{X - X_{min}}{X_{max} - X_{min}} \quad (3)$$

Where X is the soil property of interest; X_{min} and X_{max} represent the minimum and maximum of the measured soil property, respectively. Normalisation results are given in a sequence of numbers ranging from 1 representing the maximum measured value to 0 representing the minimum measured value.

As a second similarity measure, the Wilcoxon Signed-Rank test was conducted. This statistical measure was chosen above the common Paired-Samples t-test because (I) not all of the tested variables (soil properties) followed the assumption of normal distribution, and (II) because the Wilcoxon Signed-Rank test is robust to the effects of possible outliers. Calculated differences were considered to be significant if the z-value was less than the critical P-Value of 0.05 (Field, 2009).

The normality of the data was investigated with the Shapiro-Wilk test, kurtosis and skewness. If the skewness and kurtosis z-value ($P > 0.05$) is within the span of -1.96 to +1.96 and a Shapiro-Wilk P-Value above 0.05, a normal distribution can be expected. If one of the conditions is violated the data can be expected to be non-normal distributed (Field, 2009).

3.2.4. Soil sampling strategy

According to Möller et al. (2008), landscapes are determined by their landforms such as flood plains, alluvial fans, ridges and slopes. The assumption that similar landforms in an area exhibit similar soils is widely spread in the soil surveying community (McKenzie et al., 2008; Möller et al., 2008; Schaetzl & Anderson, 2005).

The rationale behind this is the assumption that the origin of landforms is controlled among others by geology and the prevailing surface and subsurface processes and that the underlying geology forms mostly the parent material of the occurred soils. Hence, landforms are expected to give an inference on soil genesis, soil formation and the forming processes (Möller et al., 2008). If complying with this line of thinking, it would be obvious to either base the soil sampling strategy on an existing geological or a soil map, or to conduct a terrain analysis dividing the area into its terrain units. As no such maps were available prior to the fieldwork and explicit terrain units could not be identified, another approach had to be chosen.

Therefore, the hillslope concept which divides the landscape into five slope positions (Schoeneberger et al., 2012) (Figure 3), was considered to represent an acceptable alternative to the other approaches for a soil sampling strategy. Slope positions are known to serve as a good predictor of soil properties and represent a comparable simple way of dividing the landscape into different units (Miller & Schaetzl, 2015; Schaetzl et al., 2005).

To aid the selection of suitable soil sampling sites, the elevation profile tool of Google Earth Pro was used. By creating cross sections, an identification of the approximate slope positions and placing of potential sampling points was possible. In the process, attention to accessibility, land cover type and spatial distribution was given. The proximity to roads and paths was expected to ensure accessibility to the soil sampling sites. Since the sampling sites were chosen based on the hillslope position and the land cover types, the sampling scheme was purposive.

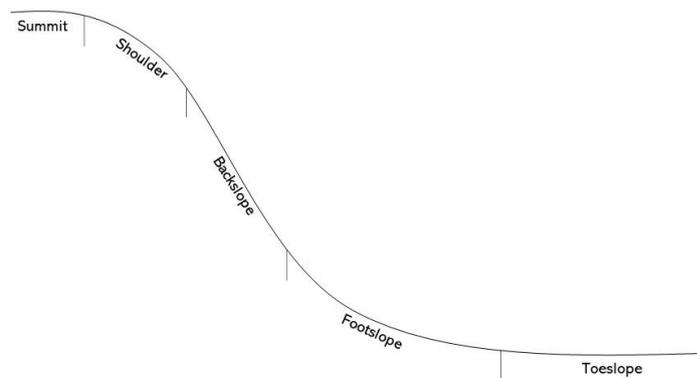


Figure 3. Slope positions in a natural landscape; adapted from Schoeneberger et al. (2012).

3.2.5. Soil sampling

Maintaining the pre-selected sampling points proved difficult after the first investigation of the watershed. Accessibility issues due to steep slopes, private property or difficult paths required an on-site selection of sampling locations (Annex 2.1). However, undisturbed and disturbed soil samples were taken on 48 locations within the land cover mixed forest, teak plantation, orchard (banana and long kong) and cropland (corn and bean). At each sampling site, GPS coordinates were recorded with a Garmin GPS receiver. The undisturbed core samples were collected at a depth of 5 cm with a steel core sampler having a diameter of 5 cm. In addition, soil depth was measured at each location with an auger (up to 1 m). Since the terrain of the watershed is characterised by steep slopes, narrow valleys and narrow hilltops, the slope positions to sample were restricted to summit, backslope and valley. A summary of the sampling points can be found in (Table 2).

Table 2. Soil samples per slope position and land cover.

		Slope Position		
		Summit	Backslope	Valley
Land Cover	Orchard	3	19	12
	Cropland	1	-	3
	Teak Plantation	-	1	1
	Mixed Forest	2	4	2

3.2.6. Laboratory analysis

The collected soil samples were analysed in the laboratory facilities of the Faculty of Engineering of the Naresuan University in Phitsanulok, Thailand. K_s , D_b , and porosity were determined based on the undisturbed surface core samples. For measuring PSD and SOM content, the disturbed samples were used. Details on the results of the laboratory analyses for each measured soil physical and chemical property along with their location information are given in Annex 2.2 and Annex 2.3.

K_s was measured using the Constant Head method as described in the operating instructions for a laboratory-permeameter (Eijkelkamp, 2013). Subsequent to the K_s measurement, D_b and porosity were measured following the method introduced by Soil Survey Staff (2014). Using the Hydrometer Method as published by the Soil Science Society of America and outlined in the Soil Survey and Laboratory Manual (Soil Survey Staff, 2014), the PSD was assessed. For the determination of the SOM content of the disturbed surface samples, the Loss-on-Ignition (LOI) method as set out by Schulte and Hopkins (1996) was applied. LOI was chosen as it represents an even more effective and simpler way to determine SOM compared to other conventional methods such as the Walkley and Black Carbon (WB-C) method, which requires additional chemicals and laboratory facilities (Paramanathan et al., 2018). Detailed descriptions of the methods and equations used can be found in Annex 2.4.

3.2.7. Slope unit delineation

For the analysis of the soil-landscape relationship, a slope unit map was generated using the GRASSGIS extension Geomorphon. According to Jasiewicz and Stepinski (2013), Geomorphon is a DEM based pattern recognition approach meant to be used for classification and mapping of landforms. Characterisations are done by the use of a local ternary operator that assigns an 8-tuple pattern making use of the symbols “-“, “0” and “+”, which describe a neighbouring cell as lower, equal or higher than the focus pixel. As the method is based on the line-of-sight principle, the neighbours are not necessarily the immediate neighbours, rather depended on the search radius L (user-defined) which varies based on the scale of interest. Where a

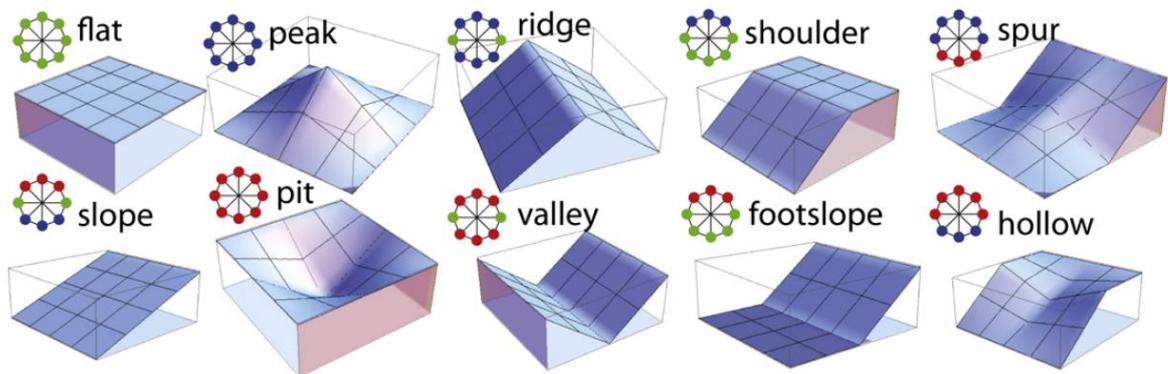


Figure 4. Common landforms recognisable with Geomorphon (Jasiewicz & Stepinski, 2013).

smaller value for L detects landforms locally and, a larger search radius L will detect landforms on several sizes. Further parameters which can be defined by the user are the flatness threshold and skip radius (Jasiewicz & Stepinski, 2013). Using a greater value for the flatness threshold and skip radius will result in an extent of the delineated plains, by reducing the influence of small irregularities (Jasiewicz et al., 2013; Veselsky et al., 2015). For simplification, Geomorphon is limited to recognize the 10 most commonly landforms, namely: flat, peak, ridge, shoulder, spur, slope, hollow, footslope, valley and pit (Figure 4).

In this study, different input values for L and the skip radius were tested and compared to the slope positions observed in the field, in order to determine the best combinations suitable for the research area and DEM resolution used. As the terrain specifications restricted the observable landforms to summit, backslope and valley (as discussed above), the output of Geomorphon was aggregated to those three units. Hence, summit and ridge were combined to form the summit unit. Shoulder, spur, slope and hollow as well as footslope are categorized as backslope unit, and lastly, the valley unit represents a combination of valley, depression and flat Geomorphons.

3.3. Land cover mapping

In this section, the methodology is described which shall satisfy the following research questions:

- Which land cover changes occurred in the study area between 2005 and 2018?
- What are the possible reasons behind these land cover changes?

Land cover maps for the years 2005 and 2018 were produced using the computational planetary-scale geospatial analysis platform GEE. The year 2005 was chosen as reference year as it represents the research area with its land cover before the great flash flood event in May 2006. As the comparative year, 2018 was selected since it represents the state of the area as observed during the fieldwork. GEE as a tool was used for the classification as it simplifies and accelerates access and analysis of remote sensing data for land cover classifications. GEE is a cloud-based online-platform accessible from any place in the world; the only requirements are a functional computer and a stable internet connection. Using the computational resources of Google’s Data Center Infrastructure and having petabytes of remote sensing datasets such as Landsat, Sentinel, Modis and non-satellite imagery, it enables parallel and fast computations on large datasets. Details on GEE can be found in Gorelick et al. (2017). Presented procedures in this chapter are based on the

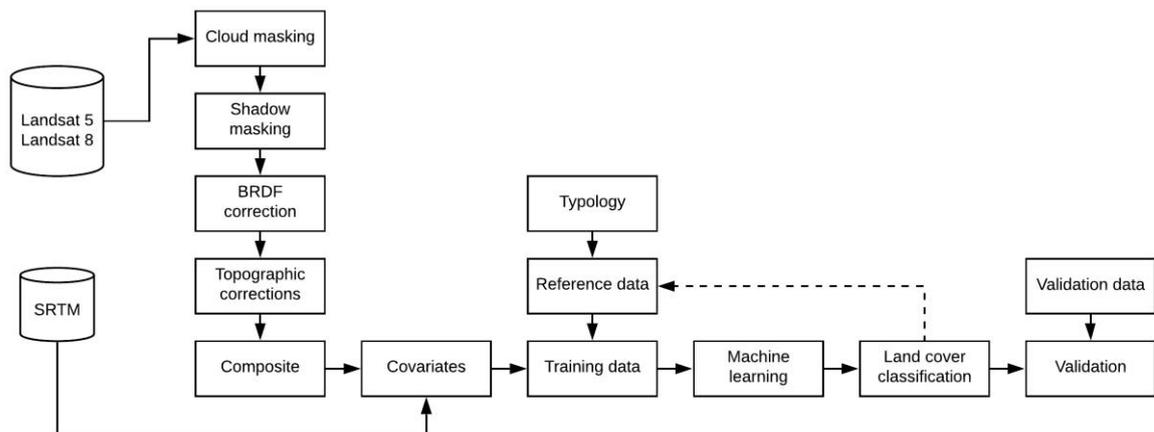


Figure 5. Workflow of the land cover classification; adapted from Saah et al. (2019).

Regional Land Cover Monitoring System (RLCMS) developed by SERVIR-Mekong as extensively described by Saah et al. (2019). Several adaptations and simplifications were done to tailor the methodology to the scope of this research. Figure 5 shows the overall workflow of the land cover classification.

3.3.1. Random forest classifier

Random forest (RF) was used as supervised classification algorithm because it allows higher mapping accuracies in comparison to conventional classifiers (e.g., maximum likelihood or simple decision tree) (Rodriguez-Galiano et al., 2012). Conventional image classification techniques often only exploit the spectral signatures and sometimes texture or pattern for pixel discrimination (Domaç & Süzen, 2006). Using RF, an enhanced differentiation between different land cover classes even in complex terrains is possible due to the incorporation of ancillary variables (Tsai et al., 2018; Rodriguez-Galiano et al., 2012). Variables can be for example spectral band indices (e.g., Normalised Difference Vegetation Index (NDVI) or Normalised Difference Water Index (NDWI)) and terrain derivatives like slope, aspect and elevation. When adding such variables, additional contextual information can be included in the classification process to improve the discrimination between different land cover classes (Domaç et al., 2006). RF found its successful application in various land cover studies for example, in Colditz (2015), Nguyen et al. (2018) and Steinhausen et al. (2018).

According to Breiman (2001), the algorithm creates a ‘forest’ by growing a number of n (user-defined) decision trees based on a random selection of a subset of data samples and a random selection of a subset of n variables. Being an ensemble classifier, RF creates the random subsets of the training data by using bootstrap aggregation (bagging) procedure (Figure 6). The output sample subsets are called bootstrap samples. Bootstrap samples are created on a random basis with the option of replacement (bagging). Thus each selected sample can be used more than once, once, or not at all within one subset (Breiman, 2001).

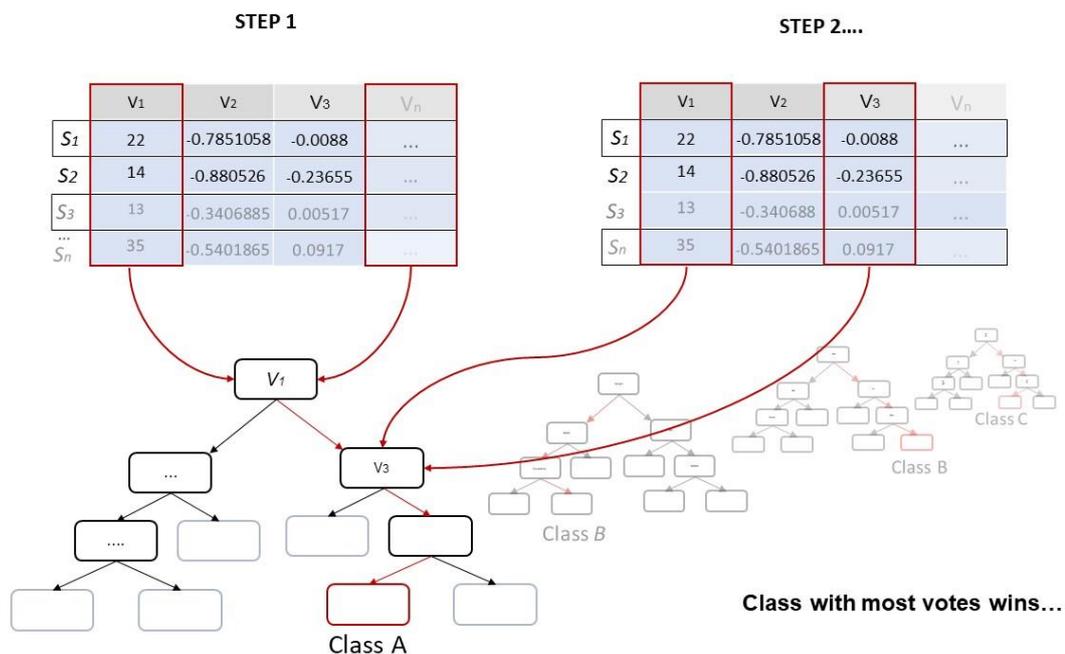


Figure 6. Random Forest classification; V = random variables and S = random samples.

Subsequently, different classifiers (trees) are built on the subsets of data, which are all a bit different even though they were created from the same original dataset. Then the final classifier is built by averaging the predictions of all sub-classifiers, in doing so it reduces the variance by taking out complexity. Hence, the stability and accuracy of the algorithm improves in the prediction making process (Bauer & Kohavi, 1999).

In other words, in RF is a pixel classified based on a set of n variables. According to Breiman (2001), the forest trees are independently grown by nodes which represent tests of these variables. The algorithm computes the information gain contributed by all the variables and splits the node accordingly with the best split contributed by the variable with the highest information gain. Each branch of a tree represents accordingly the decision of a test. Finally, each grown tree votes for a class and the class with the most votes (majority vote principle) is chosen for the respective pixel (Breiman, 2001) (Figure 6). Overall, the effectiveness of a random forest compared to ordinary decision trees lies in its variety due to bootstrapped sampling and the use of random subsets of variables at each step.

3.3.2. Land cover typology

Defining a land cover typology is a crucial step for any land cover mapping activities and is driven by its purpose and the user needs. As such, a typology is created to describe the observable (bio) physical cover of an area in a comprehensive and concise manner. By following objective criteria for the definition of clear and precise classes, various land covers can be discriminated without redundancy and overlap. This supports the observation of land characteristics and the monitoring of their dynamics. Furthermore, it enables the comparison of different land cover products and the analysis of land cover changes (Faber-Langendoen et al., 2009; Vegetation Subcommittee, 1997).

The land cover typology applied in this research study is adapted from the RLCMS. The RLCMS, in turn, is based on the Land Cover Classification System (LCCS) and associated tool developed by the Food and Agriculture Organization of the United Nations (FAO) (Di Gregorio, 2005, 2016; Di Gregorio & Jansen, 2000). The LCCS was launched with the fundamental idea to contribute to the establishment of an international land cover classification standard. Its approach is physiognomic-structural, meaning the elements are described based on their overall appearance and spatial distribution pattern (Di Gregorio, 2016). The definitions of the different land cover classes are specified below:

- **Cropland** represents a combination of irrigated or flooded rice paddy fields and dryland crops such as corn and vegetables. The majority of the cropland is covered by rice paddies which are intensively tilled (more than one cycle per year), and even being flooded, rice makes up the majority of observable surface cover. Ideally, irrigated crops and dryland crops should be classified separately as they exhibit different characteristics. However, in this research study, a separation was not possible as the area covered by dryland crops was insufficient to represent an individual class. Woody perennial crops such as fruit trees are excluded from this class (Figure 7).
- **Orchards** are defined by long term occupation of land by perennial crop trees (Blanchez, 1997). In the study area, they are mainly represented by long kong and banana plants. The fruit trees are planted in a line pattern with gaps of several meters in between. Ground cover is dense but shallow grass (Figure 7).

- **Mixed forests** include evergreen broadleaf and deciduous trees (D'Souza, 2000). The understory vegetation is very dense (Figure 7).
- **Teak plantations** consist of deciduous teak trees (*Tectona grandis*). Once a year the trees carry white flowers, arranged in dense clusters. Furthermore, the plantations are arranged in a line pattern, recognisable especially in young plantations. In mature plantations the understory vegetation is pronounced (Figure 7).
- **Urban and built-up** areas were defined as areas covered by human-made structures (e.g., buildings and roads) (D'Souza, 2000) (Figure 7).
- **Water bodies** were defined as surface water open to the sky and containing fresh water (Pekel et al., 2016) (Figure 7).

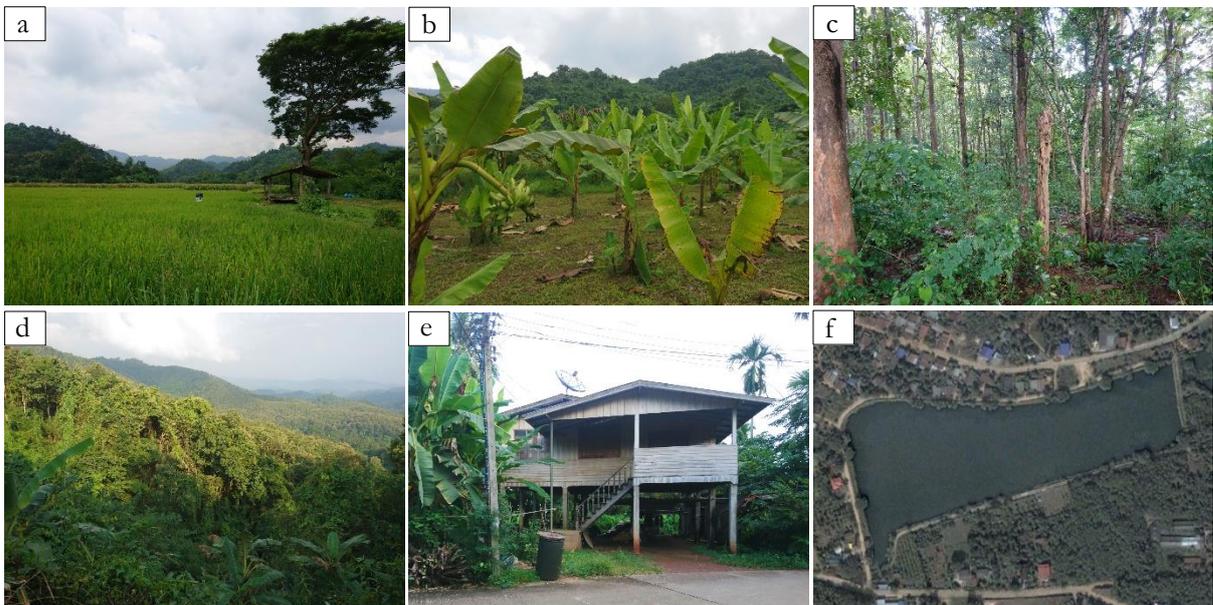


Figure 7. Land cover classes in the research areas. **a)** Cropland, **b)** Orchard, **c)** Teak Plantation, **d)** Mixed Forest, **e)** Urban, **f)** Water Bodies (Google Earth Pro, 2019).

3.3.3. Image processing

Surface reflectance products of U.S. Geological Service (USGS) Landsat 5 Enhanced Thematic Mapper (ETM) and Landsat 8 Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS) were used to create composite images for the years 2005 and 2018. ETM images include visible bands, two shortwave infrared bands as well as one thermal and one near infrared band. OLI and TIRS products provide visible bands, two short wave infrared bands, as well as two thermal bands and one near infrared band. Products from Landsat 7 ETM+ were excluded due to the Scan Line Corrector failure, which occurred after 2003.

Images were acquired for two timeframes. For 2005 the timeframe was defined from 1st of January 2005 until 20th of May 2006. This period was chosen as the flash flood event to be modelled happened on the 22nd of May 2006, and modelling efforts require a land cover map representing the condition of the area

before the event occurred. However, just two Landsat 5 images with insufficient quality (blurred) were available prior to this date in the same year. Hence, images from 2005 were included in the timeframe.

For 2018, enough images were available to create seasonal composites. The advantage of having seasonal composites is that seasonal variability of the vegetation is captured which adds additional information to the classification procedure later on (Flood, 2013; Rodriguez-Galiano et al., 2012). Therefore, seasonal composites for winter, summer and rainy season were generated. In the GEE code of Saah et al., (2019) the seasons for Thailand are defined as follows using calendar days:

- day 305 - 59 (+1 year) is defined as 'dry cool' season or winter;
- day 60 - 179 is defined as 'dry hot' season or summer; and
- day 180 - 304 (+1 year) is defined as 'rainy' season

Details all on images used for the composite of 2005 and the seasonal composites of 2018 can be found in Annex 3.

Atmospherically corrected orthorectified surface reflectance data products by USGS Landsat 5 and 8 are available in the GEE data archive. Images from Landsat Mission 5 have been atmospherically corrected using the LEDAPS algorithm (Schmidt et al., 2013) including a cloud, shadow, water and snow mask produced by CFMASK (Zhu & Woodcock, 2011). Landsat 8 images have been atmospherically corrected through the application of the Landsat Surface Reflectance Code as described by Vermote et al. (2016).

Following the image pre-processing steps of the SERVIR-RLCMS method (Saah et al., 2019), the data were subjected to further corrections to take into account for distortions caused by sensor, solar, atmospheric, and topographic effects (Young et al., 2017). Therefore, cloud and cloud shadow removal algorithms, as well as the bidirectional reflectance distribution function (BRDF), and topographic corrections have been applied. In Annex 3.1-3.4 GEE codes for image processing adapted from Saah et al. (2019) can be found.

Cloud and cloud shadow masking

The clouds were masked using the pixel-qa band and Google's cloudScore algorithm (Chastain et al., 2019). In general, cloudScore takes advantage of the spectral and thermal properties of clouds. It locates and removes pixels which are cold, and bright in all visible and infrared bands. To differentiate between clouds and snow, the Normalised Difference Snow Index (NDSI) was calculated (Salomonson & Appel, 2004). Chastain et al. (2019) provides details on the algorithm. In this study, a cloud score threshold of 20 was applied as it delivered the best results.

Cloud shadows were removed by utilising the Temporal Dark Outlier Mask (TDOM) (Housman et al., 2015). The algorithm detects dark pixels in the infrared bands (NIR, SWIR1, and SWIR2) by doing statistical outlier analysis. Pixels are identified as cloud shadow if they have not always been dark in previous or following images. If no uncontaminated pixel from previous or later images could be found as a replacement for a masked cloud or cloud shadow pixel, gaps with no data appeared. In that case, gap filling was done with pixels extracted from a composite of the previous year. The codes for cloud removal are shown in Annex 3.1 and for shadow removal in Annex 3.2.

BRDF correction

Surface reflectance measurements of Landsat images are subjected to directional effects. Those are caused by variations in solar zenith angle over time, and additional through different angles of vision during the

time of acquisition (Roy et al., 2016). The BRDF attempts to describe the directional reflectance (Roy et al., 2017). The method of Roy et al. (2016) was used for correction. GEE codes are available in Annex 3.3.

Topographic correction

Analysing any kind of remotely sensed imagery in respect to land cover types is challenging and not effective, without topographic corrections. Same objects on Earth happen to exhibit different reflectance values depending on their location in the terrain. Especially in mountainous areas terrain slope and aspect cause variations in illumination angle and reflection geometry (Colby, 1991). Compensation of those different solar illuminations can be achieved by topographic corrections (Riaño et al., 2003). In the literature, various correction approaches exist. The cosine correction by Smith et al. (1980), for example assumes that the total irradiance at a pixel is directly proportional to the cosine of the incidence angle. Hence, this model is applied when illumination is not originating from the zenith. On the other hand, the Sun-Canopy-Sensor Correction (SCS) model enhances the cosine correction by normalising the radiances from the sunlit canopy (Gu & Gillespie, 1998). However, the Modified SCS Topographic Correction method, also known as SCS+C (Soenen et al., 2005) was used which accounts for diffuse radiation caused by atmospheric and terrain sources and adjusts for terrain and at the same time preserves the sun-canopy geometry. Codes for topographic correction are available in Annex 3.4.

Composite assembling

Various methods exist for image compositing based on multiple images captured on different dates (Flood, 2013). For this study, median compositing was used. In median compositing, the median over a time series of images represented by an image collection of the defined timeframe is calculated. The output is an individual image with each pixel having the median value out of all images. By creating such a composite, it is possible to generate a representative image for the timeframe of interest without or with minimal noise caused by for example, clouds or cloud shadows (Flood, 2013).

3.3.4. Covariate layers

As already mentioned above, adding variables (covariates) to the classification enables the incorporation of contextual information to the classification procedure, which can improve the prediction of land cover classes (Cingolani et al., 2004; Domaç et al., 2006; Maselli et al., 1995; Rodriguez-Galiano et al., 2012). Covariate layers were mainly generated by exploiting the already available spectral bands of the composite images as it represents a simple, effective, and reliable form of adding contextual information (Boonprong et al., 2018). Therefore, Landsat median composite derivatives were calculated as covariates. Additionally, terrain derivatives extracted from the freely available SRTM digital elevation dataset (Farr et al., 2007) were added. In the course of this section, all included covariate layers are briefly explained, and reasoning for their incorporation is provided. The adapted GEE codes from Saah et al. (2019) for the covariate implementation are available in Annex 3.5 (seasonal composite) and Annex 3.6 (yearly composite).

Landsat derivatives

The bands of the median composites were used to calculate numerous normalised differences (NDs) and other more complex indices. The ND vector (Eq. 4) (Angiuli & Trianni, 2014) finds its application in image classification in order to extract information from spectral bands by removing errors caused amongst others by differences in space, time and acquisition (Angiuli et al., 2014).

$$ND(\text{Band}_1, \text{Band}_2) = \frac{\text{Band}_1 - \text{Band}_2}{\text{Band}_1 + \text{Band}_2} \quad (4)$$

Calculations comprised well known normalised indices such as the NDVI (Jeong et al., 2017), the NDWI (McFeeters, 1996) and the Normalised Burn Ratio (NBR)(Hislop et al., 2018). In Table 3, all applied band combinations for NDs are shown. Each of the indices was originally developed with the purpose to support the discrimination of different land covers (Angiuli et al., 2014). Additionally, two ratio bands dividing shortwave infrared 1 (SWIR1) and near-infrared (NIR), and red and SWIR1 were included as suggested by Saah et al. (2019).

Table 3. Band combinations for normalised differences; adapted from Saah et al. (2019).

Band 1	blue	green	red	NIR	SWIR ₁
Band 2	green	red	SWIR ₁	red	SWIR ₂
Band 2	red	NIR	SWIR ₂	SWIR ₁	
Band 2	NIR	SWIR ₁		SWIR ₂	
Band 2	SWIR ₁	SWIR ₂			
Band 2	SWIR ₂				

Furthermore, also the Enhanced Vegetation Index (EVI) (Jiang et al., 2008), the Soil-Adjusted Vegetation Index (SAVI) (Huete, 1988), the Index-based Built-up Index (IBI) (Xu, 2008), the Enhanced Built-up and Bareness Index (EBBI) (As-syakur, 2012), and the Tassel Cap transformation (Powell et al., 2009) were included as covariate layers.

EVI has been successfully used in land cover studies, for example, by Hussein et al. (2017) and Wardlow et al. (2007). Improvements in vegetation parameter sensitivity (e.g., Leaf Area Index (LAI)) in high biomass regions are obtained through correction of the soil background below the canopy and a reduction in atmosphere influences (Rayleigh scattering and ozone absorption) (Eq. 5) (Jiang et al., 2008).

$$EVI = G * \frac{NIR - red}{NIR + C_1 * red - C_2 * blue + L} \quad (5)$$

Where G is a gain factor (2.5); C₁ (6) and C₂ (7.5) are the coefficients of the aerosol resistance term, and L (1) is the soil-adjustment factor.

SAVI represents an advancement of the common NDVI which has limitations related to soil background brightness, according to Bausch (1993). By including a soil-adjustment factor (L =0.5), different from the one used in EVI, SAVI was found to adequately minimize soil background noise (Bausch, 1993). Therefore, SAVI is superior compared to NDVI in detecting vegetation in areas with low plant cover (Xu, 2008). This makes SAVI interesting with respect to the classification of young orchard plantations where plant cover is still developing or for other sparsely vegetated areas where intensified soil background noise can be expected. Equation 6 (Huete, 1988) was used for computation.

$$SAVI = (1 + L) * \frac{NIR - red}{NIR + red + L} \quad (6)$$

Urban areas are complex systems which lead easily to spectral confusion. By using thematic Index-derived bands rather than original images bands to construct the index, IBI suppresses background noise while retaining urban features. The index is computed by considering the three major urban components, namely soil, water and structures (Xu, 2008). This is achieved by making use of SAVI (Huete, 1988), the Modified

Normalised Difference Water Index (MNDWI) (Xu, 2006), and the Normalised Difference Built-up Index (NDBI) (Zha, 2003). After Xu (2008) the IBI is defined by equation 7.

$$IBI = \frac{NIR}{(NIR + red)} + \frac{green}{(green + SWIR_1)} \quad (7)$$

EBBI makes use of the thermal properties of urban areas to distinguish between built-up features and bare land (As-syakur, 2012). Hence, the TIR of Landsat is included (Eq. 8) (As-syakur, 2012). The index was considered as a useful covariate layer as its incorporation was expected to improve the classification in areas where agricultural land is adjacent to urban areas.

$$EBBI = \frac{SWIR_1 - NIR}{10 * \sqrt{SWIR_1 + TIR}} \quad (8)$$

The Tassel Cap transformation derivatives were additionally used as covariate layers as they were reported to be key variables for forest cover monitoring and mapping using RF by Boonprong et al., (2018). With the coefficients from Crist and Cicone (1984) and the composite bands the Tassel Cap transformation derivatives, “brightness”, “greenness”, ”fourth”, ”fifth”, and ”sixth” were calculated. Computations of the Tassel Cap angle (tcAngle) and distances (tcDist) for all pairs of brightness, greenness and wetness were done by equations 9 and 10, respectively (Powell et al., 2009).

$$tcAngle(band_1, band_2) = \arctan\left(\frac{band_1}{band_2}\right) \quad (9)$$

$$tcDist(band_1, band_2) = \sqrt{band_1^2 + band_2^2} \quad (10)$$

SRTM derivatives

Terrain indices were derived from the SRTM digital elevation dataset. Computed properties included elevation, slope, and aspect. Aspect was converted into two derivatives ‘eastness’ and ‘northness’. Eastness is a measure of the deviation from east and was calculated by the sine of the aspect, whereas the northness indicates the deviation from north, computed by the cosine of the aspect (Beers et al., 1966). The terrain indices were used as covariate layers since they have proved to be valuable environmental variables for land cover classification as shown by other studies, e.g. Brovkina et al. (2017), Domaç et al. (2006) and Koppad and Janagoudar (2013). Furthermore, also conceptually a relation between the indices and the parameter to predict (land cover) could be established:

- **Elevation** is expected to play a role for the land cover prediction, as cropland, for instance, is cultivated in lowland areas where irrigation is easier, whereas mixed forest predominantly grows at higher altitudes on summits, which are hard to access and therefore were spared by deforestation.
- **Slope** is expected to enhance the prediction of the land cover classes, as orchards were predominantly observed on sloping areas, and cropland, as well as urban areas, could be found in flat areas.
- **Aspect** is expected to influence the microclimate, including the amount of sunlight received, the temperature achievable, and the moisture retained. All those parameters might influence vegetation growth and therefore differences in land cover.

3.3.5. Reference data

The RF model was trained with reference data which were collected for 2005 and 2018 (Table 4). The reference data for 2018 were collected using a combination of high-resolution imagery (from 2018) available in GEE and visual image interpretation of the different seasonal Landsat composites. Additionally, OpenStreetMap (OSM) vector data (water bodies and road layer) were partially included as reference data.

For 2005, mainly the Landsat composite of 2005 was interpreted. Moreover, historical imagery in Google Earth Pro served as a sampling basis, and the SERVIR-RCLMS land cover map (2005) was used in a supporting manner. A greater amount of training points was collected for 2005 compared to 2018 since a larger area (both watersheds) had to be classified and OSM data could not be used for 2005 (Table 4).

Table 4. Number of training data points collected in each class.

Land cover class	No. training points (2005)	No. training points (2018)
Orchard	819	262
Cropland	385	235
Teak Plantation	210	179
Mixed Forest	711	365
Urban	446	412
Water Bodies	103	-
Total	2674	1453

3.3.6. Machine learning

The number of potential covariates is large and, in some cases, provides redundant information, for instance, when using the seasonal composites (e.g., SRTM derivatives). Eliminating covariates that provide little gains in accuracy reduces computational resources and avoids the inclusion of noise. Therefore, at each reference data point, all incorporated covariates, and each of the composites were sampled and the corresponding values evaluated in R (Breiman, 2001; Liaw & Wiener, 2002; R Core Team, 2018). In the process, a principal component analysis (PCA) was run on two measures which are commonly used to assess variable importance, Mean Decrease Accuracy and Mean Decrease Gini (Saah et al., 2019). In the former instance, the importance of a variable is assessed based on the mean decrease of the accuracy when that variable is excluded. The more the accuracy of the RF decreases, the greater is the importance of that respective variable for classifying the data. The latter measures the contribution of a variable to the homogeneity of the nodes. A higher decrease in the Gini coefficient indicates that the respective variable results in nodes with higher purity. Finally, all variables which resulted in a PCA coefficient greater than 0 were selected for the classification, as they can be considered having relative importance for the RF classifier according to Saah et al. (2019). Selected variables for both classifications can be found in Annex 3.7. Once the covariates were put in place, the RF classifier in GEE was applied with the selected bands. The model was run with 100 trees as suggested by Saah et al. (2019), and the number of variables was set to the square root of the number of variables (default of GEE). Not all reference points were included from the beginning. The classification procedure was done on a step by step basis with many runs. Meaning, some reference points were collected for each class and only then the RF model was applied. In a next step the resulting land cover map was then used for additional sampling. In this way the reference dataset was built and sample points leading to miss-classification could be excluded in the next run and replaced by others.

3.3.7. Validation data

For 2018, 315 validation data points were composed out of different sources. Firstly, 58 out of 62 land cover samples collected in the field were used as validation data points (Table 5), 6 sample points had to be excluded because the respective land cover area was less than the pixel size (30x30 m) and therefore considered as not suitable.

Table 5. Number of field validation points for each class.

Land Cover	No. of samples	No. of samples excluded
Teak Plantation	7	4
Orchard	40	0
Cropland	9	0
Mixed Forest	6	0
Total	62	4

Secondly, 1614 in GEE collected training points got a random number between 0 and 1 assigned (without recurrence). Afterwards, they were divided into two sets, one including all points with values <0.1 (10 %) and the other one containing all points with values >0.1 (90 %). The first set (<0.1) with 163 points was exported as validation data, and the second set (>0.1) was used to train the model. Thirdly, 94 additional validation points were collected using Google Earth Pro. For 2005, the 461 validation points were exclusively collected in Google Earth Pro, making use of historical imagery. Table 6 shows the distribution of the validation points among the various classes for both years.

Table 6. Validation data points for 2005 and 2018.

Land cover class	No. validation points (2005)	No. validation points (2018)
Orchard	100	71
Cropland	92	40
Teak Plantation	40	66
Mixed Forest	97	49
Urban	97	89
Water Bodies	35	0
Total	461	315

3.3.8. Accuracy assessment

In the past, the necessity of validation of image classification products was not widespread, and results were simply considered to be correct (Congalton, 1991). With the rise of digital image classification, the determination of the information value of classifications is now vital (Rwanga & Ndambuki, 2017). Hence, creating an error matrix and computing the overall classification accuracy, which presents the simplest measure to retrieve quality information, are now common practice for validation assessment in digital image classification. One who wants more detailed information can make use of the user's and producer's accuracy.

In this study, the overall accuracy and user's as well as producer's accuracies were determined. The overall accuracy considers the correctly classified reference points, which can be found on the diagonal of the error matrix and divides it by the total number of references sites, providing an accuracy value in percent (Congalton, 1991). The producer's accuracy, on the other hand, quantifies how often real features on the

ground are correctly shown on the classification product. Hence, it is computed by dividing the number of correctly classified reference points by the total number of references points of the respective class. Accordingly, the user's accuracy does indicate the reliability of the map by providing information on how often a pixel on the map represents that class on the ground. Computation is done by the division of the classified pixels with the row totals (Congalton, 1991).

Another discrete multivariate statistical technique used in this research is KHAT. Cohen (1960) introduced this method where the resulting KAPPA coefficient is another measure of accuracy. Its upper limit is 1, indicating perfect agreement, whereas results <0 are considered as poor. Detailed descriptions of all mentioned methods can be found in Congalton (1991) and Rwanga et al. (2017). With equation 11, the KAPPA coefficient was computed.

$$K = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r (x_{i+} \times x_{+i})}{N^2 - \sum_{i=1}^r (x_{i+} \times x_{+i})} \quad (11)$$

Where N = total number of observations, r = number of rows in the error matrix, x_{ii} = number of observations in row i , x_i and x_{+i} = marginal totals of row i and column i , respectively.

3.3.9. Land cover change analysis

Various methods of remote sensing based land cover change detection are available, in-depth descriptions can be found among others in Comber et al. (2016) and Mas (1999). For this study, post-classification change detection comparison was used to identify and quantify land cover changes between 2005 and 2018. This method was chosen as it represents a solid commonly used method characterized by its simplicity and the capability of comparing satellite imagery captured under different conditions and with different sensors. Examples of successful applications can be found in Kim (2016), Mallupattu and Reddy (2013) and Lubis and Nakagoshi (2011) as well as in Mas (1999). In post-classification change analysis, two or more classifications based on images from different times are examined in a comparative manner as already suggested by the name. Due to the nature of this method, it is evident that the accuracy of the results highly depends on the accuracy of the classifications used.

The ultimate goal of the land cover change analysis was to learn more about the dynamics in Ban Da Na Kham watershed, in other words, to acquire information about the "from-to" processes between the land cover categories. To fulfil this task, a change detection matrix was established. From such a matrix, the magnitude and direction of the change can be extracted. According to Lubis et al. (2011), the diagonal values in the matrix indicate areas which remained unchanged in each land cover class while off-diagonal elements provide information about occurred changes. The off-diagonal elements contain information about area increase and area loss of the different classes. Vertical values (columns) indicate an increase for a certain class, whereby horizontal values (rows) indicate a loss of area in a certain class.

3.4. OpenLisem flood model

Flash flood modelling was done in openLISEM to satisfy the objectives of this research and to answer the following research questions:

Objective 1: Comparative analysis of SG versus FD for flash flood modelling;

- Is model calibration based on historical flood marks from a nearby watershed possible?
- What are the quantitative differences of the model output using, (I) FD, and (II) SG in relation to flow dynamics?
- What is the sensitivity of the flood dynamics to different soil depths using, (I) FD, and (II) SG?

Objective 2: Analysis of the effects of land cover change on flash flood behaviour;

- Which land covers generate the highest average runoff?
- How do these land cover changes affect the runoff generation and flood dynamics?

OpenLISEM was chosen for this research study as it is an event-based integrated catchment model operating in time steps of minutes and is therefore tailored to model flash flood processes. Furthermore, it has the capability to account for detailed spatial variability in terms of land cover, soil and terrain. Moreover, processes such as groundwater flow and evapotranspiration, which are superfluous when modelling single events, are neglected (Bout et al., 2018). The applicability of openLISEM in flash flood modelling was demonstrated by previous studies conducted in different parts of the world, e.g. by Pérez-Molina et al. (2017) and Sliuzas et al. (2013) in Kampala, Uganda and Van Westen et al. (2015) in the Caribbean.

Originally the model was developed to estimate soil erosion but is now also used to assess flash floods and debris flows. Modelled processes include, amongst others interception, infiltration, runoff generation, flow, and sediment transportation. Vertical water flow in the soil, overland flow and channel flow, as well as channel flooding, are important components of flash floods. The erosion component of the model includes splash and flow detachment, as well as sediment transport and deposition (De Roo et al., 1996). Within openLISEM land cover types are represented by various resistance values (Manning's n) and hence influence the flow behaviour. The effects of soils are determined by their respective infiltration rate and infiltration capacity (porosity). Additionally, the model can include built structures such as urban areas or road networks, which represent mostly areas where no or limited infiltration takes place.

The infiltration process is very important when modelling flash floods as it determines the amount of runoff generated. In openLISEM infiltration is simulated by using the model of Green and Ampt (1911) which assumes a downward moving wetting front due to surface water infiltration. Soil layers above the wetting front are assumed to be saturated, and below the front, the soil possesses initial soil moisture. With equation 12, the potential infiltration rate is calculated and subsequently subtracted from the available surface water (Bout et al., 2018).

$$f_{\text{pot}} = -K_s \left(\psi \frac{\theta_s - \theta_i}{F} + 1 \right) \quad (12)$$

Where f_{pot} = potential infiltration rate (m s^{-1}), F = cumulative infiltrated water (m), θ_s = porosity ($\text{m}^3 \text{m}^{-3}$), θ_i = initial soil moisture content ($\text{m}^3 \text{m}^{-3}$), ψ = matric pressure at the wetting front (m)

Water that cannot infiltrate into the soil due to exceedance of infiltration rate or infiltration capacity and is moreover not stored in micro depressions is described as runoff water. This type of water can result from direct rainfall or from runoff generated on neighbouring cells. Furthermore, a Green and Ampt solution for two soil layers is used with their characteristics; these are named soil layer SL1 and SL2 (see section below on Soil Data).

Overland flow is simulated by either kinematic flow, diffusive flow or dynamic flow (Saint-Venant) (Bout et al., 2018). When using kinematic flow, the water is routed via a predefined converging flow network. Velocity is determined by friction forces and gravitational force. In the diffusive flow approximation, velocity is additionally controlled by the hydraulic gradient. However, for this study, the dynamic flow was used as it additionally takes the momentum terms into consideration. Furthermore, Bout et al. (2017) proved that using the dynamic flow most accurate flow simulations could be achieved. According to Bout et al. (2017), channel flow is generally decoupled from overland flow and routed by the kinematic wave. Whenever runoff water reaches the channel it is added to the channel water. Flow in the channel is always computed using a kinematic wave. In case of a channel overflow, this water is added to the overland flow water (Bout et al., 2017).

3.4.1. Data preparation

OpenLISEM requires input data generated from four main data categories, namely rainfall, soil, land surface and terrain properties (Table 7). To develop a modelling environment as close as possible to field conditions, the parametrisation is done by including FD whenever available. By nature, this is limited due to time and other resources. Hence, remote sensing products like a DEM, satellite images, as well as pre-existing GIS-data and literature values support the input generation. Below the methodology used to generate the subordinated input data for each category is elaborated.

Table 7. Main input data required for openLISEM.

Category	Input data	Comments
Rainfall	Text file	<ul style="list-style-type: none"> • Single rainfall event based on daily rainfall from 22nd of May 2006
Topography	DEM	-
	Slope map	<ul style="list-style-type: none"> • DEM derived
	Local drain direction map (LDD)	<ul style="list-style-type: none"> • DEM derived
	Channels	<ul style="list-style-type: none"> • Channels were generated based on LDD map; dimensions based on field measurements and Google Earth Pro
Surface	Land cover	<ul style="list-style-type: none"> • Generated land cover maps for 2005 and 2018
	Manning's n	<ul style="list-style-type: none"> • Based on literature
	Roughness	<ul style="list-style-type: none"> • Based on literature
	Vegetation cover	<ul style="list-style-type: none"> • Based on field measurements
	Leaf Area Index	<ul style="list-style-type: none"> • Based on empirical equation
	Road network	<ul style="list-style-type: none"> • Generated based on OpenStreetMap data
Soil	Saturated hydraulic conductivity	<ul style="list-style-type: none"> • Laboratory analysis and SoilGrids
	Porosity	<ul style="list-style-type: none"> • Laboratory analysis and SoilGrids
	Initial soil moisture	<ul style="list-style-type: none"> • Initial soil moisture based on antecedent rainfall and literature
	Matric suction	<ul style="list-style-type: none"> • Matric suction calculated based on initial soil moisture and soil texture
	Depth	<ul style="list-style-type: none"> • Based on topographical factors

Rainfall Data

Daily rainfall data for the period from 1951 to 2018, provided by the Thai Meteorological Department for the meteorological station in Uttaradit city (Station ID: 351002), was processed. Based on Gumble analysis, the return period (RP) of the extreme event in 2006 was determined. Subsequently, Intensity Duration Frequency (IDF) curves for Uttaradit province published by Rittima et al. (2013) were used to generate 10-minute rainfall intensities (Annex 4). Lastly, to compile a design storm, the Alternating Block method as introduced by Yen and Chow (1980) was utilised.

Digital elevation model

The ALOS Palsar DEM with a spatial resolution of 12.5 m was obtained from the Alaska Satellite Facility (ASF). With the tool gdalwarp the DEM was resampled to 15 m. Resampling with bilinear interpolation was done for two reasons, (I) to speed up the modelling process but retaining sufficient spatial detail at the same time, and (II) to smoothen the DEM. Smoothing represented a positive side effect of the resampling, as the original DEM exhibited a “LEGO-like” surface due to the fact that elevation values are given as integers, not as fractional numbers. Subsequently, a simple pit filling algorithm was applied to restore flow pathways.

Most of the ALOS Palsar radiometric terrain corrected products are created from SRTM DEM source, which has a vertical accuracy of less than 16 m (ASF, 2015). Furthermore, C-band radar, which is used to create these DEMs, does not penetrate well through areas of dense vegetation. Hence, these areas exhibit elevations related to the top of the canopy (according to email correspondence with ASF).

For obvious reasons, it is expected that a DEM which does not relate to the bare soil surface might cause considerable problems in hydrological modelling. Hence, it was decided to run openLISEM in the calibration watershed with three different versions of the ALOS Palsar DEM, and additionally with the freely available SRTM DEM from USGS. This choice was made to investigate differences in the model output and to test the applicability of the DEMs for hydrological modelling in the watershed. The three versions were assembled as follows: (I) original DEM (DEM_o), (II) DEM without vegetation (DEM_v), and (III) DEM partially manipulated (DEM_m). Generation of DEM_v was accomplished by subtraction of on-site measured vegetation heights from DEM_o . The average vegetation heights used can be found in Table 8. For DEM_m elevation values were partially manipulated in narrow valleys along with the drainage network in order to restore a functional drainage behaviour of the terrain as vegetation is expected to generate artificial dams blocking the water flow.

Channels

Channel layers were created for both watersheds, for Ban Da Na Kham and for Laplae, based on the DEM_o derived LDD network. Validation of the channel location was subsequently done based on Google Earth Pro. In the case of deviating channel locations caused by errors in the DEM, channels were manually digitised in ArcGIS and PCRaster. This was mainly necessary for the lowland areas where rivers flow in flat terrain. Besides channel location, also channel dimensions are an important input for openLISEM. Therefore, dimensions were observed and measured during fieldwork at several locations in order to have benchmarks for the channel parametrisation (Annex 5 & Annex 5.1). The channel dimensions for Laplae watershed were approximated based on visual field observations and measurements in Google Earth Pro.

Surface parameters

A diverse set of surface parameters is necessary to represent the different land cover types in the modelling environment (Bout et al., 2018). Manning's n and surface roughness (RR) were parametrized based on

literature values. Vegetation overstory density (cover) was measured in the field using a spherical densiometer (Model A), following the instructions of Lemmon (2008). Additionally, the average vegetation height per land cover class was estimated based on field observations. To estimate the storage capacity of the canopy, the LAI was calculated accordingly with equation 13. In Table 8, all surface parameters used are shown.

$$LAI = \frac{\ln(1 - \text{cover})}{-0.4} \quad (13)$$

Table 8. Surface parameter for openLISEM.

Land Cover	Manning's n	Source	RR*	Cover (%)	Height (m)	LAI
Orchard	0.025	Morgan et al. (1998)	1	57	9.7	2.1
Cropland	0.04	Chow (1959)	0.1	47	0.9	1.6
Teak Plantation	0.1	Arcement and Schneider (1989)	1	80	21.5	4.0
Mixed Forest	0.2	Hessel et al. (2003)	1	78	14.3	3.8
Urban	0.0678	Kalyanapu et al. (2009)	0.5	2	9.7	0.6
Water	0	-	0.1	0	0.0	0.0
Channel	0.035	Chow (1959)	-	-	-	-

* RR adapted from Bout et al. (2018)

Soil data

Two sets of soil data were compiled, (I) FD analysed in the laboratory, and (II) SG data. Soil properties are changing with increasing depth, as for instance SOM content and microbial activity are declining (Hassler, 2013). To be able to represent the soil landscape as realistic as possible, it was decided to use a two-layer soil system. As no opposing information available, it was decided to close the soil at the bottom and therewith prevent free drainage due to percolation.

Setup - FD:

For the surface layer (SL1) which possesses a thickness of 5 cm, K_s and porosity were parameterised based on values derived from the soil analysis (Section 3.2.6). This value for thickness was chosen as it corresponds to the thickness of the collected soil samples and additionally matches the distribution of the observable SOM content and the presence of bioturbation. Only surface samples were collected. Therefore, no immediate values for K_s and porosity for a second soil layer (SL2) were available. This lack was remedied by using the PTFs by Saxton and Rawls (2006). As input for the PTFs, the PSD retrieved from laboratory analysis was used. Since SHPs of deeper soil layers are not expected to be influenced by the land cover and land use but rather by texture, it was decided to generate SL2 based on the aggregated PSD within each slope unit and to set SOM to 0 %. Since the soil depth of SL2 is expected to vary in space, its generation is discussed separately below.

Setup - SG:

For the SG setup PSD, D_b , gravels and soil organic carbon (SOC) (g kg^{-1}) layers for depths of 5 cm (SL1) and 100 cm (SL2) were converted into K_s , porosity and matric suction maps based on the PTFs from Saxton and Rawls (2006).

Initial Soil Moisture and Matric Suction:

Firstly, analysis of the historical daily rainfall data revealed that almost 135 mm of rainfall occurred in the two preceding days of the event to model (Table 9).

Table 9. Daily rainfall prior, during (red square) and after the event.

Date (May 2006)	14	15	16	17	18	19	20	21	22	23	24
Rainfall (mm)	25.6	17.4	0.5	0	7.9	0	14.5	120.4	263.7	13.4	8

Secondly, a post-disaster assessment report of the event in 2006 suggested that the tremendous landslides and flash floods were caused by a saturation of the topsoil layer (30-40 cm) (Usamah & Arambepola, 2013). Based on this information, the initial soil moisture content was set to 0.90 of the porosity to ensure saturation during the rainfall event. However, this parametrisation introduces an unavoidable wrong assumption that the soils are homogeneously saturated regardless of other influencing factors such as K_s , land cover and soil depth. OpenLISEM only allows up to two soil layers when using Green and Ampt. Furthermore, a gradual decrease of initial soil moisture corresponding to increasing depth can only be implemented when the SWATRE module is used. This model is based on the full Richards equation for multiple layers and beyond the scope of this research.

Subsequently, the matric suction (Psi) was calculated accordingly based on the Brooks and Corey Model (Brooks & Corey, 1964) (Eq. 14).

$$\text{Psi} = S_e^{(-\lambda)} h_b \quad (14)$$

Where S_e = effective saturation, h_b = bubbling or air entry pressure head, and λ = is the pore size distribution index. Values for h_b and λ depend on soil texture and were adapted from Saxton et al. (2006). Psi maps for openLISEM were generated based on the slope unit map as soil texture is expected to relate rather to the terrain than to the land cover. Correspondingly, for computation of Psi, the PSD of all soil measurements within one slope unit were aggregated and an overall texture class estimated. Subsequently, h_b and λ were extracted from Saxton et al. (2006).

Soil Depth:

Soil depth data was not available in the research area. During the fieldwork, soil depth measurements were taken at the 48 soil sampling sites using a soil auger. These measurements indicated soil depths of at least 1 m at each location, but at the same time, the measurement activities were limited to 1 m because of the length of the auger. Due to these limitations and having the knowledge that soil depth varies within the landscape, the approach of Kuriakose et al. (2009) was adopted to generate a soil depth map. This approach takes topographic factors such as altitude, slope, wetness index and profile curvature into account to approximate soil depths within the landscape. A sort of calibration was done based on the field measurements, meaning that at each measurement point the soil depth has at least 1 m depth. However, the lack of a minimum and maximum soil depth, as well as the lack of validation data, led to a large unavoidable uncertainty. In order to investigate this uncertainty, four additional soil maps were generated each having 1 additional meter of soil. In the course of the modelling, openLISEM was subsequently run with the SHPs of FD and SG with varying soil depths to investigate how flow dynamics change in the watershed with an increase in soil depth. The PCRaster script to generate the soil depth maps is available in Annex 6.

3.4.2. Model calibration

After the flash flood event on the 22nd of May 2006 the community of Laplae painted marks on the power poles along the major roads in the city of Si Phanommat to indicate the flood height and flood extent (Figure 8). Using a measuring tape and GPS receiver, data on the historical flood heights were captured in 45 locations (Figure 8). Based on these measurements, attempts were taken to calibrate the model.

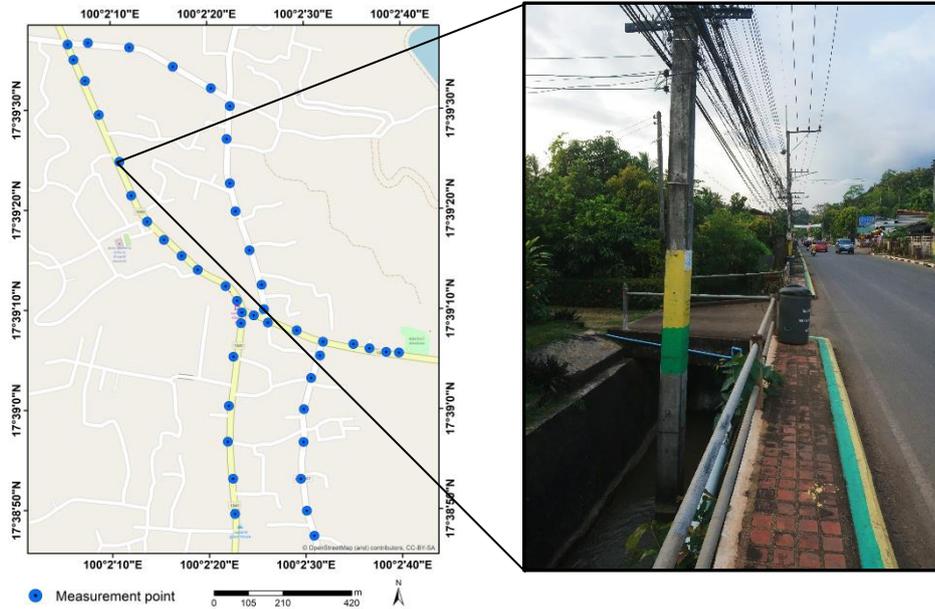


Figure 8. Locations of historical flood height measurements in Si Phanommat city (left) and flood mark painted on a power pole (right).

For this purpose, the model was run in Laplae watershed using the three versions of the ALOS Palsar DEM and the SRTM DEM. Subsequently, the number of points with a simulated flood depth ≥ 0.3 m was calculated to identify the DEM having the best agreement in terms of the presence of flood. The water depth of 0.3 m was selected as lower limit as every inundation below that threshold was expected to represent the effects of pure precipitation rather than by channel flooding.

4. RESULTS AND DISCUSSIONS

4.1. Soil analysis

In this subsection, the results of the general statistical analysis of soil data are described. Firstly, the results of the slope unit delineation are presented, as it served as an input for further analysis. Secondly, the FD was put in relation to land cover types and slope positions to investigate possible effects and how they are reflected by the soil properties. Thirdly, an examination of SG based on the land covers and slope positions was done to see if similar patterns compared to the FD can be observed. In the last part, similarity analysis between the FD and SG is presented, making use of the Cosine Similarity measure and the Wilcoxon Signed-Ranked test. In the course of this section, the following research questions will be dealt with:

- How well do SG and FD correlate?
- How do the soil properties of (I) FD, and (II) SG relate to the main land cover types?
- How do the soil properties of (I) FD, and (II) SG relate to the terrain?

4.1.1. Slope units

Slope unit delineation was aided by the GRASSGIS extension called “Geomorphon”. Best results were achieved applying a search radius of 2000 m, a skip radius of 15 m, and a flatness threshold of 0. Figure 9 shows the spatial distribution of the slope units in the research area, with the narrow valley floors and similarly narrow summits. The accuracy assessment of the slope position delineation yielded an overall accuracy of 79 % (Table 10). Misclassification is expected to be caused by the deviation of the GPS location due to poor signal within the forested and mountainous areas. Also, the resolution and quality of the DEM need to be taken under consideration as a potential source for wrongly classified pixels. According to Miller et al. (2015), the identification of small variations in topography and the correct classification of slope positions does depend on the DEM quality and resolution, used for the classification. This seems to be especially important in areas where valley floors and summits are located relatively close to each other, due to steep sloping terrain, as observed in the research area. Where one pixel of the DEM represents the elevation of the summit, an adjacent pixel could already provide information on the valley bottom. Furthermore, should the DEM quality issues as described in Section 3.4.1 be taken into consideration. Having the vegetation height on top, partly wrong classified pixels can be expected.

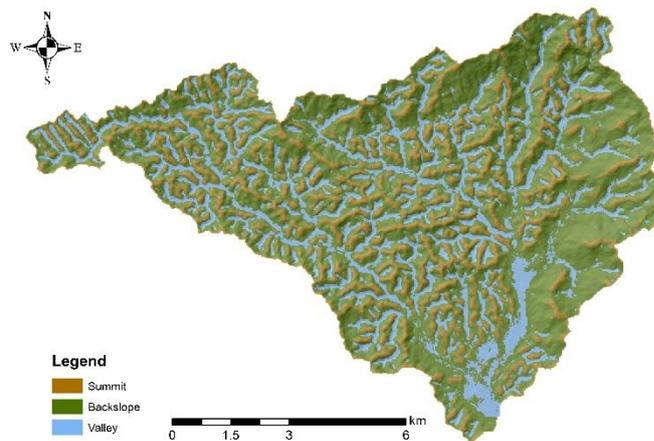


Figure 9. Slope position map.

Table 10. Confusion matrix of slope unit classification.

Geomorphon	Field Observations			Row Total	User Accuracy (%)
	Backslope	Summit	Valley		
Backslope	18	4	2	24	75
Summit		4		4	100
Valley	4		16	20	80
Column Total	22	8	18	48	Total Accuracy 79%
Producer Accuracy (%)	81.8	50	88.8		

4.1.2. Influence of land cover on soil properties

Various descriptive statistics were computed for the FD (Table 11 & Figure 10). Striking are the field measurements of K_s , with values ranging from 0.1 mm h⁻¹ up to 10784 mm h⁻¹. It is worth to note that K_s values above 2000 mm h⁻¹ were considered as arbitrary high and therefore excluded from further analysis. Details on the omitted data can be found in Annex 2.3.

Table 11. Statistics of soil properties per land cover.

Land Cover		Variable							n
		Clay (%)	Silt (%)	Sand (%)	K_s (mm/h)	Por (%)	D_b (g/cm ³)	SOM (%)	
OR	Mean	18	42	40	411	54	1.2	4.7	34
	SD	11	7	14	446	6	0.2	1.3	
CR	Mean	21	37	42	65	42	1.5	2.5	4
	SD	7	9	15	23	6	0.2	0.4	
TK	Mean	7	42	51	594	51	1.3	4.6	2
	SD	7	5	2	177	7	0.2	2.0	
MF	Mean	30	40	30	722	61	1.0	5.2	8
	SD	10	6	14	750	8	0.2	1.8	

OR = Orchard, CR = Cropland, TK = Teak Plantation, MF = Mixed Forest, K_s = Saturated Hydraulic Conductivity, Por = Porosity, D_b = Bulk Density, SOM = Soil Organic Matter, SD = Standard Deviation, n = number of samples

In the research area, it becomes obvious that mixed forest represents the land cover class with the highest max K_s values (1986 mm h⁻¹), followed by perennial crops comprising long kong trees and banana plants with 1782 mm h⁻¹, and teak plantations with 719 mm h⁻¹ (Figure 10). Both orchards and mixed forest show also a notable high standard deviations with 446 mm h⁻¹, and 750 mm h⁻¹, respectively, indicating great variation within the investigated territory (Table 11). The exceptional high K_s values measured might be explained by the intensive bioturbation in the form of abandoned root channels and holes excavated by insects living in the near-surface layer as well as high SOM content that were observed during the fieldwork. This observed near-surface soil layer condition is additionally reflected by high porosity values of up to 76 %, the corresponding D_b of 0.6 g cm⁻³, and SOM content of on average more than 6 % (Figure 10 & Table 11). The findings are in line with observations made by Hassler et al. (2011) and Hassler (2013), who investigated K_s variability in the tropics and reported measurements of more than 1400 mm h⁻¹ within forested areas. In contrast, Hassler (2013) reported the lowest measurements of 0 mm h⁻¹ in all land cover classes, referencing to great variability.

In order to get an overview of the data and to find possible linkages, K_s was correlated with ancillary variables such as D_b , SOM content and PSD. This was done using all data points based on the land cover or slope position without any subdivision. K_s was kept as a constant variable during correlation because it is known that other ancillary variables influence K_s (Hassler et al., 2013). Subsequently, the same procedure was applied, but in each land cover class (Table 12).

Since the macro-porosity of soils has a great impact on D_b , which again has a positive relationship with K_s (Ahuja et al., 2010), it can be argued that drivers affecting K_s similarly affect D_b . Examples for such drivers are abandoned root channels or microorganism activity. Furthermore, SOM content has a positive relationship with soil structure (Lado et al., 2004). Thus, it is expected that K_s shows an increase with increasing SOM content. Also, the percentage of sand, silt and clay may influence the K_s , whereas after García-Gutiérrez et al. (2018) the main determinant in respect to K_s is the percentage of sand.

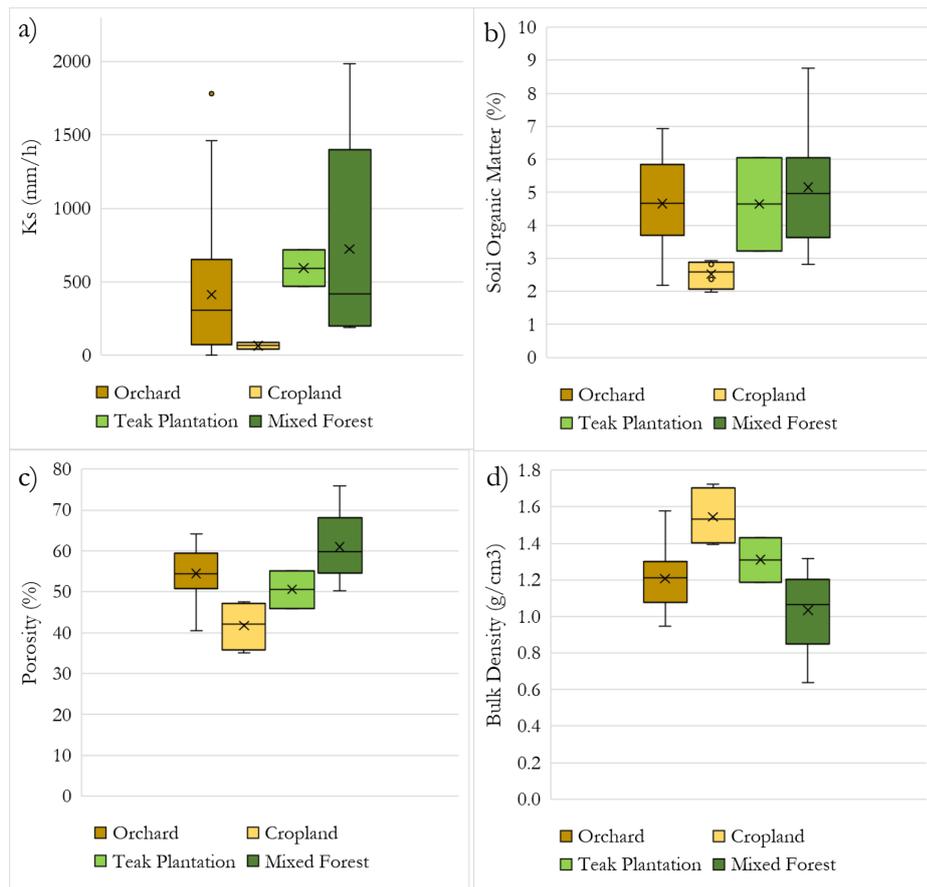


Figure 10. Variability of soil properties per land cover; **a)** Saturated Hydraulic Conductivity, **b)** Soil Organic Matter, **c)** Porosity, **d)** Bulk Density.

A statistically significant correlation with $P \leq 0.01$ and a medium correlation coefficient of $r = 0.5$ was found between K_s and D_b considering the complete dataset. At the same time, correlations with other variables were rather low and not statistically significant.

Further, it is unlikely that the texture class influences the K_s of the collected soil samples indicated by the absence of correlation between the percentage of sand, silt and clay. For the tropics, Hassler et al. (2011) compared spatial means of K_s between different soil map units and did not observe any differences.

Likewise, Sobieraj et al. (2002) faced difficulties with explaining changes in K_s along a catena and Hassler (2013) also failed to identify a correlation between differences of K_s and sand, silt or clay content. In line and with the thoughts of the above-mentioned authors, it can be concluded, that it is very likely that biotic factors predominate the formation of soil properties within the study area.

In the next step, K_s measurements were correlated with the ancillary variables within different land cover classes. Since the available K_s measurements for cropland was reduced to 3 after applying the threshold mentioned above, and the teak plantation class was just represented by two samples, the correlation was done within the orchard and mixed forest class only (Table 12). Subsequently, a statistically significant result were found within the mixed forest class with $P \leq 0.05$ and a high correlation of $r = -0.92$ with D_b (Table 12). Additionally, SOM and K_s correlated very well with $r = 0.75$, but no statistical significance was exhibited (Table 12). This confirms the observations from Lado et al. (2004) that SOM positively influences the soil structure and therewith K_s , and at the same time diminishes compacted soil conditions.

Table 12. Correlation (r) of ancillary variables with K_s per land cover.

Variable	Orchard	LK	Mix	Banana	Cropland	TP	Mixed Forest
D_b (g/cm ³)	n.s.	n.s.	n.s.	n.s.	n.a.	n.a.	-0.92*
Clay (%)	0.37*	n.s.	n.s.	n.s.	n.a.	n.a.	n.s.
Silt (%)	n.s.	n.s.	n.s.	n.s.	n.a.	n.a.	n.s.
Sand (%)	n.s.	n.s.	n.s.	n.s.	n.a.	n.a.	n.s.
SOM (%)	n.s.	n.s.	n.s.	n.s.	n.a.	n.a.	0.75

D_b = Bulk Density, SOM = Soil Organic Matter, LK = Long Kong, Mix = Banana and Long Kong, TP = Teak Plantation, n.s. = not significant, n.a. = not available, * = statistically significant at $P < 0.05$ level

More than 55 % of the analysed orchards are located in steep sloping terrains, which favour the loss of surface material due to erosional processes (Schaeztl et al., 2005), even in tropical forest-like ecosystems (Sidle et al., 2006). Thus, it can be assumed that the considerably high variability expressed by the standard deviation of each orchard class and the significant relationship with $P \leq 0.05$ and $r = 0.37$ between K_s and clay particles originates partly from erosion (Table 12 & Table 13). Loss of the litter cover and an increase of soil surface exposed to precipitation could lead to a poorly developed A horizon and thereby potentially transform an alleged surface sample in a sub soil sample according to Hassler (2013). This result urges the possibility that the influence of land cover is decreasing with increasing soil depth and that soil properties of deeper soil layers, e.g. below 5 cm depth, might rather be controlled by soil texture.

Interesting is the fact that in the orchard class, just weak correlations were found between K_s and SOM and without any statistical significance (Table 12). This raises the question of why a strong relationship was observed within the mixed forest class (Table 12). Nemes et al. (2005) discussed similar contradictory interactions of SOM and soil properties. An explanation could be a possible enhancement of the flow of water through the soil in the mixed forest class by improved porosity. On contrary to that, the SOM in the orchard class could cause a filling of pores in the soil, leading to tortuous and thin pathways decelerating the water flow. This would result rather in water retention by SOM, explaining the weak correlation of SOM and K_s within the orchard class. Another reason could be that the floor of the mixed forest was covered by small seedlings and dense undergrowth which are potentially also a source of soil structure enhancement next to the SOM. In conclusion, SOM content seems to have a varying effect on the soil water characteristics within different land cover classes in the study area and even within just one land cover class.

In order to gain a better understanding of the variability within the orchard class, the class was divided into subclasses based on fruit trees; namely long kong, banana and long kong (mix), and pure banana.

The disaggregation held interesting insights as the K_s , D_b and SOM differ within the subclasses (Table 13). It can be observed that mean K_s and mean SOM content decrease from pure long kong plantation with K_s of 521 mm h⁻¹ and SOM of 4.9 % to areas with only banana trees planted showing a K_s of 140 mm h⁻¹ and a SOM of 4.1 % (Table 13). In contrast to that, the mean D_b is increasing from 1.1 g cm⁻³ to 1.3 g cm⁻³ indicating increasing compaction (Table 13). Also, the composition of sand, silt and clay is going through a transition with having a higher mean percentage of clay (23 %) within the pure long kong plantations decreasing to a mean of 11 % in the pure banana plantations. At the same time, the sand component is rising more than 16 % to a mean of 51 % (Table 13). The finding entails that soil texture is not necessarily the predominant factor causing the alteration of the soil physical properties within the fruit tree classes. Rather it can be assumed that it is the source and proportion of organic material determining the changes in the physical properties. Hence, overripe fruits dropped by the long kong trees, as observed during the fieldwork could be one cause for the differences in SOM content among the orchard types. This phenomenon was investigated by several studies, e.g. see De Paz et al. (2018), Ohm et al. (2007) and Hamer and Marschner (2005). Leaking liquid from the fallen fruits enriches the soil with sugar, which intensifies the microbial growth and consequently enhances the decomposition rate of the fallen leaves. Further, an acceleration of the leaf decomposition was observed by De Paz et al. (2018), not only triggered by the absolute sugar content of the fruits, rather by factors such as fruit moisture and micronutrient content. This corresponds with the properties of long kong fruits being fleshy with approximately 84 g water per 100 g long kong (Tilaar et al., 2008).

Table 13. Soil properties per orchard subclass.

Land Cover		Variable							n
		Clay (%)	Silt (%)	Sand (%)	K_s (mm/h)	Por (%)	D_b (g/cm ³)	SOM (%)	
LK	Mean	23	45	32	521	57	1.1	4.9	34
	SD	10	5.5	12.8	409	4.2	0.1	1.3	
Mix	Mean	16	41	43	470	55	1.2	4.6	4
	SD	8.6	6.4	10.4	581	5.4	0.1	1.5	
BN	Mean	11	38	51	140	49	1.3	4.1	2
	SD	10.1	9.2	11	134	6.2	0.2	1.0	

LK = Long Kong, Mix = Banana and Long Kong, BN = Banana, n = number of samples, K_s = Saturated Hydraulic Conductivity, Por = Porosity, D_b = Bulk Density, SOM = Soil Organic Matter, SD = Standard Deviation

In contrast, K_s in the cropland is moderate, with an average value of 65 mm h⁻¹ and a standard deviation of 23 mm h⁻¹ (Table 11). Comparable high is D_b with a mean of 1.5 g cm⁻³ and a maximum of 1.7 g cm⁻³ corresponding to a relatively low SOM content with a mean of 2.5 % and a maximum of 2.9 % (Table 11 & Figure 10). Increased D_b can also be an indication of disrupted pores by cultivation, as outlined by Kizilkaya and Dengiz (2010).

It should be noted that predominantly areas with corn were sampled while most of the cropland in the research area is represented by paddy fields. Unfortunately, paddy fields could not be sampled, being already tilled and flooded. A drastic drop of mean K_s within the cropland class is expected if the soils of paddy fields would be taken into account. This expectation is supported by Wijaya et al. (2009) and Boumana et al. (1994) who investigated K_s within flooded paddy fields. Their results show values ranging from 0.0125 mm h⁻¹ up to a maximum of approximately 0.208 mm h⁻¹ for poorly permeable puddled plow soles.

4.1.3. Field data related to the terrain

In order to identify possible linkages between slope positions and soil properties descriptive statistics such as mean and standard deviation were computed for the classes summit, backslope and valley, and box plots were created (Table 14 & Figure 11). Additionally, soil properties were correlated against slope gradient and elevation, and K_s with ancillary variables such as D_b , porosity and SOM (Table 15 & Table 16).

The correlation with the slope gradient did not yield any significant results. Whereas elevation showed a moderate ($r=0.45$) but significant ($P\leq 0.01$) correlation with clay content signifying the occurrence of higher clay content in higher altitudes (Table 15). This can also be observed in Table 14 and Figure 11, where summits represent with a mean of 23 % and a max of 42 % the slope position with the highest clay content. At the same time, sand content seems to increase in lower terrain as indicated by a moderate but significant correlation with $P\leq 0.05$ and $r=-0.3$ (Table 15) and the highest mean sand content being 43 % in valleys (Table 14).

Clay particles are difficult to detach, and their cohesion can be enhanced by organic matter rich soils (Gilley, 2005), like the once found on summits and backslopes (Table 14). Once clay particles are detached, they can be transported easily by overland flow. On the contrary, sand particles generally lack cohesiveness and are thus detached easier. Because of their size, sand particles require more energy to be transported (Gilley, 2005), hence, their greater occurrence might be related with an increase in flow accumulation in lower altitudes of the watershed. Furthermore, it can be assumed that eroded silt and clay particles were already transported out of the watershed and deposited in downstream areas. Apart from that, the occurrence of sand, silt and clay varies not much among the different positions in the upper most layer. Slight differences are detectable but not as much as that they could be linked to erosional and dispositional processes as described by Malo et al. (1974).

Table 14. Soil properties per slope unit.

Position		Variable							n
		Clay (%)	Silt (%)	Sand (%)	K_s (mm/h)	Por (%)	D_b (g/cm ³)	SOM (%)	
SU	Mean	23	41	36	570	58	1.1	4.7	6
	SD	13.8	10.4	20.6	599.8	7.6	0.2	1.3	
BS	Mean	20	43	37	434	56	1.2	4.9	24
	SD	11.9	6.1	14.2	490.5	7.2	0.2	1.6	
VA	Mean	18	39	43	398	51	1.3	4.0	18
	SD	10.0	7.0	11.5	457.1	7.2	0.2	1.2	

VA = Valley, BS = Backslope, SU = Summit, K_s = Saturated Hydraulic Conductivity, Por = Porosity, D_b = Bulk Density, SOM = Soil Organic Matter, SD = Standard Deviation, n = number of samples

By correlating K_s values with ancillary variables within each slope unit a significant relationship ($P\leq 0.01$; $r=-0.75$) with D_b was detected in the backslope class, as well as a strong dependency of K_s on SOM content on summits ($r=0.81$) (Table 16). The latter can be traced back to the fact that >83 % of the samples collected within the summit class were located within the land cover mixed forest and orchard. SOM's dominating impact on soil physical properties in those classes was discussed in the above section.

In terms of D_b a slight increase of the mean and max values is inherent starting in the catena from the summit (mean = 1.1 g cm^{-3} ; max = 1.4 g cm^{-3}) via the backslope (mean = 1.2 g cm^{-3} ; max = 1.6 g cm^{-3}) to the valley (mean = 1.3 g cm^{-3} ; max = 1.7 g cm^{-3}) (Table 14 & Figure 11). This can be attributed to a transition of land covers along the catena and to compaction in agricultural areas in the valleys.

Table 15. Correlation of terrain derivatives with soil properties.

Variable	Elevation	Slope
$D_b \text{ (g/cm}^3\text{)}$	n.s.	n.s.
Clay (%)	0.45**	n.s.
Silt (%)	n.s.	n.s.
Sand (%)	-0.3*	n.s.
SOM (%)	n.s.	n.s.

*/** = statistically significant at $P \leq 0.05/P \leq 0.01$ level, n.s. = not significant, D_b = Bulk Density, SOM = Soil Organic Matter

Table 16. Correlation of K_s per slope unit with soil properties.

Variable	Summit	Backslope	Valley
$D_b \text{ (g/cm}^3\text{)}$	n.s.	-0.75**	n.s.
Clay (%)	n.s.	n.s.	n.s.
Silt (%)	n.s.	n.s.	n.s.
Sand (%)	n.s.	n.s.	n.s.
SOM (%)	0.81	n.s.	n.s.

** = statistically significant at $P \leq 0.01$ level, n.s. = not significant, D_b = Bulk Density, SOM = Soil Organic Matter, K_s = Saturated Hydraulic Conductivity

In summary, a more dominant impact of the different slope positions on the soil PSD and therefore, on the soil physical properties, was expected. It is evident that the soil properties are not varying too much along the catena. Especially the soil properties of the summit and backslope class are alike. When comparing soil properties on slope units (Table 14 & Figure 11) with soil properties in different land cover types (Table 11, Table 13 & Figure 10) it seems as if values become normalised by aggregation, and specifications attributable to land covers get lost. The suspicion that land cover and land use, as well as biotic factors, are the prevailing determinants of the soil properties in the upper soil layer as discussed earlier seems to be confirmed. This result coincides with the findings of Hassler (2013), Hassler et al. (2011) and Sobieraj et al. (2002) who discovered comparable conditions in tropical soils along slope transects.

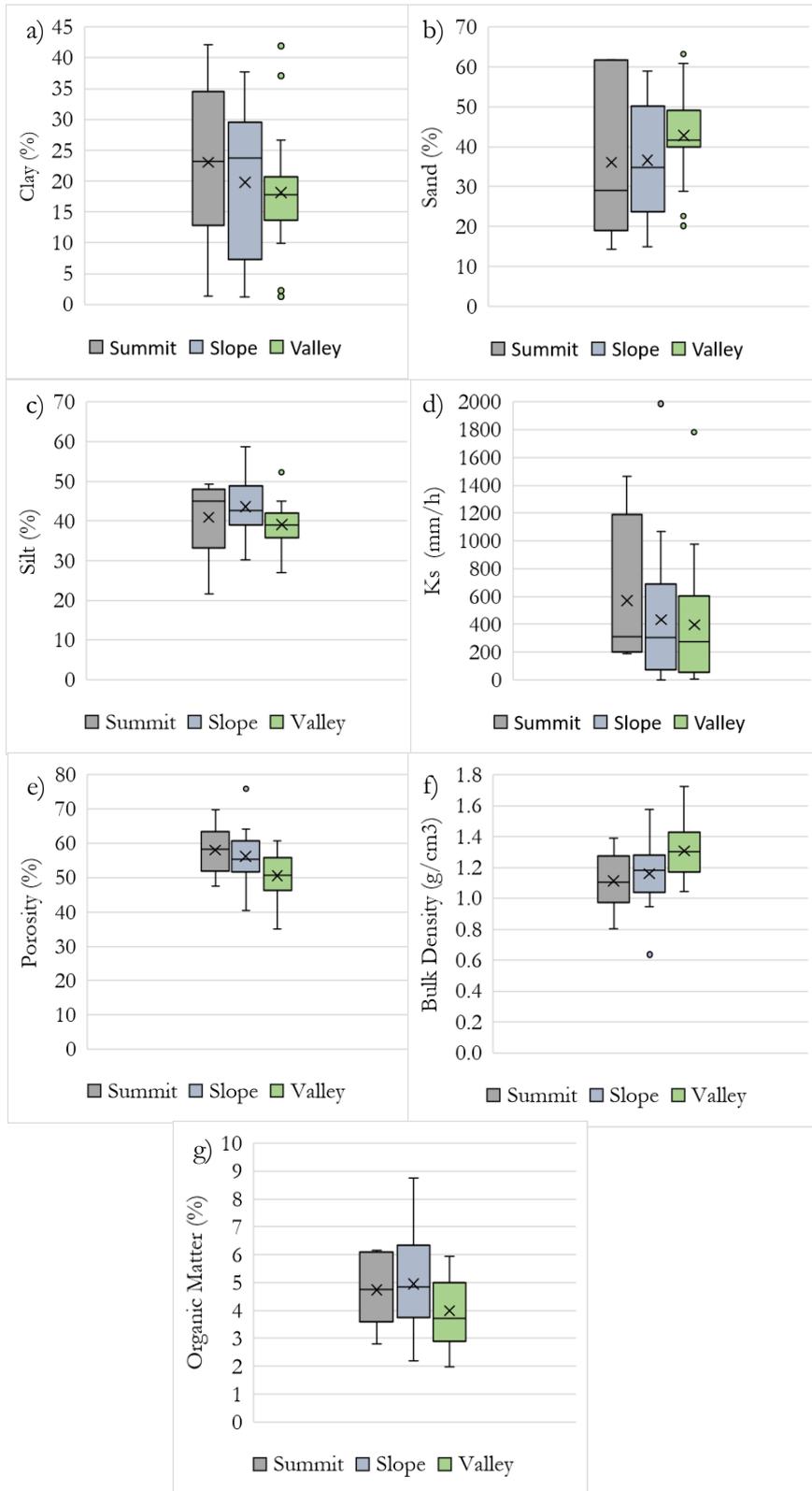


Figure 11. Variability of soil properties per slope unit; **a)** clay, **b)** sand, **c)** silt, **d)** Saturated Hydraulic Conductivity, **e)** Porosity, **f)** Bulk Density, **g)** Organic Matter Content.

4.1.4. SoilGrids variability in the landscape

Descriptive statistics for all soil properties were calculated, for the ones originally extracted from the SG layers and for the ones which were created by the PTFs of Saxton et al. (2006), such as K_s and porosity. The data can be found in Annex 2.5. Statistics were computed for each type of land cover and the three slope positions (Table 17 & Table 18). It is evident that spatial variability is not represented by the SG dataset, neither in different land cover classes nor on the slope positions. The largest standard deviation was computed for K_s with 8.9 mm h^{-1} and the lowest can be found for D_b with 0 g cm^{-3} and porosity with 0% , which indicates a very low or not existing variability (Table 17). The same applies to the PSD, the percentage of sand, silt and clay is almost identical in all of the three slope position (Table 18). This contradicts the investigations of Malo et al. (1974) who observed different processes taking place along a catena, leading spatially to different PSDs, causing differences in soil properties.

Significant are the continuously high D_b values being lowest in the mixed forest on the summit with 1.3 g cm^{-3} and greatest in almost all land covers with 1.6 g cm^{-3} (Table 17 & Table 18). Those high values seem to be generic and do not take into account the impact of land cover types and land use practices. Thus, they do not represent the mostly porous and loose surface soil layer caused by abandoned root channels, microorganism activity and other biotic factors as observed during the fieldwork. Further, they contradict the findings of Hassler (2013) who reported for tropical forest soils D_b of 0.8 g cm^{-3} and correspond more to the findings of Muñoz et al. (2007) who reported D_b of greater 1.5 g cm^{-3} for very degraded soils in Chile.

The mean SOM contents of 5-6.5 % having its highest values with 7 % and 8 % in the mixed forest and orchard class respectively appears valid for a tropical area although a lack of spatial variability is recognised (Table 17). This is identifiable by a mean SOM content of 6 % in agricultural areas, however, ignoring the process of SOM declination by deforestation activities and continuous cultivation as it is described by Ross (1993) (Table 17). SG makes use of different soil covariates to predict soil properties (Hengl et al., 2017). One of these covariates is the land cover product GlobalCover30 of the year 2010 by Chen et al. (2015). This could be another reason for missing spatial variability of SOM content, as this product is clearly outdated, and land cover changes are therefore not taken into consideration.

It is evident that the PTFs by Saxton et al. (2006) considers SOM as soil structure enhancing factor. Therefore, an increase in SOM will always result in greater porosity and K_s and a decline in D_b . Because different soils and different types of SOM might have different intercorrelation, this feature is questionable. Where one type of SOM influences soil aggregation and associated pore space distribution (Hudson, 1994) another one would cause clogging of available pores and thereby reduces the flow of water (Nemes et al., 2005). Signs of this behaviour were, for example, visible during the assessment of the FD where K_s exhibited a strong positive correlation with SOM in the mixed forest. In contrast, no similar correlation was found within the orchard class.

K_s values produced by SG and the PTFs are largely controlled by the PSD, which is visible in Table 17. Increasing clay by a few percent will result instantly in a decrease of K_s . Therefore, this combination might potentially achieve good results when applied for estimating soil-water characteristics of deeper soil layers where the influence of surface processes is not dominant, and texture takes over control.

Using the PTFs of Saxton et al. (2006) to predict soil properties such as porosity and K_s is convenient but should be used with caution. The PTFs are based on the original PTFs introduced by Saxton et al. (1986) and were considered as the best performing PTFs during an extensive review conducted by Gijssman et al.

(2003). However, the output of a PTF is largely depended on the soil database used for its definition, and thus, its predictions reflect the interactions of its input soils (Nemes et al., 2005).

Table 17. Statistics of SoilGrids soil properties per land cover.

Land Cover		Variable							n
		Clay (%)	Silt (%)	Sand (%)	K _s (mm/h)	Por (%)	D _b (g/cm ³)	SOM (%)	
Orchard	Mean	26	32	42	36	59	1.5	6	34
	SD	1.7	1.4	1.8	7.0	2.4	0.1	1	
	Min	22	29	37	17	53	1.4	4	
	Max	30	35	46	57	65	1.6	8	
Cropland	Mean	27	32	41	34	59	1.5	6	4
	SD	3.0	1.2	1.9	4.5	0.9	0.1	0.4	
	Min	23	32	39	31	58	1.4	5	
	Max	29	34	43	41	59	1.6	6	
Teak Plantation	Mean	30	32	38	22	56	1.5	5	2
	SD	0.7	1.4	0	0.2	0	0	0.2	
	Min	29	31	38	22	56	1.5	5	
	Max	30	33	38	22	56	1.5	5	
Mixed Forest	Mean	28	32	40	29	57	1.5	5	8
	SD	2.6	1.3	1.5	7.0	2.0	0.1	0.9	
	Min	24	31	38	23	55	1.3	4.5	
	Max	30	35	41	41	61	1.6	7	

K_s = Saturated Hydraulic Conductivity, Por = Porosity, D_b = Bulk Density, SOM = Soil Organic Matter

Table 18. Statistics of SoilGrids soil properties per slope unit.

Position		Variable							n
		Clay (%)	Silt (%)	Sand (%)	K _s (mm/h)	Por (%)	D _b (g/cm ³)	SOM (%)	
Summit	Mean	27	32	41	37	60	1.5	6.5	6
	SD	2.0	1.2	1.0	4.7	1.7	0.1	0.9	
	Min	24	31	39	32	57	1.3	4.8	
	Max	29	34	42	41	61	1.6	7.2	
Backslope	Mean	27	32	41	33	58	1.5	5.8	24
	SD	2.0	1.3	2.2	8.9	2.8	0.1	1.2	
	Min	24	29	37	17	53	1.4	3.6	
	Max	30	35	46	57	65	1.6	8.3	
Valley	Mean	26	33	41	34	58	1.5	5.8	18
	SD	2.2	1.5	1.8	6.0	2.0	0.1	0.9	
	Min	22	30	38	22	56	1.4	4.5	
	Max	30	35	44	42	62	1.6	7.4	

K_s = Saturated Hydraulic Conductivity, Por = Porosity, D_b = Bulk Density, SOM = Soil Organic Matter

4.1.5. Similarity analysis of SoilGrids and field data

In order to quantify the similarity of SG data and FD, the Cosine Similarity was used as similarity measure. To make a statement if FD and SG are similar from a statistically perspective, the Wilcoxon Signed-Rank test was conducted.

Since SG purest form is the PSD, the first assessment was done just comparing the percent of sand, silt and clay of the two datasets. In this manner, it can be prevented to include an uncertainty caused by other predicted soil physical properties through the use of PTFs. The Cosine Similarity for the summed PSD values scored a similarity of 0.98, indicating a high level of similarity. In the next step, the same procedure was done but considering different slope positions. The highest results were found in the valley and summit class with a cosine of 0.99 and the lowest in the backslope class with 0.97 (Table 19). However, all of them show high similarity within the fractions of sand, silt and clay.

In the next step, the other soil properties (normalised) such as K_s , D_b and SOM were compared. This time based on the different land cover types as it was found that land cover has a great impact on these properties within the research area (Section 4.1.2). Table 20 shows that high similarity was yielded in each land cover class, being lowest in the cropland class with 0.68. Additionally, different combinations of soil properties were tested with the Cosine Similarity for each site. The results can be found in Annex 6. Overall, almost always a high level of similarity was found. Each combination gave an average similarity of 0.77. The lowest result for a single site was 0.42, which can still be considered as moderate similar (Annex 7).

Table 19. Cosine Similarity per position.

Position	Cosine Similarity
Summit	0.99
Backslope	0.97
Valley	0.99

Computation was done based on the particle size distribution

Table 20. Cosine Similarity per land cover.

Land Cover	Cosine Similarity
Orchard	0.94
Cropland	0.68
Teak Plantation	0.97
Mixed Forest	0.99

Computation was done based on the soil properties: K_s , D_b and SOM

Achieving high similarity throughout the datasets seems to be odd considering the different behaviours of the datasets as discovered in the previous sections. On the one hand, having the FD which is highly variable and on the other hand having SG showing almost no variability in the landscape (Section 4.1.2 & 4.1.4). Hence, an obvious assumption is that the Cosine Similarity does not represent an appropriate similarity measure when comparing soil data due to a lack of sensitivity. Differences in properties must be very high to have a significant effect.

Before the Wilcoxon Signed-Rank test could be computed, both datasets had to be assessed in terms of normality by deploying the Shapiro-Wilk test, kurtosis and skewness. A normal distribution is given if the z-values (at $P > 0.05$) of skewness and kurtosis fall within the span of -1.96 to +1.96 and if the Shapiro-Wilk P-Value is above 0.05 (Field, 2009). Table 21 shows the results of the normality test, indicating that not all property pairs were normally distributed. Examples for that are the K_s values (FD) and the silt and SOM values (SG) with a Shapiro Wilk P-Value of 0, 0.001 and 0.037, respectively (Table 21). As not all properties were normally distributed the Wilcoxon Signed-Rank test was seen as an appropriate choice.

Table 21. Results of the test of normality.

Dataset	Property	Skew Z-Value	Kurtosis Z-Value	Shapiro Wilk P-Value
Field Data	Sand	0.43	-1.063	0.085
	Silt	-0.321	0.631	0.416
	Clay	0.026	-0.718	0.203
	D _b	0.005	0.91	0.767
	SOM	0.408	-0.132	0.33
	K _s	4.817	27.45	0.000
SoilGrids	Sand	-0.041	0.059	0.253
	Silt	-0.46	-0.553	0.001
	Clay	0.221	-0.615	0.101
	D _b	-0.05	0.193	0.744
	SOM	-0.08	-0.867	0.037
	K _s	0.154	0.855	0.151

K_s = Saturated Hydraulic Conductivity, D_b = Bulk Density, SOM = Soil Organic Matter

Table 22 shows the results of the Wilcoxon Signed-Rank test. Sand was the only physical property which yielded a P>0.05, declaring sand as the only property having statistical similarity when comparing FD and SG. The other properties had P-Values consistently below the critical value of 0.05. Therefore, it can be concluded that there is a significant difference inherent.

Positive and negative ranks indicate the direction of reported difference (Table 22). Hence, the majority of negative ranks within D_b, SOM and Clay indicate that the FD has significantly lower values compared to SG, suggesting an overestimation by SG. In contrast, when inspecting K_s and Silt, it is evident that the FD is significantly higher, indicating an underestimation of the properties by SG.

These results imply that SG data should be used carefully. Especially when used for hydrological modelling on a watershed scale where differences in water holding capacity or infiltration are decisive. For instance, clay content and D_b are clearly overestimated by SG (Table 22). This, in turn, will greatly affect other soil properties such as K_s. Having predominately an underestimation of K_s values predicted with the PTFs (Table 22) the model will generate surface runoff when rainfall intensities exceed the infiltration rate although infiltration would take place in reality. Similarly, when looking at SOM, an overestimation could lead to a wrong water holding capacity of the soil. Thus, rainfall intensities below K_s might infiltrate in the soil whereas a soil with less SOM could already be saturated and hence surface runoff would occur much earlier.

Table 22. Output of Wilcoxon Signed-Rank test on SoilGrids derived properties.

Property	Standard Error	Z-Score	P-Value	Positive Ranks (n)	Negative Ranks (n)
Sand	97.5	-1.1	0.272	21	27
Silt	97.5	5.44	0.000	43	5
Clay	97.5	-3.3	0.001	14	34
D _b	97.5	-5.7	0.000	4	44
SOM	97.5	-4.3	0.000	13	35
K _s	97.5	5.6	0.000	41	7

K_s = Saturated Hydraulic Conductivity, D_b = Bulk Density, SOM = Soil Organic Matter

4.1.6. Summary of soil analysis

Spatial variability and the effects of land cover were investigated for (I) FD and (II) SG. Subsequently, the similarity of both datasets was assessed using Cosine Similarity and the Wilcoxon Signed-Rank test. The result of the independent assessment of the FD revealed that the surface soil layer in the watershed and its physical and chemical properties is highly affected by the respective land cover class and land use activities. Those effects are so pronounced that even clear patterns between different types of fruit trees in the orchard class could be established. Furthermore, SHPs such as K_s and porosity seem to be mainly determined by the presence of intensive bioturbation. Soil texture, on the other hand, plays a subordinated role. In contrast to that, the influence of the terrain and the different processes taking place is neglectable, as soil properties do not vary much along the catena. The assessment of SG data revealed its almost non existing spatial variability within the watershed. Neither effects of different land cover types nor of changing slope positions are identifiable. Furthermore, PTFs should be used carefully as they reflect the interactions of soils used for their creation and do not necessary represent features which are location specific.

The similarity analysis of the two soil datasets gave contradicting results. Cosine Similarity, on the one hand, yielded a high similarity of both FD and SG, regarding soil properties within the three slope positions, and different land covers. The Wilcoxon Signed-Rank test, on the other hand, declared sand as the only property having a statistical similarity. Since obvious differences between both datasets were observed during the independent assessment, Cosine Similarity is expected not to be a suitable similarity measure for soil properties. The Wilcoxon Signed-Rank test furthermore indicated that fundamental soil properties such as percentage clay and SOM, as well as D_b , appeared to be underestimated by SG compared to field data which had a great effect on SHPs predicted by the PTFs and potentially far-reaching consequences for the subsequent hydrological modelling.

Based on the obtained findings, several conclusions for the model parametrization can be drawn. Firstly, since the properties of the surface soil layer are determined by their respective land cover, soil property maps for openLISEM should be rather based on the land cover map than on the slope position map. Secondly, it can be expected that intensive bioturbation just takes place in the most upper soil layer and therefore that SHPs of soil layers below are mainly determined by texture. Hence, the introduction of a second soil layer with SHPs predicted with PTFs seems an obvious choice.

4.2. Land cover analysis

In this subsection, the results of the land cover classification and land cover change analysis are presented. The land cover of Ban Da Na Kham watershed was mapped for the years 2005 and 2018 and subsequently the appeared land cover changes identified. In the course of this section, the following research questions are addressed:

- Which land cover changes occurred in the study area between 2005 and 2018?
- What are the possible reasons behind these land cover changes?

The mapping activities were done in GEE employing RF as a classifier and using Landsat 5 and 8 imagery (Section 3.3). Main land cover classes identified during the fieldwork, which were also used for the classification, included orchard, cropland, teak plantation, mixed forest and built-up areas.

Orchards are represented by a mix of different perennial crop trees such as long kong, durian and banana. Long kong and banana are the predominant trees within that class and their plantation area extents from the valleys up to the slopes until the edges of the forest. It appears that a continuous expansion of the orchard plantations is taking place. Young fruit trees are planted between the natural trees in an agroforestry system leading to fuzzy boundaries of the two classes. At the same time, natural trees were noticed being carved to cut off the nutrient supply in order to ultimately let the trees die. Cropland is mostly situated in the wider valleys in the low-lying land, where the main streams are located thus simplifying irrigation. Flooded rice paddy fields hold the majority share in this class. Apart from that, corn and beans were cultivated. Interesting is the distribution of the teak plantations, being arranged in higher altitudes surrounding the watershed, areas characterized by clay-rich soils (Section 4.1.3). This might be related to the fact that soils with high clay content support the growth of teak trees as discussed by Pramono et al. (2015). Teak as a tropical hardwood tree species is grown for high-quality timber production and represents one of the livelihoods of the local population. The trees are planted in a line pattern to ease plantation maintenance (Pramono et al., 2015). Urban areas can be found in two different patterns and are located in the main valleys of the research area (Figure 12). Firstly, as isolated clusters, each featuring one village community, and second as buildings lining the roads mostly coupled with shops where local products such as harvested fruits and wooden goods are sold. In Table 23 the land cover area distribution among the different classes for both years are shown.

Table 23. Land cover area for 2005 and 2018.

Land cover	Area 2005 (km ²)	Area 2018 (km ²)
Orchard	34.71	39.80
Cropland	2.02	3.12
Teak Plantation	4.73	5.46
Mixed Forest	45.13	36.91
Urban	0.32	1.60

The accuracy assessment of the land cover map for 2018 yielded an overall accuracy of 89.2 % and a Kappa coefficient of 0.86 (Table 24). According to Rwanga et al. (2017), a Kappa of >0.81 signalizes a very strong agreement. The weakest producer accuracy of 66.9 % was found in the cropland class, where 6 out of 40 reference points were classified as orchards (Table 24). One reason for this misclassification could be an early stage orchard, represented by young trees with a not yet fully developed leaf canopy. In this case, most of the reflection of the pixel would originate from the soil surface rather than from the fruit tree itself. Another possible explanation could be a freshly cut banana plantation, leading equally to extensive soil background noise. This could easily be associated with a cropland area, which can be for example, brownish coloured when the paddy is ripe. Similar misclassifications were observed in the orchard class, with 8 reference points classified as cropland (Table 24).

Fuzzy boundaries between orchards, mixed forests and teak plantations are expected to be the main reason for wrongly classified pixels in the mixed forest and teak plantation class. Reference pixels of the teak plantation class, classified as cropland can be linked to faded teak blossom and surface cover of brownish coloured foliage, occurring during the dry season. Looking at the urban class, confusions appeared with orchards and cropland pixels. This might be attributable due to the close vicinity of the urban area to the orchards and croplands, thus potentially causing propagation of mixed pixels. Furthermore, similarities between the colours of rooftop materials and cropland areas are expected to partly cause such confusions.

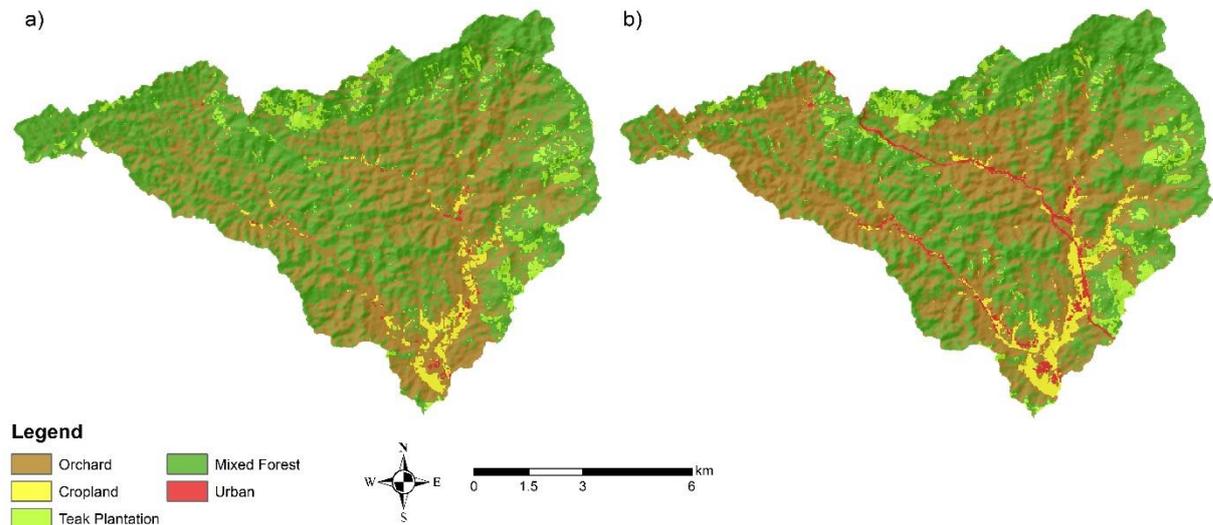


Figure 12. Land cover map Ban Da Na Kham watershed, **a)** 2005 and **b)** 2018.

For 2005 a larger area was classified in order to simultaneously create a land cover map for (I) Ban Da Na Kham watershed (Figure 12) and (II) Laplae watershed (Annex 8). The accuracy of the land cover maps for 2005 is 87.8 %, with a lower but still strong Kappa coefficient of 0.84 compared to 2018 (Table 25). Lowest producer accuracy (77 %) was obtained in the orchard class where confusions mainly happened with pixels of the mixed forest class (Table 25). This might be due to the less developed orchard plantations compared to 2018. Data for 2005 indicate that fruit trees were mostly still situated in the bottom part of the narrow tributary valleys, where they were easily covered by the leaf canopy of the mixed forest. In 2005, a water class had to be added into the classification procedure as a small lake, and several larger ponds were identified in Laplae watershed (Annex 8). In the water class confusions mostly appeared with urban areas and cropland. The former might have two reasons, firstly the proximity of roads and buildings to the lake and secondly blue coloured corrugated iron roofs from some of the buildings. Water, classified as cropland, might be caused by small ponds located on the agricultural areas or by flooded paddy fields where the rice plants not yet represent the majority of the observable surface cover. Remaining misclassifications are similar to the ones from the land cover map of 2018 and were discussed above.

Attempts for further improvement of the classifications could involve different actions. Various studies focusing on land cover mapping emphasized the exploitation of auxiliary information for classification accuracy improvements (Saah et al., 2019; Khatami et al., 2016; Zhu et al., 2016; Sluiter & Pebesma, 2010; Franklin & Wulder, 2002). In this study, only ancillary information extracted from elevation datasets and optical imagery were used (Section 3.3.4). Additional information for example distances to roads and buildings obtained from sources such as OSM (Saah et al., 2019), or the incorporation of texture layers, are additional popular ways to yield classification improvements (Coburn & Roberts, 2004). In order to forestall confusions between classes especially between classes characterised by trees, such as mixed forest, teak plantation and orchards, the inclusion of canopy cover and tree height as covariate layer could have also been supportive as shown by Saah et al. (2019). Canopy cover information could be retrieved from Landsat products using a regression and classification tree procedure (Hansen et al., 2011). A method for tree height estimation is presented by Hansen et al. (2016), who used besides of Landsat time-series multi-spectral data, GLAS (Geoscience Laser Altimeter System) also height data as input for their regression tree model.

Table 24. Accuracy assessment of land cover map 2018.

Classified Image	Reference Data					Row Total	User Accuracy (%)
	Cropland	Mixed Forest	Orchard	Teak Plantation	Urban		
Cropland	33		8	2	2	45	73.3
Mixed Forest		46	2	3		51	90.2
Orchard	6	2	60	2	4	74	81.1
Teak Plantation		1		59		60	98.3
Urban	1		1		83	85	97.6
Column Total	40	49	71	66	89	315	Total Accuracy 89.2 %
Producer Accuracy (%)	66.9	93.9	84.5	89.4	93.3		Kappa coefficient 0.86

Table 25. Accuracy assessment of the land cover map 2005.

Classified Image	Reference Data						Row Total	User Accuracy (%)
	Cropland	Mixed Forest	Orchard	Teak Plantation	Urban	Water		
Cropland	86				5	2	93	92.5
Mixed Forest	1	91	22	5	1	1	121	75.2
Orchard	5	3	77		4	1	90	85.5
Teak Plantation		2	1	35			38	92.1
Urban		1			87	2	90	96.6
Water						29	29	100
Column Total	92	97	100	40	97	35	461	Total Accuracy 87.8 %
Producer Accuracy (%)	93.5	93.8	77	87.5	89.7	82.9		Kappa coefficient 0.84

Additionally, it would have been interesting to investigate the influence of the independent variables on the prediction of each of the land cover classes in order to assess the relative importance independently. However, this would require the application of the random forest model to each class (class versus other), which could not be accomplished in this research study due to time constraints.

4.2.1. Land cover change analysis

Post-classification change detection comparison revealed that substantial changes in different land covers took place in Ban Da Na Kham watershed over the study period (Figure 13). Changes included urban developments, deforestation activities and expansion of cultivated areas. The relative changes of each land cover class within the period between 2005 and 2018 are depicted in Table 26. The greatest increase took place in the urban class with 394.31 % (1.28 km²), followed by cropland with 54.55 % (1.10 km²) and teak plantation with 15.55 % (0.74 km²), as well as in the orchard class with 14.68 % corresponding to 5.10 km². Mixed forest was the only category where a net decline occurred with 18.2 %.

As mentioned above the largest change in terms of area procent was found in the urban class (Table 26). This is partially due to urban development in terms of an increase in buildings on the expenses of orchards (0.77 km²), mixed forest (0.38 km²) and cropland (0.16 km²). Nevertheless, the greatest change can be attributed to the expansion of the main connecting road between Uttaradit province and the neighbouring province Phrae, crossing the research area from north to south. The road was extended from originally two to four lanes (Figure 14). Therefore, the constructed highway now contributes to the urban class in the classification since its width exceeds 30 m. Thus, is it recognisable on the Landsat images having a spatial resolution of 30 m. For the classification of 2005, it was not possible to map the existing smaller road with

this pixel size as its spectral reflectance got lost due to the reflectance of other land covers represented in the same pixel. Additionally, the road was often covered by the leaf canopy of trees lining the roadside, which led to mixed pixels (Figure 14). Hence, the recorded change in Table 26 was, in reality, much smaller than measured based on the land cover maps.

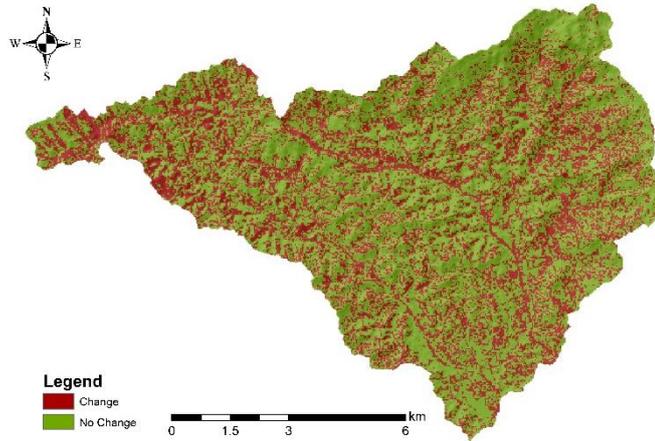


Figure 13. Change map of Ban Da Na Kham watershed.

Changes in cropland seem to have different reasons. As indicated in Table 26, the increase of cropland area is mainly due to the conversion of orchards (1.22 km²) and mixed forest (0.2 km²). Thereby, the change from orchards to cropland might be attributed to a rotational cropping system. As it is very unlikely that fruit trees such as long kong were cut down, it can be expected that those areas were previously occupied by banana plantations which got pared back. Furthermore, as outlined in section 3.3.9, the results of the change analysis depend on the classification accuracy. Hence, an exacerbated increase in the area could also be caused by class confusions, which indeed happened between the cropland and orchard class as described above.

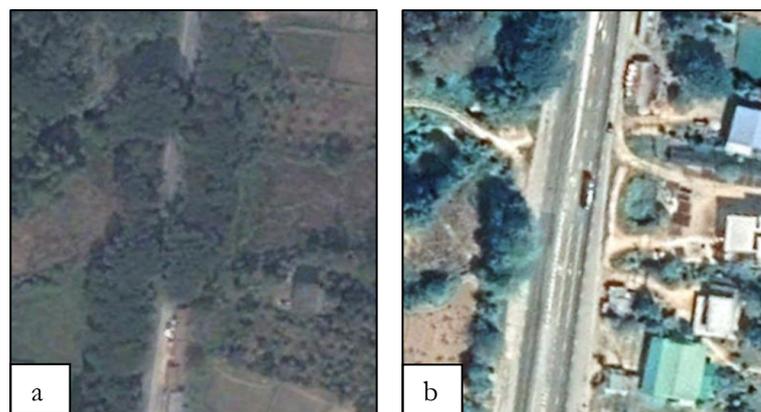


Figure 14. Main road in 2005 **a)** and 2018 **b)**; Source Google Earth Pro.

When comparing the land cover maps of the Ban Da Na Kham watershed of the years 2005 and 2018 a rivalry of the mixed forest and orchard class in terms of area coverage can be noticed (Figure 12). In 2005, mixed forest occupies most of the watershed with 51.9 %, whereas in 2018 orchards overtook mixed forest for being the most prominent land cover with 45.8 % (Table 26). Where the mixed forest partly spread out towards the valley bottoms in 2005, in 2018 the class is mostly forced back to the steeply sloping terrain and the summits while experiencing as only class a total area reduction of 8.21 km² (Figure 12). The land cover

change analysis reveals that the large decrease of mixed forest was caused by deforestation followed by conversion to orchards (15.48 km²) and teak plantations (2.27 km²) (Table 26). Similar observations were reported by Baicha (2016) who investigated land use dynamics in the Nan province which borders on the north-east of the Uttaradit province, which has a comparable terrain as the Ban Da Na Kham watershed. In the Nan province, the natural forest decreased by 3558.07 km² between 1995 until 2012 having the greatest reduction in the period between 2009 and 2012 with 2550.08 km², which corresponds partly to the period examined in this research. In the same period perennial crops and orchards experienced their largest expansion with together more than 1600 km² (Baicha, 2016). Those findings and the findings of the present study contradict with the Forestry Act from 1989 by the Government of Thailand enforcing a ban on commercial deforestation at the expenses of natural forest (FAO, 2009). Especially since the Forestry Act states that all areas with slopes of 35 % and above need to be preserved as forest in order to prevent surface runoff generation and erosion (FAO, 2009). Nevertheless, during the fieldwork for this study occupation of steeply sloping terrain by teak and fruit trees were observed.

Increasing by an area of 5.09 km² orchard was the land cover category increasing the most (Table 26). While new orchards were created, previous orchards became abandoned (Table 26). Similarly, areas previously cultivated with fruit trees got reconquered by nature. Thus, new mixed forests emerged (Table 26). The rationale behind such an interchange could be a hydrometeorological disaster like the one in 2006, where vast amounts of landslides destructed the landscape, making areas inaccessible for farming (Boonyanuphap, 2013; Usamah et al., 2013).

Two potential reasons explaining the increase in teak plantations were identified. Firstly, the conversion from orchards (0.21 km²) and cropland (0.01 km²) might be related to environmental policies which were enforced to reduce agricultural pressure on upland soils. According to Forsyth (2007) past reforestation activities (teak and pine) conducted by both governmental authorities and non-governmental organizations were encouraged with the aim to counteract runoff generation and erosional processes by protecting the soil surface. Furthermore, these plantations were supposed to decrease the need for farming activities as new livelihood opportunities in the form of plantation maintenance were created (Forsyth, 2007). Secondly, the increase of teak plantation on the expenses of natural forest (2.27 km²) might be attributable to illegal logging activities and subsequent reforestation.

Table 26 Land cover change matrix of Ban Da Na Kham watershed (2005 to 2018).

Year	Land Cover	2018 (km ²)					Total	(%)
		Orchard	Cropland	Teak Plantation	Mixed Forest	Urban		
2005 (km ²)	Orchard	23.72	1.22	0.21	8.78	0.77	34.71	39.9
	Cropland	0.24	1.61	0.01	0.00	0.16	2.02	2.3
	Teak Plantation	0.33	0.07	2.98	1.33	0.02	4.73	5.4
	Mixed Forest	15.48	0.20	2.27	26.79	0.38	45.13	51.9
	Urban	0.03	0.02	0.00	0.00	0.28	0.32	0.4
	Total	39.80	3.12	5.46	36.91	1.60	86.91	
	(%)	45.8	3.6	6.3	42.5	1.8		
	Change (km ²)	5.09	1.10	0.74	-8.21	1.28		
	Change (%)	14.68	54.55	15.55	-18.20	394.31		

For conclusion the results of the land cover change analysis confirm the trend of converting natural habitats into agricultural land and urban areas which was also reported by other studies in different region of the world e.g. Baicha (2016), Barasa and Perera (2018), Lin and Wei (2008), Panahi et al. (2010) and Sajikumar et al. (2015).

4.3. Flash flood modelling

The focus of this subsection is to address the questions of how different soil information influence the flow dynamics of the flash flood event of the 22nd of May 2006. Therefore, openLISEM was run with (I) detailed FD, and (II) SG data. In the second part of this section, the influence of land cover changes in Ban Da Na Kham watershed on flash flood dynamics is investigated. Prior, the generated input data is briefly described. In this section, the following research questions will be addressed:

Objective 1: Comparative analysis of SG versus FD for flash flood modelling;

- Is model calibration based on historical flood marks from a nearby watershed possible?
- What are the quantitative differences of the model output using, (I) FD, and (II) SG in relation to flow dynamics?
- What is the sensitivity of the flood dynamics to different soil depths using, (I) FD, and (II) SG?

Objective 2: Analysis of the effects of land cover change on flash flood behaviour;

- Which land cover generates the highest average runoff?
- How do these land cover changes affect the flood dynamics?

4.3.1. Final model parametrisation

The input data to run openLISEM were generated in the data preparation phase (Section 3.4.1). In the following, the results of the rainfall analysis and the final soil layer parametrisation are outlined.

Rainfall

Based on Gumble analysis, an RP of 69 years was determined for the event on the 22nd of May 2006 (Annex 9). Since no reliable information about the exact duration of the rainstorm was available, the design storm was created based on the maximum rainfall depth of 263.7 mm which was recorded on the day of the event. This resulted in a storm lasting 250 minutes, having a peak intensity of 236 mm h⁻¹ (Figure 15). These intensities seem to be high, but no reference intensities recorded in this time scale could be found for validation. However, ADPC (2006) connected the event to a low-pressure system which had developed due to the cyclone Chan Chu, which devastated the Dan Nag province in Central Vietnam simultaneously. This process could be a possible explanation for such an extreme rainfall event.

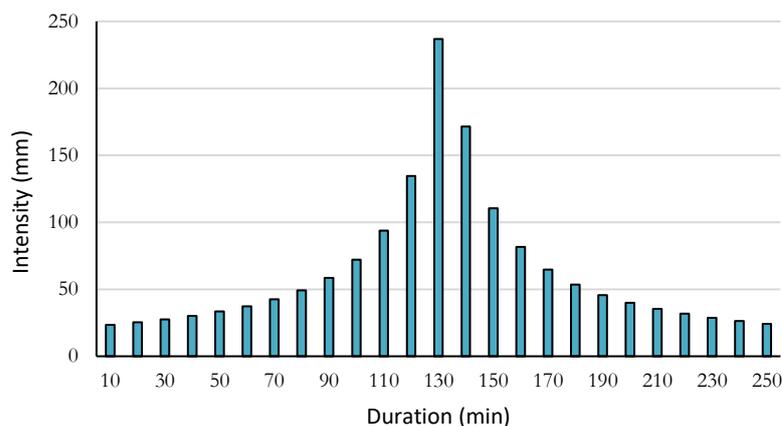


Figure 15. Design storm rainfall graph.

Soil data

The results of the soil analysis in Section 4.1 suggested that soil physical properties of the upper layer are rather influenced by the land cover than by slope position. Hence, the land cover map was used as a mapping base to generate soil property maps for openLISEM. The final parametrisation of SL1 for the FD setup can be found in Table 27. Analysing this data shows that mixed forest and teak plantations possess the highest rates of K_s ($>500 \text{ mm h}^{-1}$) and cropland and urban areas the lowest values ($<1 \text{ mm h}^{-1}$). K_s in urban areas was parametrized based on the assumption that those areas are characterised by surface sealing and compaction. This choice is supported by studies concerning urban hydrology by, for example, Gregory et al. (2006) and Ossola et al. (2015). Urban porosity values were generated based on the average porosity of the other classes. Results of SL2 are shown in Table 28 with a gradual increase in K_s going from the summits (5.6 mm h^{-1}) to the valleys (10.19 mm h^{-1}). The slightly higher porosity of $0.4 \text{ cm}^3 \text{ cm}^{-3}$ on summits can be linked to the average texture class being clay loam (Table 28). The Psi for the texture classes clay loam (Summit) and loam (Slope and valley) were calculated with 28.1 cm and 22.27 cm, respectively. For SG continuous soil property layers were generated. For SG SL1 K_s and porosity range from $81\text{-}15 \text{ mm h}^{-1}$ and $0.68\text{-}0.51 \text{ cm}^3 \text{ cm}^{-3}$, respectively. Whereas in SG SL2, K_s ranges from $13.4\text{-}0.15 \text{ mm h}^{-1}$ and porosity from $0.54\text{-}0.39 \text{ cm}^3 \text{ cm}^{-3}$.

Table 27. Soil hydraulic properties SL1.

Land Cover	K_s (mm/h)	Porosity (cm^3/cm^3)
Orchard	413	0.54
Cropland	0.208*	0.42
Teak Plantation	593.7	0.51
Mixed Forest	722	0.61
Urban	0.4	0.52

* Wijaya et al. (2009), K_s = Saturated Hydraulic Conductivity

Table 28. Soil hydraulic properties SL2 and texture class per slope position.

Slope Position	K_s (mm/h)	Porosity (cm^3/cm^3)	Texture Class
Summit	5.6	0.4	Clay Loam
Backslope	7.27	0.39	Loam
Valley	10.19	0.39	Loam

K_s = Saturated Hydraulic Conductivity

Soil depth:

Using the approach of Kuriakose et al. (2009), four soil depth maps were generated based on topographic factors. The depth scenarios are compiled as follows: D1 (min: 0.36 m; max: 2.04 m), D2 (min: 1.36 m; max: 3.04 m), D3 (min: 2.36 m; max: 4.04 m), D4 (min: 3.36 m; max: 5.04 m). In Figure 16, an example soil map (D1) is shown. Larger soil depths are given in the major valleys and on the foot of large hillslopes. Shallow soil can be found on steeply sloping terrain predominantly in the north-eastern corner of the study area.

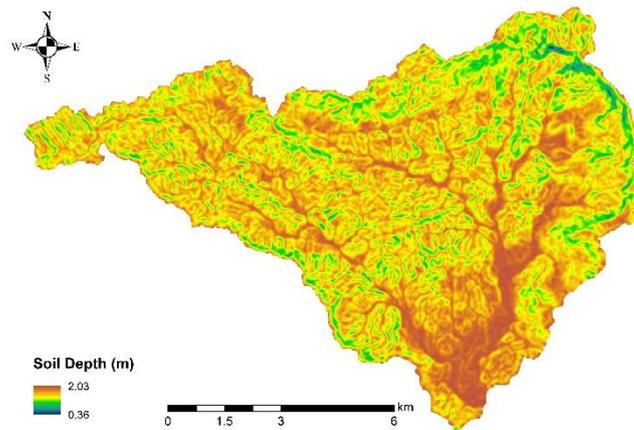


Figure 16. Soil depth map for scenario D1.

4.3.2. Calibration and the effects of different elevation models

Analyzing the research data revealed that a calibration based on historical flood height measurements was not achievable for various reasons. With none of the employed DEMs, it was possible to generate a flood pattern comparable to the measured flood heights in Laplae. All the DEMs utilized produced rather sobering results with unrealistic flood heights of more than 10 m up to 23 m in the mountainous areas (Figure 17). Simulations with DEM_v yielded the highest agreement in terms of points flooded with 44 %. DEM_m achieved 40 % and DEM_o 36 % (Annex 10). Simulations with the SRTM DEM had the lowest agreement with 24 % (Annex 10). However, all simulations showed consistently, that wide parts of the Laplae city remained dry, and areas which were previously flooded in 2006 received only a few centimetres of water. In other locations, an overestimation of several meters was simulated. Those mismatches are expected to be caused by the poor penetration of the C-band radar which applies besides of vegetation also for structures such as buildings. Hence, in the DEM, urban areas will exhibit elevations related to the top of the roofs rather than to the road surfaces. Therefore, flood waters must first overcome the approximate height of the buildings before other areas within the city can be inundated.

In general, the flood distribution pattern of DEM_o and DEM_m are alike. Using DEM_m drainage of flood waters is slightly better compared to DEM_o. However, both simulations still show large accumulations of flood waters in the upper mountainous valleys (Figure 17). In agricultural areas in the southern lowland, the flood spreads in a lake-like pattern having water depths of up to 10 m (Figure 17).

Using DEM_v and the SRTM DEM the flood distribution is controlled by small local depressions. However, simulations with the SRTM DEM indicate a much greater dispersion of the flood water, especially in the lowland (Figure 17). In the simulation with DEM_v northern mountainous areas show a flood pattern similar to DEM_o and DEM_m but with reduced intensity. In the lowland areas, floods are comparably small and with shallow water depths (0-1 m) (Figure 17). This might be due to improved drainage in the cropland area caused by the vegetation corrections.

Rice fields possess, in general, a flat ground. However, the modelling revealed various water accumulations within the cropland area leading to large ponds with water depths of several meters (Figure 18). It is assumed that isolated orchards, natural trees and bushes surrounding the paddy fields might affect the radar and therewith causing errors in the DEM. Hence, it can be concluded that this represents potentially the main reasons for these artificial ponds.

Another reason which possibly impeded the calibration attempts is the complexity of the 2006 flash flood event. Apart from the flooding, also debris flows and landslides occurred in the research areas (Boonyanuphap, 2013). Both are processes which can have a substantial effect on flood dynamics (Marchi, 2017). Therefore, openLISEM might not be the best choice to model such a complex situation. A solution could represent a multi-hazard model which is able to model hazard interactions. However, the drawback of such a model is the amount of detailed data required (Bout et al., 2018).

Overall, it appears that problems caused by incorrectly detected elevations caused by the top of vegetation and rooftops create a DEM which shows many small depressions that are likely not present in reality, and that have a great influence on the flood behaviour. These potential errors could not be solved by the corrections applied. This effect not only affects the local flood depth, but also the flow connectivity: water that is stored in upstream locations will not continue to flood the area that was flooded in reality. Hence, a realistic representation of the flood situation in 2006 was not possible with any of the DEMs. All DEMs result in unrealistic flood depths, in wrong locations and too shallow or too deep (more than 10 m). DEM_o and DEM_m simulate the most realistic flood pattern in terms of spread. In contrast, SRTM DEM performs worst.

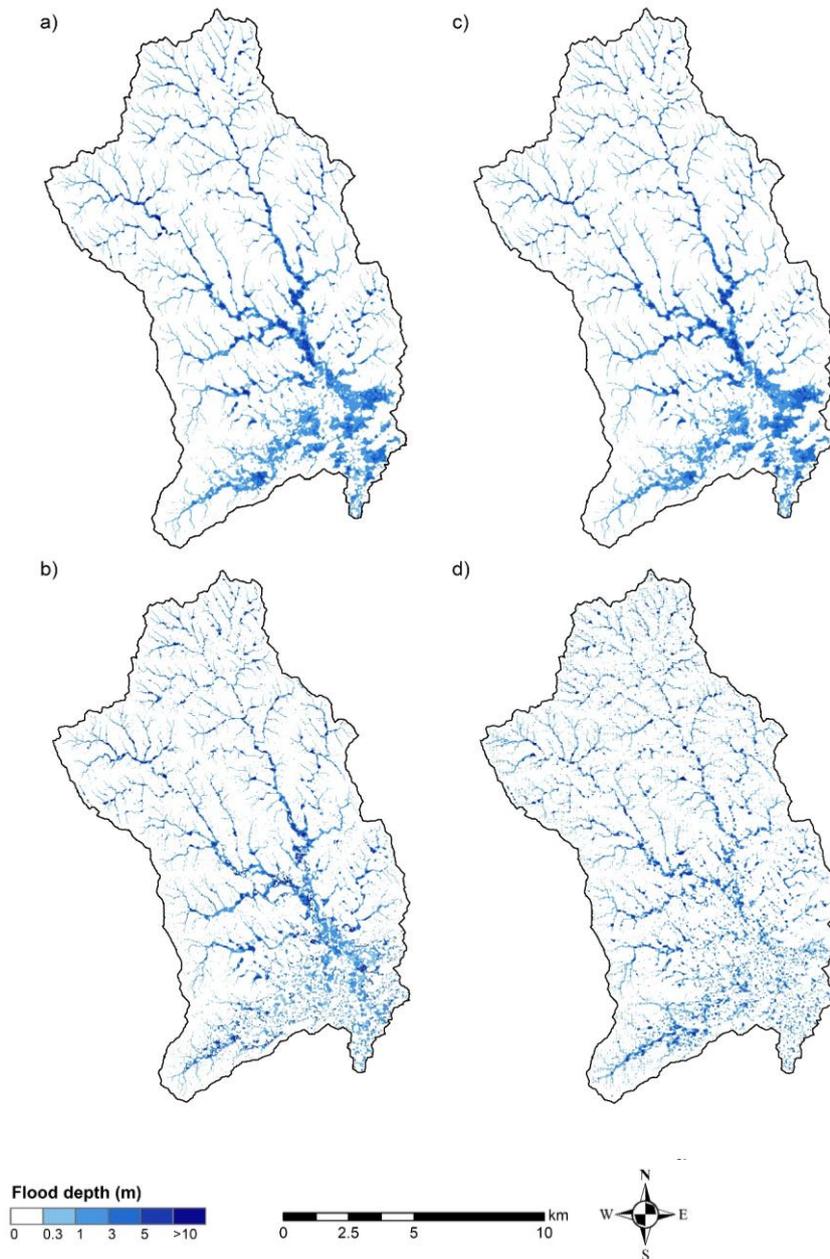


Figure 17. Flood simulation results in Laplae watershed using different DEMs: a) DEM_o; b) DEM_v; c) DEM_m and d) SRTM DEM.

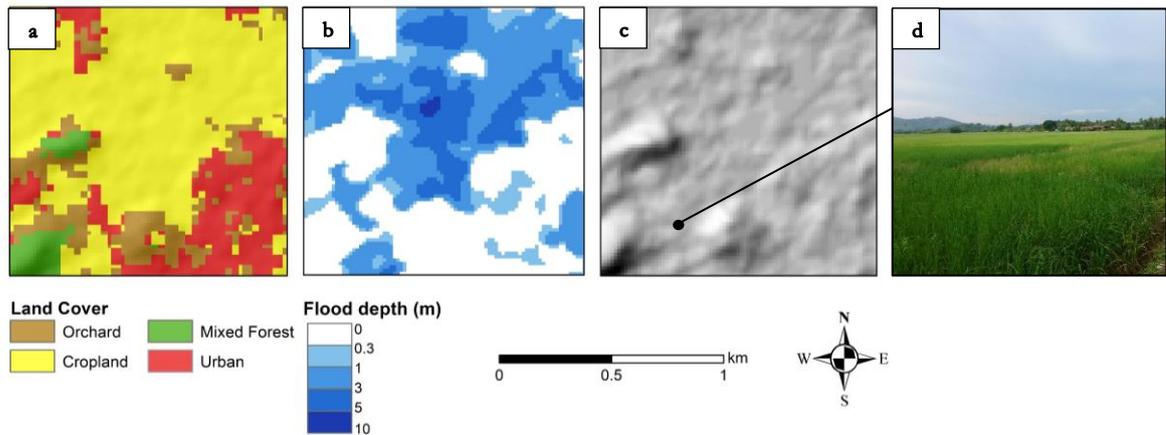


Figure 18. Example of DEM irregularities in cropland areas; **a)** section of land cover map, **b)** flood simulation with deep water ponds in parts of the cropland, **c)** hillshade showing undulating surface of cropland area, **d)** photo taken with view in north-east direction standing at black dot **c)**.

4.3.3. Sensitivity of flood dynamics to soil information

Eight model simulations were run to assess the sensitivity of flood dynamics in terms of (I) soil data source, and (II) soil depth. Additionally, the differences between SG and FD were calculated for each soil depth scenario. Employing SG as input much higher flood volumes were simulated for all three soil depths with 28.0 (D1), 27.2 (D2), 26.5 (D3) and 25.5 (D4) versus 24.0 (D1), 19.5 (D2), 14.2 (D3) and 9.0 million m³ (D4) (Table 29). This is caused by the infiltration behaviour, reflected by the amount of infiltration and overall runoff percentage generated. Runoff percentage was calculated as total outflow/total precipitation. Utilizing SG data, greater flood extents are simulated in comparison to FD with an increase of 8 %, 25 %, 55 % and 110 % for D1, D2, D3 and D4, respectively (Table 29). Using SG, the overall peak discharge is considerably higher for each soil depth scenario and the differences become more significant with each meter of additional soil (Table 29). However, in general, peak discharges are decreasing for both soil data sources with increasing soil depth. Using FD, the decrease is considerable larger with -62 % (D4 – D1) compared to SG with -27 % (D4 – D1) (Table 29). Similar behaviour is observable for the runoff percentage. In this case, the lowest runoff percentage simulated for SG in scenario D4 is still 6 % larger than the highest runoff percentage in FD-D1 with 48 % (Table 29).

Reasons for these observations are different mechanisms set into function for each soil data source. When using FD, the hydrological behaviour of the watershed is controlled by the storage capacity of the soil, including porosity and soil depth due to high infiltration rates of the porous SL1. Runoff is generated when SL1 reaches saturation, and rainfall intensities exceed the infiltration rate of SL2. In contrast, using SG, the hydrology of the watershed is predominantly determined by the infiltration rate and only to a limited extent by the soil storage. Correspondingly, during the rainfall, the rain intensity will already exceed the lower infiltration rates of SL1 and runoff is generated, even though the soil might still have empty pores.

Table 29. Water balance for eight model simulations using field data (FD) and SoilGrids (SG) data by using four soil depth scenarios; (SG-FD) represents the difference between SG and FD for each soil depth scenario.

Watershed Parameters	FD				SG				D1	D2	D3	D4
	D1	D2	D3	D4	D1	D2	D3	D4	(SG-FD)	(SG-FD)	(SG-FD)	(SG-FD)
Total Infiltration (mm)	62	102	142	183	15	25	35	45	-76%	-75%	-75%	-75%
Total outflow (mm)	127	90	61	38	172	162	153	143	35%	80%	151%	276%
Runoff (%)	48	34	23	14	65	61	58	54	35%	80%	151%	276%
Flood vol. (million (m ³))*	24	19.5	14.2	9	28	27.2	26.5	25.5	17%	39%	87%	183%
Flood area (km ²)*	11.8	10.1	8	5.8	12.8	12.6	12.4	12.2	8%	25%	55%	110%
Peak discharge (thousand (l/s))	214	154	104	81	343	312	264	249	60%	103%	154%	207%

*Values correspond to a flood threshold (flow height reported as flood) of 0.3 m, D1 = min: 0.36 m; max: 2.04 m, D2= min: 1.36 m; max: 3.04 m, D3= min: 2.36 m; max: 4.04 m, D4= min: 3.36 m; max: 5.04 m

The simulated flood distribution pattern when using FD and SG is similar and follows the terrain. Flooding primarily takes place along main channels and in the wide valleys (Figure 19). In lowland area in the south where cropland is the predominant land cover, flood depths range from 0.5 m – 5 m, whereas 1 – 3 m is the class occupying the largest area (Figure 19). Those comparable low flood depths can be linked to the wide geometry of the terrain enabling flood water spread, and to wider channels which ensure more discharge. In the northern part of the watershed in higher altitudes flood depths are increasing having mostly more than 3 m and often more than 10 m, due to reduced drainage in narrow valleys. In the tributaries, flood waters are shallower (0.3-1 m) as either infiltration or drainage to the main channels takes place (Figure 19). Apart from that, isolated very deep lakes with water depths of more than 10 m can be observed (Figure 19). Those water accumulations are formed by a lack of drainage connectivity, introduced by the poor quality of the DEM as discussed in section 4.3.2. Nevertheless, occasional natural sinks possibly capable of storing a few meters of storm water were observed during the fieldwork but not in the extent simulated.

Although the flood patterns are similar, simulations with SG lead to recognisable larger and deeper floods throughout the watershed (Figure 19). Table 30 presents the changes in area related to the flood depth classes with increasing soil depth for FD and SG. Obviously, the flooded area is shrinking with deeper soil independently from the soil data source (Table 30). Interesting, however, is which flood depth classes are affected most. For instance, using SG, areas with comparable shallow floods (0.3-1 m) experience the least total decrease with 0.5 % whereas in case of FD the same class is reduced by 45.3% (Table 30). In scenario FD-D4, the flooded area is very limited and mostly represented by shallow flood levels where formerly areas with large water depths could be found (Table 30). In contrast, simulations with SG show a minimal decrease at all flood levels. In scenario D4, there are 1.16 km² still occupied by water heights between 5 and 10 m and even 0.15 km² with >10 m (Table 30). This underlines the great sensitivity of the FD in regards to soil depth information, due to its ability to decide between a large-scale watershed-wide flood (D1) and small-scale localised floods when using D4 (Figure 19). SG continuously overestimates flood extent and depth. The effects of soil depth seem to be minor. Nevertheless, the sensitivity of SG to soil depth information should not be underestimated even it appears weak in comparison.

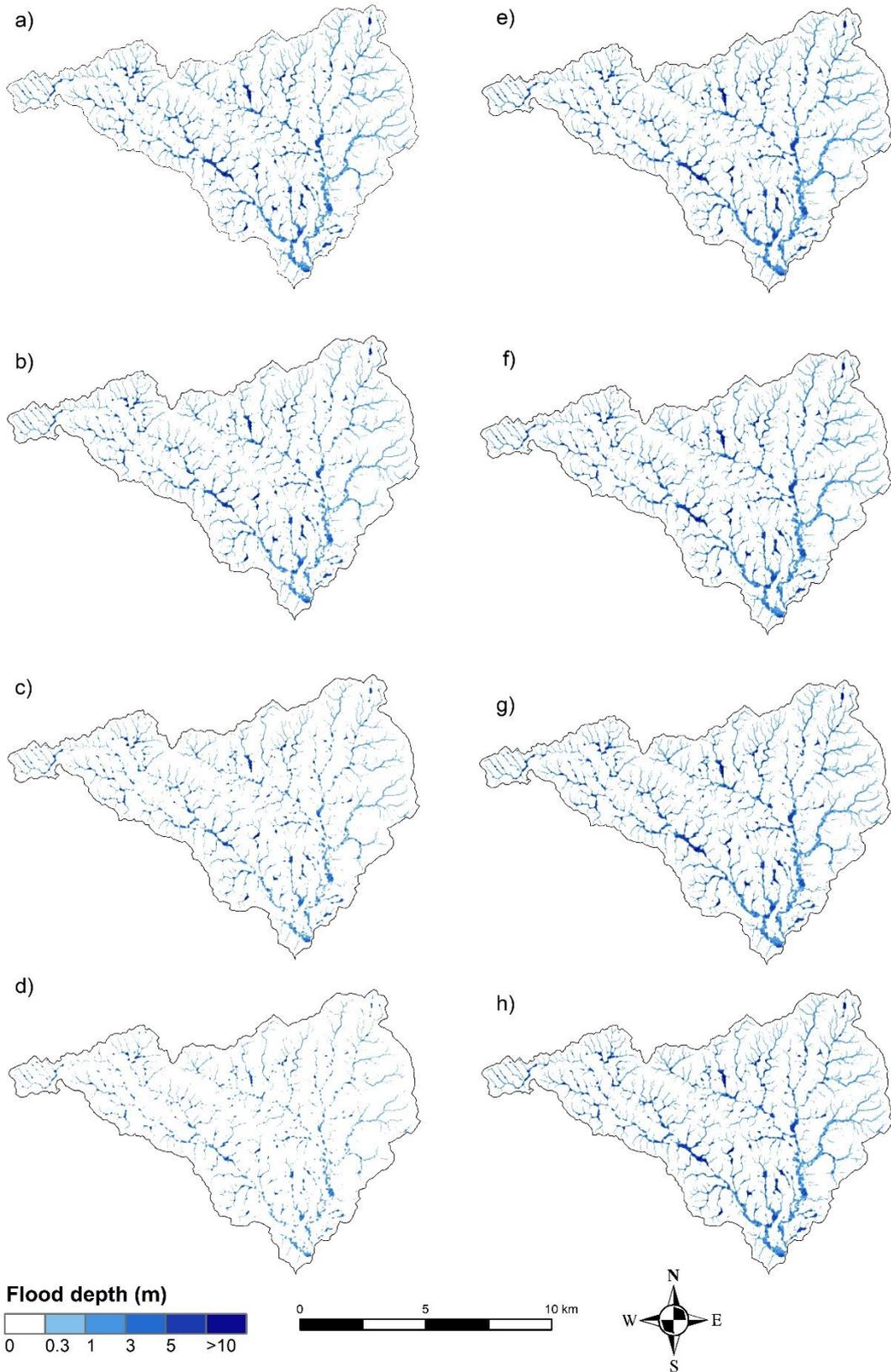


Figure 19. Maximum flood depth (m) for field data (FD) and SoilGrids (SG) for different soil depths: **a)** FD-D1, **b)** FD-D2, **c)** FD-D3, **d)** FD-D4, **e)** SG-D1, **f)** SG-D2, **g)** SG-D3 and **h)** SG-D4, D1= min: 0.36 m; max: 2.04 m, D2= min: 1.36 m; max: 3.04 m, D3= min: 2.36 m; max: 4.04 m, D4= min: 3.36 m; max: 5.04 m.

Table 30. Flood depth related to soil data information and total reduction from D1 to D4.

Flood depth (m)	FD (km ²)				Total (D4-D1)	SG (km ²)				Total (D4-D1)
	D1	D2	D3	D4		D1	D2	D3	D4	
0.3	5.4	4.8	3.9	3.0	-45.3%	5.52	5.53	5.51	5.49	-0.5%
1	3.8	3.2	2.6	2.0	-47.4%	4.2	4.1	4.0	3.9	-6.4%
3	1.4	1.2	0.95	0.6	-56.2%	1.7	1.61	1.55	1.52	-8.6%
5	1.1	0.8	0.5	0.2	-78.6%	1.3	1.19	1.21	1.16	-11.2%
>10	0.1	0.09	0.04	0.01	-96.0%	0.2	0.17	0.16	0.15	-14.0%

FD= Field Data, SG = SoilGrids D1= min: 0.36 m; max: 2.04 m, D2= min: 1.36 m; max: 3.04 m, D3= min: 2.36 m; max: 4.04 m, D4= min: 3.36 m; max: 5.04 m

Flood duration is influenced by the terrain. In areas, with high flood waters, the duration exceeds 10 h and often stays more than 60 h (Figure 20). Therefore, simulations with SG show prolonged flood durations compared to FD, caused by less and slower infiltration, more runoff and thus more water which needs to evacuate.

The dependency of flood duration related to the terrain is especially visible in mountainous pits where flood waters got locked and remained for more than 60 h (Figure 20). In wide cultivated valleys with flood depths between 0.5 – 5 m floods last long as well (Figure 19 & Figure 20). The low flow velocity is due to the almost flat terrain and therefore, slow drainage (Figure 20). This behaviour is intensified by openLISEM as channel inflow is controlled by surface flow velocity, which is often almost 0 m s⁻¹ in those areas. It appeared that the channels did not always transport water with their full capacity even if they passed deeply flooded areas.

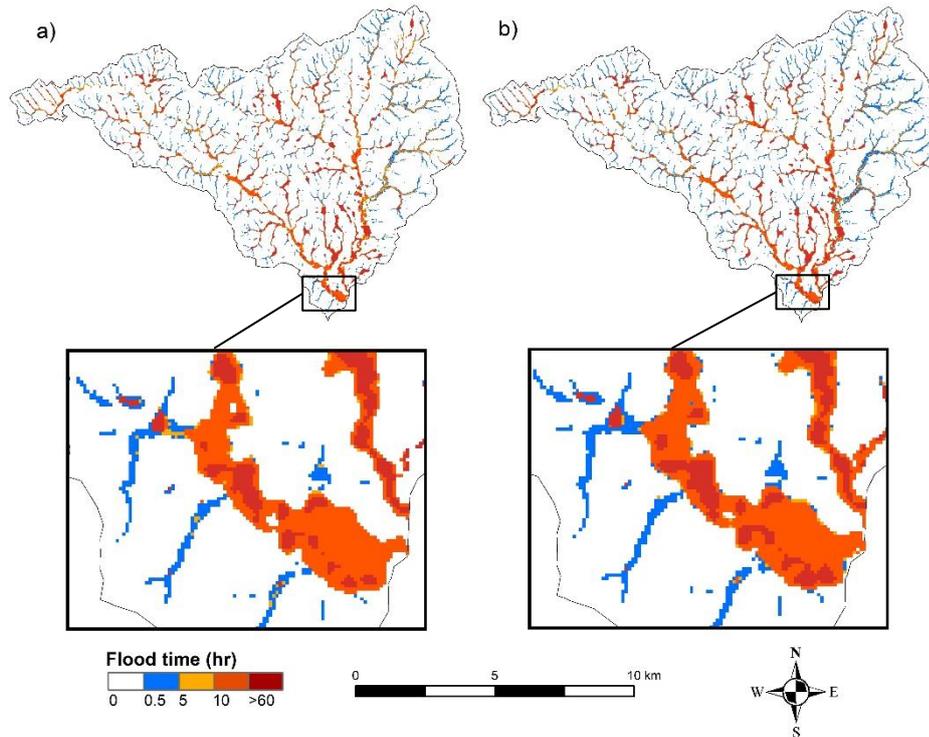


Figure 20. Maximum flood duration (hr) using field data **a)** and SoilGrids **b)** with soil depth scenario D1 (min: 0.36 m; max: 2.04 m). Zoomed area shows how flood time changes when using different soil information.

In contrast, short flood durations can be found in tributaries, in areas with steeply sloping terrain and shallow flood levels as both conditions promote quick drainage (Figure 19 & Figure 20). Nevertheless, also areas in the northern parts of the watershed, indicate long flood durations even in parts characterized by steep slopes which should counteract flood accumulations. These simulated hydrological behaviours can be attributed to the problematic DEM as surface water is erroneously slowed down and finally blocked in the middle of the slope.

Infiltration is increasing with each additional meter of soil added (Table 29). This is the case for both FD and SG. However, it is worth to note with each meter of soil added, the average infiltration of the FD is increased by approximately 40 mm whereas the average infiltration utilizing SG is always approximately 75% less for each soil depth scenario (Table 29).

Figure 21 shows infiltration maps for both soil datasets under D1 scenario. Using FD, it is observable that the infiltration pattern depends on the soil depth distribution. After a careful investigation, it appears that areas with high infiltration values (70 – 85 mm) are predominantly found in wider valleys and occasionally on summits (Figure 21). The former shows higher infiltration values for two reasons, (I) water from upslope areas is accumulated, and (II) soil is deeper. This leads inevitably to more infiltration capacity. However, in valleys also variations in infiltration can be observed, depending on their location in the terrain and the predominant land cover. In the main valleys where urban areas are situated, and crops are cultivated, the infiltration is low (0-40 mm) due to compaction of SL1. Correspondingly the available storage cannot be fully exploited (Figure 21).

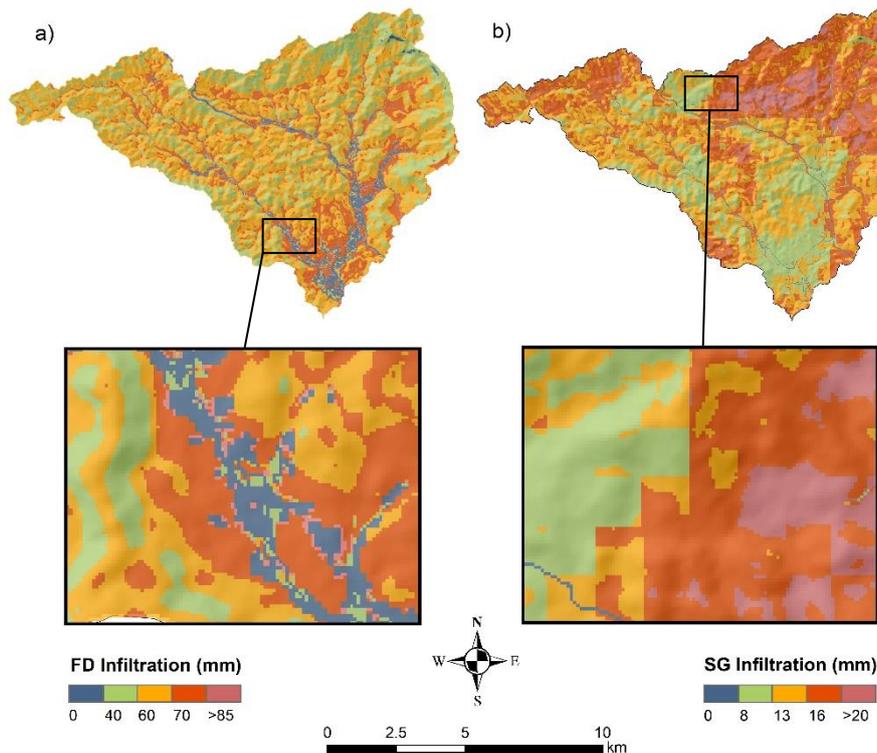


Figure 21. Total infiltration (mm) for field data **a**) and SoilGrids **b**) using soil depth scenario D1 = (min: 0.36 m; max: 2.04 m). Zoomed section of **a**) shows high infiltration values next to impermeable areas. Zoomed section of **b**) shows artificial border between low and high infiltration values.

Summits experience less infiltration since water is in this case only received by rainfall and not by runoff. Additionally soils are shallower. Variations of infiltration on summits vary between 60-85 mm caused by the difference in soil depth and to a small extent by changes in land cover (Figure 21). The rather low K_s of SL2 (5.6 mm h⁻¹) on summits seems to be counteracted by high flow resistance values of the teak plantations and mixed forest, thus reducing the flow velocity of surface water and promoting infiltration.

On slopes, more runoff is generated, and infiltration is decreased. This effect is caused by shallow soil depths, which limit infiltration and an increasing slope gradient, which accelerates runoff flow velocities. Additionally, the predominant cultivation in terms of fruit trees in sloping areas causes less infiltration due to its low flow resistance values. Consequently, besides of the limited infiltration capacity, the surface flow velocity is enhanced.

Nevertheless, almost all areas possess infiltration values > 40 mm (Figure 21). The porous upper soil layer with high K_s enables absorption of even high rainfall intensities, only limited by the available pore space. Large amounts of runoff generated on compacted surfaces (built-up and cropland) evacuate to adjacent pixels causing very high infiltration values of more than 85 and even up to 505 mm (Figure 21). On the one hand, this behaviour is in line with the hydrological principles; on the other hand, those values are too high and therefore might be caused by numerical errors within openLISEM.

Infiltration of SG is considerably lower compared to the results of the FD (Figure 21). The infiltration pattern is clearly determined by the distribution of SGs SOM and D_b in the research area (Annex 11). Furthermore, high infiltrations (>20 mm) can be found predominantly in the north-eastern part of the watershed, in the higher altitudes at the foot of the great northern slopes, defining the upper watershed border (Figure 21). Those areas are spots of runoff accumulation being characterised by low slope gradient, which favours infiltration. Interestingly a small spatial variability of SG is visible. This is shown as areas with low infiltrations (8 -13 mm) correspond very roughly the distribution of built-up areas and cropland (Figure 21). Both are characterised by high D_b and therefore, lower K_s and reduced porosity (Annex 11).

Analysing the “sub-patterns” reveals, that these are comparable to the ones produced by the FD. Thus, it can be argued that SGs infiltration seems additionally influenced by slope gradient and associated soil depth. Therefore, it is expected that the same mechanisms as described above, apply. In addition, in the northern part of the watershed, several sharp boundaries between areas with high and low infiltration are visible. Those infiltration patterns do not follow any logic as they do not correspond to any land cover boundary, nor to the soil depth distribution, and also not to the slope gradient (Figure 21). This pattern seems to be artificial as such boundaries are not expected to be found in natural landscapes where soil and its properties change more gradually. The same line pattern can be found on SGs SOM layer and is most likely attributable to a misprediction (Annex 11).

4.3.4. Effects of land cover change on flood dynamics

Table 31 shows the results of two model simulations for Ban Da Na Kham watershed, which were run in order to assess the effects of land cover changes on runoff and flood dynamics. In the period between 2005 to 2018, there was an expansion of predominantly orchards on the expenses of mixed forest. Further smaller changes comprised an increase in urban and cropland areas, as well as an increase in teak plantations (Section 4.2.1). The results contain interesting insights, for example, flood volume and inundated area decreased from 24.3 million m³ and 12.1 km² to 24 million m³ and 11.8 km² from 2005 to 2018 (Table 31).

Key figures such as total infiltration and runoff percentage experienced neglectable or no changes. Besides, it appears that the peak discharge increased by 1.9 %, and the peak time of the discharge decreased from 348 to 315 min (Table 31). Thus, those results indicate an accelerated watershed response time and an increase of storm water transportation due to the conversion from mixed forest to orchards.

Table 31. Water balance of Ban Da Na Kham watershed for 2005 and 2018.

Watershed Parameters	Land Cover 2005	Land Cover 2018
Total Infiltration (mm)	63	62
Total outflow (all flows) (mm)	127	127
Runoff (%)	47.7	47.7
Flood volume (in million (m ³))*	24.3	24
Flood area (km ²)*	12.1	11.8
Peak discharge (in thousands (l/s))	210	214
Peak time discharge (min)	348	315

*Values correspond to a flood threshold (flow height reported as flood) of 0.3 m

Other studies concerning land cover change and its associated effects on watershed hydrology, investigated similar transitions of forested areas into cultivated and urban land. Regardless of the simulated time scale, whether they are event-based (Barasa et al., 2018), seasonal (Lin et al., 2008) or perennial (Sajikumar et al., 2015), they emphasise an increase of peak discharges. Interestingly, a shift in peak time, as observed in this study was not reported. Sajikumar et al. (2015) detected similar a small increase in peak discharge. They claimed that similarities of land cover characteristics of classes which underwent the change (forest to plantation) mainly caused this effect.

In this study, the similarities in the main classes (mixed forest and orchard) are given by their effects on the underlying soil, reflected by high K_s values in SL1 which ensures full infiltration until saturation state independent from the class (Section 4.3.1). Differences, on the other hand, are the available pore space in the soil and the land cover characteristic flow resistance (Sections 3.4.1 & 4.3.1). Quantification of alterations of SHPs from land cover changes requires actual monitoring of those properties in space and time (Hassler, 2013). Studies from Nyberg et al. (2011), Peng et al. (2012) and Zimmermann et al. (2006), which investigated the transition of SHPs due to afforestation of cultivated land or vice versa, showed that it can take several decades until conditions close to full recovery or complete declination are reached, as soil exhibits some sort of memory effect. However, this process was ignored in this research as differences between orchards and mixed forest were considered as neglectable small. Furthermore, monitoring of SHPs is an activity which was technical, not feasible and would have been beyond the scope of this research study.

Runoff can be explained by two theories, infiltration excess and saturation excess. The former assumes runoff generation as soon as rainfall intensity exceeds the infiltration rate of the soil (Horton, 1933). The concept of saturation excess, on the other hand, describes runoff initiation caused fully saturated soil (Dunne & Black, 1970). In this study infiltration rates of SL1 are higher than the rainfall intensities, except for cropland and urban areas. For SL2 the opposite applies with a maximum K_s of 10.19 mm h⁻¹ in valleys (Section 4.3.1). Thus, during the modelled rainfall event first SL1 will become saturated as it has limited storage capacity due to its shallow depth of only 5 cm and almost saturated initial condition. Subsequently, infiltration and runoff are determined by the infiltration rate of SL2.

Table 32 shows the runoff amounts per land cover class for both years. The runoff amounts were calculated as precipitation - (infiltration + interception). It appears that the runoff experienced just a slight increase (0.6 %) from 2005 to 2018. It is evident that urban areas and cropland exhibit the highest average runoff amounts (>200 mm), in both years. Furthermore, both classes experience an increase in runoff (Table 32). In contrast, orchards, teak plantation and mixed forest show a considerably lower average runoff with approximately 200 mm (Table 32). A decrease in average runoff from 2005 to 2018 is only noticeable in the orchard and teak plantation class with <1 % for both.

The contribution of cropland area to the total runoff increased by 1.3 %. However, the runoff in cropland is difficult to quantify as rice fields are flooded during the rice production period, and no runoff as such is generated in reality. To model the flooded condition of the rice fields, those areas were parametrised as nearly flat areas, which are almost impermeable with K_s of <1 mm h⁻¹ (Section 4.3.1). Nevertheless, it was not possible to replicate the storing effect which would become effective due to the bunds surrounding the fields. In reality, the water which would overflow the bunds due to an exceedance of the storage capacity could potentially be considered as runoff.

Interesting is, that in areas of mixed forest, which is the only category where an area net decline occurred, the average runoff increased (Section 4.2.1 & Table 32). Associated to the changing land cover, the land cover distribution over the terrain has also changed. This has implications on the runoff initiation since soil depth, and with it, the available pore space varies in the terrain. Especially since the initial soil moisture was assumed to be homogenous, consequently, areas having less pore space and/or shallow soil will generate more surface water. Therefore, the increase in average runoff in the mixed forest class is based on the deforestation which predominately took place in the valleys and on the foot of the slopes. Both are areas which have (I) more storage capacity due to deeper soils, (II) higher K_s values of SL2, and (III) decelerated surface water velocities. Correspondingly, infiltration is promoted, and less runoff generated. In, 2018, mixed forest predominantly occurs in higher altitudes, on narrow summits and on steep slopes. Especially on steep slopes, more runoff is initiated as they possess comparably shallow soils. Additionally, an increasing slope gradient accelerates flow velocities. Consequently, infiltration is not only limited by the storage capacity of the soil but also by the enhanced surface flow velocities. This runoff terrain dependency is also transferable to the other land cover classes explaining changes in average runoff amounts. Despite the fact that soil moisture is not homogenous in the terrain and therefore the runoff amounts might deviate in reality, it is expected that the results give a fairly good indication of the land cover runoff relation within the watershed.

Table 32. Runoff amounts per land cover for 2005 and 2018.

Land Cover	2005			2018		
	Avg. infiltration (mm)	Avg. runoff (mm)	Runoff (%)*	Avg. infiltration (mm)	Avg. runoff (mm)	Runoff (%)*
Orchard	66.3	197.1	74.7	66.5	196.8	74.6
Cropland	28.4	235.1	89.1	26.4	237.1	89.9
Teak Plantation	63.4	199.4	75.6	63.5	199.3	75.6
Mixed Forest	63.1	199.8	75.8	62.3	200.6	76.1
Urban	21.3	242.4	91.9	20.1	243.6	92.4

* Runoff percentage is calculated as average runoff/precipitation

Besides causing alterations in runoff amounts, land cover changes also affected the subsequent flood and its distribution. This is due to the land cover characteristic flow resistance. Table 33 presents various flood depth classes with the corresponding inundated area for both years. Mixed forest is characterised by dense

undergrowth and is parameterised accordingly with a high flow resistance. Associated with the increased cultivation of orchards having considerably less flow resistance, as their ground is almost continuously covered by short grass, surface water accumulations in mountainous areas got reduced. Hence, the overall extent of surface water decreased due to the enhanced drainage. This process is evident as water depths of <5 m decreased from 21.0 km² (2005) to 20.4 km² (2018). In contrast, the water levels increased by ≥5 m from 11.7 (2005) km² to 11.9 km² (2018) (Table 33). This increase is shown in Figure 22 and mainly took place in the lowland areas.

Table 33. Flood depths and their corresponding extent due to land cover change.

Flood depth (m)	2005 Area (km ²)	2018 Area (km ²)	Change (%)
<0.3	10.1	9.8	-3
0.3	5.52	5.4	-2
1	3.94	3.7	-4
3	1.46	1.43	-2
5	1.04	1.06	1
>10	0.13	0.14	6

Changes in water depth also led to changes in flood durations. While in mountainous areas flood durations decreased with a reduction in water depth, an increase of the maximum flood durations in the lowland area is noticeable (Figure 23). Hence, in wide valleys where crop cultivation is practised and most of the urban settlements can be found flood water often remains for more than 10 h and in various locations also for more than 60 h (Figure 23). In conclusion, it is inherent that changes in land cover did not affect the flood dynamics of the watershed considerably, which could be expected as changes from a hydrological point of view were minor.

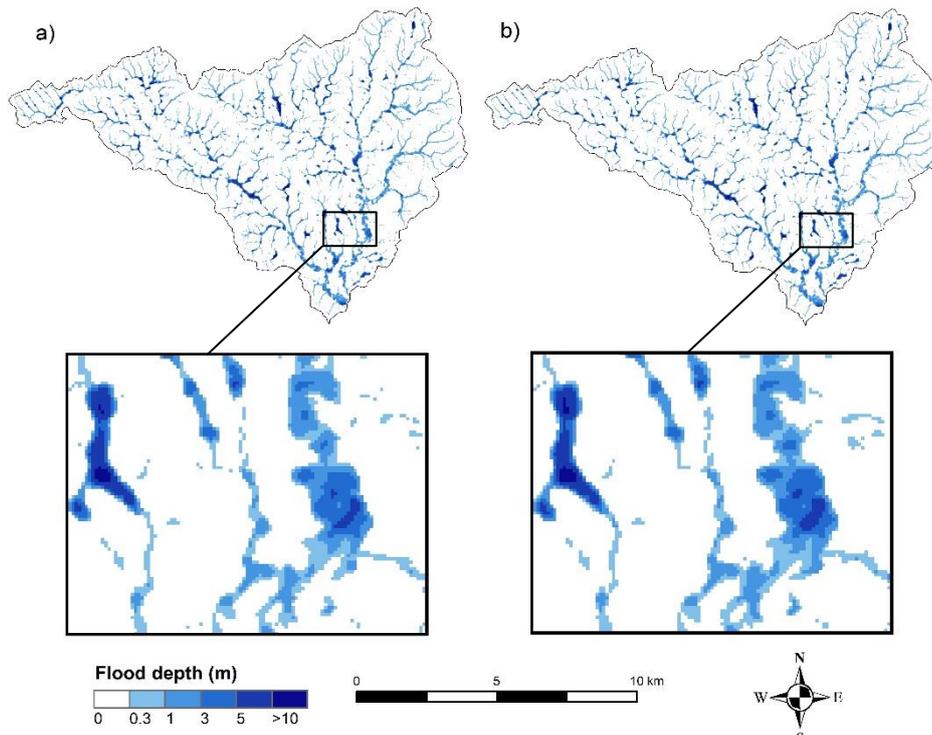


Figure 22. Maximum flood depth for 2005 **a)** and 2018 **b)**. Zoomed in area indicates an increase in maximum flood depth in the lowland area due to land cover changes.

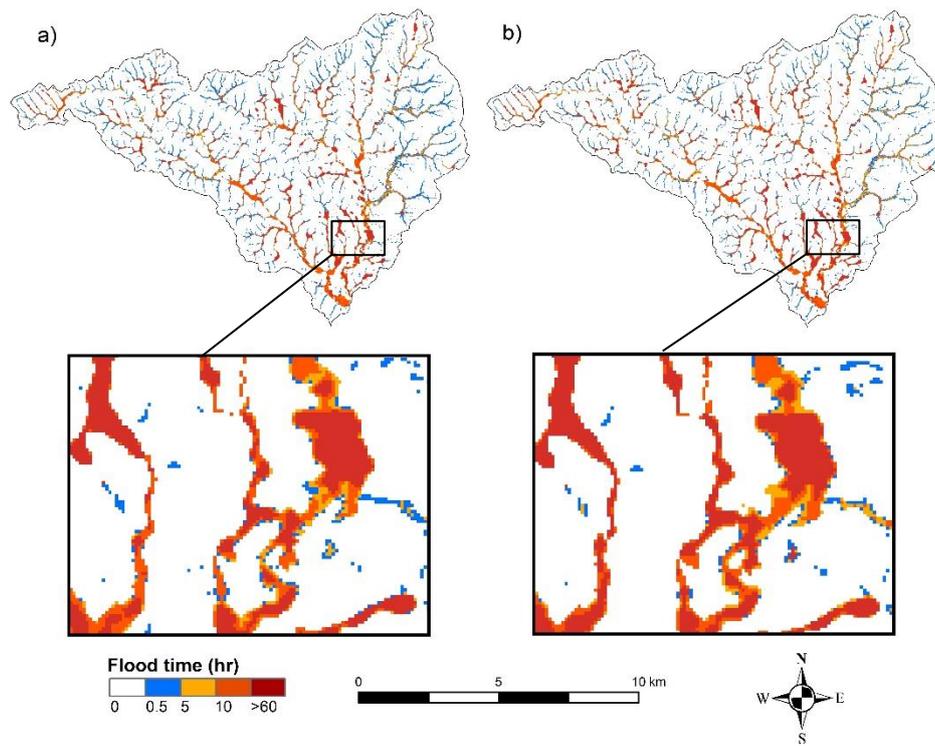


Figure 23. Flood duration for 2005 **a)** and 2018 **b)**. Zoomed in area indicates prolonged floods in the lowland areas due to land cover changes.

5. CONCLUSION AND RECOMMENDATIONS

This research analysed two soil datasets, (I) detailed field data and (II) SoilGrids to assess their similarity and to determine how sensitive the flood dynamics are in relation to their soil hydraulic properties when applied in an integrated flood model, to make a statement if SoilGrids represents a reasonable alternative in data scarce environments. In the second part, the land cover was mapped for two year, and land cover changes identified, subsequently the effects of those changes on runoff generation and flash flood behaviour were investigated.

The results imply that the field data and SoilGrids do not share many commonalities. Collected field data is highly influenced by land cover, whereas SoilGrids variability is limited throughout the watershed. SoilGrids is tailored to a global scale and therefore, does not necessarily take local phenomena into account. Hence, applying SoilGrids should be considered carefully. Especially when used for hydrological modelling on a watershed scale where differences in water holding capacity or infiltration are decisive. Soil physical properties such as clay content and bulk density were overestimated by SoilGrids. This, in turn, had great effects when predicting soil hydraulic properties such as saturated hydraulic conductivity using the pedotransfer functions of Saxton and Rawls (2006). Having predominantly an underestimation of saturated hydraulic conductivity led to an increased surface runoff during the modelling as rainfall intensities exceeded the infiltration rate although infiltration would have taken place in reality. Similarly, the overestimation of soil organic matter content by SoilGrids led to wrong water holding capacities of the soil. Those overestimations of soil physical and chemical properties and the subsequent false estimation of soil hydraulic properties by the pedotransfer functions ultimately had far-reaching consequences on the hydrology. An alteration in the runoff generation mechanism partly appeared, and flood extent, depth and duration increased considerably. Increasing soil depth affected both datasets similarly by promoting infiltration and reducing surface water. However, using field data, the flood dynamics were more sensitive to changes in soil depth. Overall, it can be concluded that the choice of soil input data (soil hydraulic properties and soil depth) had a great influence on both the quantity and spatial variability of infiltration which consequently affected runoff and flood dynamics.

The land cover change analysis confirmed the almost omnipresent trend of converting natural habitats into cultivated and urbanised areas. Subsequent flood simulations with changed land cover information indicated a slight intensified hydrological watershed response in the form of increased runoff amounts and greater peak discharge, as well as a reduction of the peak discharge time. Effects on general flood dynamics were minor as from a hydrological point of view no essential changes occurred. However, it can be concluded that national policies such as the Forestry Act enforced in 1989 by the Government of Thailand, aiming to stop commercial forest exploitation show no measurable effect in the research area. Consequently, strengthening of local level policy implementation and monitoring is urgently needed to ensure environmental protection and human security.

Furthermore, this study provides valuable insights into the implications of hydrological modelling in (tropical) data scarce environments beyond the original objectives of this research study, which might be of interest for future studies. **Firstly**, the findings reveal that complications which can arise with freely available elevation data in densely vegetated areas nearly outweighs the effects of using predicted soil hydraulic properties, retrieved from soil physical and chemical property layers of SoilGrids. The correct representation of the terrain is of at least equal importance for realistic flood simulations as soil data. It is not only a matter

of wrong water depths because of errors in the elevation, but the lack or the alteration in flow connectivity completely changes the flood dynamics. Without a precise digital elevation model, flood modelling is only useful in a relative sense (intercomparing of scenarios) but not when used for a flood hazard analysis which requires an accurate flood location. **Secondly**, in tropical areas where processes influencing the structure of soil (e.g., bioturbation) can be expected, it is advisable to develop a soil sampling strategy based on the prevailing land cover types rather than on the terrain or an existing soil texture map. Since biotic macropores might have a dominating influence on soil hydraulic properties instead of the soil texture. **Thirdly**, it is suggested to further subdivide the land cover class orchard for soil sampling and classification purposes to prevent soil property aggregation as experienced in this study. **Lastly**, additional sampling of subsurface soil (B horizon) is recommended as soil properties are expected to considerably change due to the absence of surface layer particularities (e.g., bioturbation) and soil hydraulic properties representing only the most upper layer do not suffice when modelling extreme events, that will have deeper infiltration.

In the course of this research, various **limitations** had to be faced. **Firstly**, the ruggedness of the terrain interfered a well-balanced soil sample distribution, hence summits and steep slopes and with them the mixed forest and teak plantation class are underrepresented in the soil analysis. **Secondly**, the representation of flooded rice fields in the model encountered difficulties. In the study area, the rice fields are surrounded by bunds having an outlet which controls the water level by enabling drainage during rainfall events. Therefore, during an extreme event, the rice fields would first function as some sort of reservoir storing the stormwater before water is released. Those interactions are complicated to model and need detailed investigations in order to make an adequate representation possible. Hence, in openLISEM, the rice fields had to be represented in a simplified way, which influenced the flood dynamics in the lowland areas. **Lastly**, the greatest limitation of this research is the quality of the DEM, which adversely affected flood dynamics and thereby led to an obstruction of the calibration efforts undertaken. Correction attempts and smoothing did not harvest any improvements, resulting in a wrong flood pattern. However, the goal of this research was not to do a hazard analysis, which would have been a futile exercise with any hydrological model when using this elevation model. Therefore, as the infiltration process and therefore, the overall runoff quantities can be expected to be correct, the quality of the digital elevation model does not invalidate the conclusions for the objectives set.

Future research should continue exploring SoilGrids data applicability for hydrological modelling as it represents a valuable source of information. To be able to make a profound statement about SoilGrids quality, it is necessary to conduct studies in various regions of the world. This will help to investigate its quality dependency on factors such as terrain, climate and vegetation.

List of References

- ADPC. (2006). *Rapid Assessment: Flashflood and Landslide Disaster in the Provinces of Uttaradit and Sukbothai Northern Thailand*. Bangkok. Retrieved from http://www.adpc.net/enewsjuly/Uttaradit_rapidassessment.pdf
- Ahuja, L. R., Naney, J. W., Green, R. E., & Nielsen, D. R. (2010). Macroporosity to Characterize Spatial Variability of Hydraulic Conductivity and Effects of Land Management. *Soil Science Society of America Journal*, 48(4), 699–702. <https://doi.org/10.2136/sssaj1940.036159950004000c0132x>
- Angiuli, E., & Trianni, G. (2014). Urban Mapping in Landsat Images Based on Normalized Difference Spectral Vector. *IEEE Geoscience and Remote Sensing Letters*, 11(3), 661–665. <https://doi.org/10.1109/LGRS.2013.2274327>
- Aon Benfield. (2012). *2011 Thailand Floods Event Recap Report*. Chicago. Retrieved from http://thoughtleadership.aonbenfield.com/Documents/20120314_impact_forecasting_thailand_flood_event_recap.pdf
- Arcement, G. J., & Schneider, V. R. (1989). *Guide for Selecting Manning's Roughness Coefficients for Natural Channels and Flood Plains* (Water-Supply Paper No. 2339). Retrieved from <https://pubs.usgs.gov/wsp/2339/report.pdf>
- Arrouays, D., Grundy, M. G., Hartemink, A. E., Hempel, J. W., Heuvelink, G. B. M., Hong, S. Y., ... Zhang, G.-L. (2014). GlobalSoilMap: Toward a Fine-Resolution Global Grid of Soil Properties. In *Advances in Agronomy* (Vol. 125, pp. 93–134). Academic Press. <https://doi.org/10.1016/B978-0-12-800137-0.00003-0>
- As-syakur, A. R., Adnyana, I. W. S., Arthana, I. W., & Nuarsa, I. W. (2012). Enhanced Built-Up and Bareness Index (EBBI) for Mapping Built-Up and Bare Land in an Urban Area. *Remote Sensing*, 4(10), 2957–2970. <https://doi.org/10.3390/rs4102957>
- ASF. (2015). *ASF radiometric terrain corrected products*.
- Baicha, W. (2016). Land use dynamics and land cover structure change in Thailand (as exemplified by mountainous Nan Province). *Geography and Natural Resources*, 37(1), 87–92. <https://doi.org/10.1134/s1875372816010121>
- Barasa, B. N., & Perera, E. D. P. (2018). Analysis of land use change impacts on flash flood occurrences in the Sosiani River basin Kenya. *International Journal of River Basin Management*, 16(2), 179–188. <https://doi.org/10.1080/15715124.2017.1411922>
- Bauer, E., & Kohavi, R. (1999). An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants. *Machine Learning*, 36(1/2), 105–139. <https://doi.org/10.1023/A:1007515423169>
- Bausch, W. C. (1993). Soil background effects on reflectance-based crop coefficients for corn. *Remote Sensing of Environment*, 46(2), 213–222. [https://doi.org/10.1016/0034-4257\(93\)90096-G](https://doi.org/10.1016/0034-4257(93)90096-G)
- Beers, T. W., Dress, P. E., & Wensel, L. C. (1966). Notes and observations: aspect transformation in site productivity research. *Journal of Forestry*, 64, 691–692. <https://doi.org/10.2307/2653667>
- Beon, M.-S., Cho, K. H., Kim, H. O., Oh, H.-K., & Jeong, J.-C. (2017). Mapping of Vegetation Using Multi-Temporal Downscaled Satellite Images of a Reclaimed Area in Saemangeum, Republic of Korea. *Remote Sensing*, 9(272), 2–17. <https://doi.org/10.3390/rs9030272>
- Blanchez, J. (1997). Forest Resources and Roundwood supply in the Asia Pacific Countries : Situation and Outlook to the Year 2010. *Asia - Pacific Forestry Sector Outlook Study Working Paper Series*. Retrieved from http://nkcs.org.np/dfrs/cfl/opac_css/index.php?lvl=notice_display&id=9114
- Bonfante, A., & Bouma, J. (2015). The role of soil series in quantitative land evaluation when expressing effects of climate change and crop breeding on future land use. *Geoderma*, 259–260, 187–195. <https://doi.org/10.1016/J.GEODERMA.2015.06.010>
- Boonprong, S., Cao, C., Chen, W., & Bao, S. (2018). Random forest variable importance spectral indices scheme for burnt forest recovery monitoring-multilevel RF-VIMP. *Remote Sensing*, 10(807), 2–15. <https://doi.org/10.3390/rs10060807>
- Boonyanuphap, J. (2013). Cost-benefit analysis of vetiver system-based rehabilitation measures for landslide-damaged mountainous agricultural lands in the lower Northern Thailand. *Natural Hazards*, 69(1), 599–629. <https://doi.org/10.1007/s11069-013-0730-y>

- Botula, Y.-D., Van Ranst, E., & Cornelis, W. M. (2014). Pedotransfer functions to predict water retention for soils of the humid tropics: a review. *Revista Brasileira de Ciência Do Solo*, 38(3), 679–698. <https://doi.org/10.1590/S0100-06832014000300001>
- Boumana, B. A. M., Wopereis, M. C. S., Kropff, M. J., Ten Berge, H. F. M., & Tuong, T. P. (1994). Water use efficiency of flooded rice fields II. Percolation and seepage losses. *Agricultural Water Management*, 26, 291–304. Retrieved from <https://core.ac.uk/download/pdf/29341530.pdf>
- Bout, B., & Jetten, V. G. (2017). The validity of flow approximations when simulating catchment-integrated flash floods. *Journal of Hydrology*, 556, 674–688. <https://doi.org/10.1016/j.jhydrol.2017.11.033>
- Bout, B., & Jetten, V. G. (2018). OpenLISEM Multi-Hazard Land Surface Process Model - Documentation & User Manual. Enschede: ITC Faculty of Geo-Information Science and Earth Observation of the University of Twente.
- Bout, B., Lombardo, L., van Westen, C. J., & Jetten, V. G. (2018). Integration of two-phase solid fluid equations in a catchment model for flashfloods, debris flows and shallow slope failures. *Environmental Modelling & Software*, 105, 1–16. <https://doi.org/10.1016/j.envsoft.2018.03.017>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5–32. Retrieved from <https://link.springer.com/content/pdf/10.1023%2FA%3A1010933404324.pdf>
- Bronstert, A., Niehoff, D., & Gerd, B. (2002). Effects of climate and land-use change on storm runoff generation : present knowledge and modelling capabilities. *Hydrological Processes*, 16(2), 509–529. <https://doi.org/10.1002/hyp.326>
- Brooks, R. H., & Corey, A. T. (1964). Hydraulic Properties of Porous Media. *Hydrology Papers*, 3, 1–25. Retrieved from https://mountainscholar.org/bitstream/handle/10217/61288/HydrologyPapers_n3.pdf?sequence
- Brovkina, O., Zemek, F., Novotný, J., Heřman, M., & Štěpánek, P. (2017). Analysing changes in land cover in relation to environmental factors in the districts of Znojmo and Třebíč (Czech Republic). *European Journal of Environmental Sciences*, 7(2), 108–118. <https://doi.org/10.14712/23361964.2017.9>
- CFE-DM. (2018). *Disaster Management Reference Handbook Thailand*. Joint Base Pearl Harbor - Hickam: Center for Excellence in Disaster Management and Humanitarian Assistance. Retrieved from <https://www.cfe-dmha.org>
- Chastain, R., Housman, I., Goldstein, J., Finco, M., & Tenneson, K. (2019). Empirical cross sensor comparison of Sentinel-2A and 2B MSI, Landsat-8 OLI, and Landsat-7 ETM+ top of atmosphere spectral characteristics over the conterminous United States. *Remote Sensing of Environment*, 221, 274–285. <https://doi.org/10.1016/j.rse.2018.11.012>
- Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., ... Mills, J. (2015). Global land cover mapping at 30 m resolution: A POK-based operational approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, 103, 7–27. <https://doi.org/10.1016/j.isprsjprs.2014.09.002>
- Chen, L., Van Westen, C. J., Hussin, H., Ciurean, R. L., Turkington, T., Chavarro-rincon, D., & Shrestha, D. P. (2016). Integrating expert opinion with modelling for quantitative multi-hazard risk assessment in the Eastern Italian Alps. *Geomorphology*, 273, 150–167. <https://doi.org/10.1016/j.geomorph.2016.07.041>
- Chow, V. T. (1959). *Open-Channel Hydraulics*. (H. E. Davis, Ed.) (Internatio). New York: McGraw Hill Book Company.
- Cingolani, A. M., Renison, D., Zak, M. R., & Cabido, M. R. (2004). Mapping vegetation in a heterogeneous mountain rangeland using landsat data: An alternative method to define and classify land-cover units. *Remote Sensing of Environment*, 92(1), 84–97. <https://doi.org/10.1016/j.rse.2004.05.008>
- Coburn, C. A., & Roberts, A. C. B. (2004). A multiscale texture analysis procedure for improved forest stand classification. *International Journal of Remote Sensing*, 25(20), 4287–4308. <https://doi.org/10.1080/0143116042000192367>
- Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, 20(1), 37–46. Retrieved from <https://journals-sagepub-com.ezproxy2.utwente.nl/doi/pdf/10.1177/001316446002000104>
- Colby, J. D. (1991). Topographic Normalization in Rugged Terrain. *Photogrammetric Engineering & Remote Sensing*, 57(5), 531–537. Retrieved from <https://pdfs.semanticscholar.org/4f0c/45d90bbae073872c2d4cd73aa4fa2e6da4c9.pdf>

- Colditz, R. R. (2015). An evaluation of different training sample allocation schemes for discrete and continuous land cover classification using decision tree-based algorithms. *Remote Sensing*, 7(8), 9655–9681. <https://doi.org/10.3390/rs70809655>
- Comber, A., Balzter, H., Cole, B., Fisher, P., Johnson, S., & Ogotu, B. (2016). Methods to Quantify Regional Differences in Land Cover Change. *Remote Sensing*, 8(176), 1–19. <https://doi.org/10.3390/rs8030176>
- Congalton, R. G. (1991). A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data. *Remote Sensing of Environment*, 37, 35–46. Retrieved from <https://pdfs.semanticscholar.org/d7be/d062683df1fd6723fb9c0c1d26feddc8c133.pdf>
- Crist, E. P., & Cicone, R. C. (1984). A Physically-Based Transformation of Thematic Mapper Data-The TM Tasseled Cap. *IEEE Transactions on Geoscience and Remote Sensing*, 22(3), 256–263. Retrieved from http://www.geo.umass.edu/courses/geo594q/Crist1984IEEE_TasseledCap.pdf
- D'Souza, G. (2000). *Forest Cover Mapping & Monitoring with NOAA-AVHRR & other Coarse Spatial Resolution Sensors* (No. 29). Rome. Retrieved from www.fao.org/fo
- De Paz, M., Gobbi, M. E., & Raffaele, E. (2018). Fallen fruits stimulate decomposition of leaf litter of dominant species in NW Patagonia shrublands. *Plant and Soil*, 425(1–2), 433–440. <https://doi.org/10.1007/s11104-018-3590-0>
- De Roo, A. P. J., Wesseling, C. G., & Ritsema, C. J. (1996). LISEM: a single event physically-based hydrologic and soil erosion model for drainage basins. I: Theory, input and output. *Hydrological Processes*, 10(April 1995), 1107–1117.
- Devia, G. K., & Ganasri, B. P. (2015). A Review on Hydrological Models. *Aquatic Procedia*, 4, 1001–1007. <https://doi.org/10.1016/J.AQPRO.2015.02.126>
- DHI. (2017). Mike 11 - Reference Manual. DHI. Retrieved from http://manuals.mikepoweredbydhi.help/2017/Water_Resources/Mike_11_ref.pdf
- Di Gregorio, A. (2005). *Land cover classification system : classification concepts and user manual : LCCS* (No. 2). Rome. Retrieved from <http://www.fao.org/3/y7220e/y7220e00.htm>
- Di Gregorio, A. (2016). *Classification Concepts Land Cover Classification System Software version 3* (No. 3). Food and Agriculture Organisation of the United Nations. Retrieved from www.fao.org/
- Di Gregorio, A., & Jansen, L. J. M. (2000). *Land cover classification system (LCCS): classification concepts and user manual*. Rome: Food and Agriculture Organization of the United Nations. Retrieved from <http://www.fao.org/3/x0596e/x0596e00.htm>
- Domaç, A., & Süzen, M. L. (2006). Integration of environmental variables with satellite images in regional scale vegetation classification. *International Journal of Remote Sensing*, 27(7), 1329–1350. <https://doi.org/10.1080/01431160500444806>
- Dunne, T., & Black, R. D. (1970). Partial Area Contributions to Storm Runoff in a Small New England Watershed. *Water Resources Research*, 6(5), 1296–1311.
- Eijkelkamp. (2013). Laboratory-Permeameters Operating Instructions. Giesbeek: Eijkelkamp Agrisearch Equipment.
- Faber-Langendoen, D., Tart, D. L., & Crawford, R. H. (2009). Contours of the Revised U.S. National Vegetation Classification Standard. *Bulletin of the Ecological Society of America*, 90(1), 87–93. <https://doi.org/10.1890/0012-9623-90.1.87>
- FAO. (2009). *Thailand forestry outlook study*. Bangkok. Retrieved from <http://www.fao.org/docrep/014/am617e/am617e00.pdf>
- Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., ... Alsdorf, D. (2007). The Shuttle Radar Topography Mission. *Reviews of Geophysics*, 45, 1–33. <https://doi.org/10.1029/2005RG000183>
- Field, A. (2009). *Discovering Statistics Using SPSS* (3rd ed.). London: Sage Publications. <https://doi.org/10.14359/51686441>
- Flood, N. (2013). Seasonal composite landsat TM/ETM+ Images using the medoid (a multi-dimensional median). *Remote Sensing*, 5(12), 6481–6500. <https://doi.org/10.3390/rs5126481>
- Forsyth, T. (2007). Sustainable livelihood approaches and soil erosion risks: Who is to judge? *International Journal of Social Economics*, 34(1–2), 88–102. <https://doi.org/10.1108/03068290710723381>
- Franklin, S. E., & Wulder, M. A. (2002). Remote sensing methods in medium spatial resolution satellite data land cover classification of large areas. *Progress in Physical Geography*, 26(2), 173–205. <https://doi.org/10.1191/0309133302pp332ra>
- García-Gutiérrez, C., Pachepsky, Y., & Martín, M. Á. (2018). Technical note: Saturated hydraulic conductivity and textural heterogeneity of soils. *Hydrol. Earth Syst. Sci.*, 22, 3923–3932. <https://doi.org/10.5194/hess-22-3923-2018>

- Gijsman, A. J., Jagtap, S. S., & Jones, J. W. (2003). Wading through a swamp of complete confusion : how to choose a method for estimating soil water retention parameters for crop models. *European Journal of Agronomy*, 18, 77–106. [https://doi.org/10.1016/S1161-0301\(02\)00098-9](https://doi.org/10.1016/S1161-0301(02)00098-9)
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, 202, 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>
- Green, W. H., & Ampt, G. A. (1911). Studies on Soil Physics. *The Journal of Agricultural Science*, 4(1), 1–24. <https://doi.org/10.1017/S0021859600001441>
- Gregory, J. H., Dukes, M., Jones, P., & Miller, G. (2006). Effect of urban soil compaction on infiltration rate. *Journal of Soil and Water Conservation*, 61(3), 117–124.
- Grunwald, S., Thompson, J. A., & Boettinger, J. L. (2011). Digital Soil Mapping and Modeling at Continental Scales: Finding Solutions for Global Issues. *Soil Science Society of America Journal*, 75(4), 1201–1213. <https://doi.org/10.2136/sssaj2011.0025>
- Gu, D., & Gillespie, A. (1998). Topographic Normalization of Landsat TM Images of Forest Based on Subpixel Sun–Canopy–Sensor Geometry. *Remote Sensing of Environment*, 64(2), 166–175. [https://doi.org/10.1016/S0034-4257\(97\)00177-6](https://doi.org/10.1016/S0034-4257(97)00177-6)
- Hamer, U., & Marschner, B. (2005). Priming effects in soils after combined and repeated substrate additions. *Geoderma*, 128(1–2), 38–51. <https://doi.org/10.1016/J.GEODERMA.2004.12.014>
- Han, E., Ines, A., & Koo, J. (2015). *Global High-Resolution Soil Profile Database for Crop Modeling Applications*. Washington DC. Retrieved from <https://harvestchoice.org/publications/global-high-resolution-soil-profile-database-crop-modeling-applications>
- Han, J., Kamber, M., & Pei, J. (2013). Getting to Know Your Data. In *Data Mining Concepts and Techniques* (3rd ed., pp. 66–79). Amsterdam: Elsevier. Retrieved from <http://ir.obihiro.ac.jp/dspace/handle/10322/3933>
- Hansen, M. C., Egorov, A., Roy, D. P., Potapov, P., Ju, J., Turubanova, S., ... Loveland, T. R. (2011). Continuous fields of land cover for the conterminous United States using Landsat data: first results from the Web-Enabled Landsat Data (WELD) project. *Remote Sensing Letters*, 2(4), 279–288. <https://doi.org/10.1080/01431161.2010.519002>
- Hassler, S. K. (2013). *Saturated hydraulic conductivity in the humid tropics : sources of variability, implications for monitoring and effects on near-surface hydrological flow paths*. University of Potsdam. Retrieved from <http://opus.kobv.de/ubp/volltexte/2013/6686/>
- Hassler, S. K., Lark, R. M., Milne, A. E., & Elsenbeer, H. (2011). Exploring the variation in soil saturated hydraulic conductivity under a tropical rainforest using the wavelet transform. *European Journal of Soil Science*, 62, 891–901. <https://doi.org/10.1111/j.1365-2389.2011.01400.x>
- Hengl, T., Heuvelink, G. B. M., Kempen, B., Leenaars, J. G. B., Walsh, M. G., Shepherd, K. D., ... Tondoh, J. E. (2015). Mapping Soil Properties of Africa at 250 m Resolution: Random Forests Significantly Improve Current Predictions. *PLOS ONE*, 10(6), 1–26. <https://doi.org/10.1371/journal.pone.0125814>
- Hengl, T., Mendes De Jesus, J., Heuvelink, G. B. M., Gonzalez, M. R., Kilibarda, M., Blagotić, A., ... Blagotić, M. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE*, 12(2), 1–40. <https://doi.org/10.1371/journal>
- Hessel, R., Jetten, V., & Guanghui, Z. (2003). Estimating Manning's n for steep slopes. *CATENA*, 54(1–2), 77–91. [https://doi.org/10.1016/S0341-8162\(03\)00058-4](https://doi.org/10.1016/S0341-8162(03)00058-4)
- Hislop, S., Jones, S., Soto-Berelev, M., Skidmore, A., Haywood, A., & Nguyen, T. H. (2018). Using landsat spectral indices in time-series to assess wildfire disturbance and recovery. *Remote Sensing*, 10(3), 1–17. <https://doi.org/10.3390/rs10030460>
- Horton, R. E. (1933). The Role of Infiltration in the Hydrologic Cycle. *Transactions, American Geophysical Union*, 446–460.
- Housman, I., Tanpipat, V., Biswas, T., Clark, A., Stephen, P., Maus, P., & Megown, K. (2015). *Monitoring Forest Change in Southeast Asia : Case Studies for USAID Lowering Emissions in Asia's Forests*. Retrieved from <http://www.leafasia.org/sites/default/files/public/resources/USAIDLEAF-RSAC-ForestLossReport-November2015-Revised.pdf>
- Hudson, B. D. (1994). Soil organic matter and available water capacity. *Journal of Soil and Water Conservation*, 49(2), 189–194.
- Huete, A. R. (1988). A soil-adjusted vegetation index (SAVI). *Remote Sensing of Environment*, 25(3), 295–309. [https://doi.org/10.1016/0034-4257\(88\)90106-X](https://doi.org/10.1016/0034-4257(88)90106-X)

- Hussein, S. O., Kovács, F., & Tobak, Z. (2017). Spatiotemporal assessment of vegetation indices and land cover for Erbil City and its surrounding using MODIS imageries. *Journal of Environmental Geography*, 10(2), 31–39. <https://doi.org/10.1515/jengeo-2017-0004>
- IFRC. (2016). *World Disasters Report. International Federation of Red Cross and Red Crescent Societies* (Vol. 91). <https://doi.org/10.2307/2998097>
- IPCC. (2014). *Summary for Policymakers. Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. <https://doi.org/10.1017/CBO9781107415324>
- Jasiewicz, J., & Stepinski, T. F. (2013). Geomorphons — a pattern recognition approach to classification and mapping of landforms. *Geomorphology*, 182, 147–156. <https://doi.org/10.1016/j.geomorph.2012.11.005>
- Jiang, Z., Huete, A. R., Didan, K., & Miura, T. (2008). Development of a two-band enhanced vegetation index without a blue band. *Remote Sensing of Environment*, 112(10), 3833–3845. <https://doi.org/10.1016/J.RSE.2008.06.006>
- Kalyanapu, A. J., Burian, S. J., & Mcpherson, T. N. (2009). Effect of land use-based surface roughness on hydrologic model output. *Journal of Spatial Hydrology*, 9(2).
- Khatami, R., Mountrakis, G., & Stehman, S. V. (2016). A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research. *Remote Sensing of Environment*, 177(May), 89–100. <https://doi.org/10.1016/j.rse.2016.02.028>
- Kim, C. (2016). Land use classification and land use change analysis using satellite images in Lombok Island, Indonesia. *Forest Science and Technology*, 12(4), 183–191. <https://doi.org/10.1080/21580103.2016.1147498>
- Kizilkaya, R., & Dengiz, O. (2010). Variation of land use and land cover effects on some soil physico-chemical characteristics and soil enzyme activity. *Zemdirbyste-Agriculture*, 97(2), 15–24. Retrieved from <https://pdfs.semanticscholar.org/2f5b/b343e51e503b68b6675b6c7a40a0fd8ddd1b.pdf>
- Koppad, A. G., & Janagoudar, B. S. (2013). Vegetation analysis and land use land cover classification of forest in Uttara Kannada district India using remote sensign and GIS techniques. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, 42(4W5), 121–125. <https://doi.org/10.5194/isprs-archives-XLII-4-W5-121-2017>
- Kuriakose, S. L., Devkota, S., Rossiter, D. G., & Jetten, V. G. (2009). Prediction of soil depth using environmental variables in an anthropogenic landscape, a case study in the Western Ghats of Kerala, India. *Catena*, 79(1), 27–38. <https://doi.org/10.1016/j.catena.2009.05.005>
- Lado, M., Paz, A., & Ben-Hur, M. (2004). Organic Matter and Aggregate-Size Interactions in Saturated Hydraulic Conductivity. *Soil Science Society of America Journal*, 68(1), 234–242. <https://doi.org/10.2136/sssaj2004.2340>
- Lemmon, P. E. (2008). *Using Forest Densimeters - Documentation*. Jackson: Forestry Suppliers, Inc.
- Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, 2(3), 18–22. <https://doi.org/10.1159/000323281>
- Lin, Y., & Wei, X. (2008). The impact of large-scale forest harvesting on hydrology in the Willow watershed of Central British Columbia. *Journal of Hydrology*, 359(1–2), 141–149. <https://doi.org/10.1016/j.jhydrol.2008.06.023>
- Looy, K. Van, Bouma, J., Herbst, M., Koestel, J., Minasny, B., Mishra, U., ... Looy, V. (2017). Pedotransfer Functions in Earth System Science: Challenges and Perspectives. *AGU Publications*. <https://doi.org/10.1002/2017RG000581>
- Lubis, J. P. G., & Nakagoshi, N. (2011). Land Use and Land Cover Change Detection using Remote Sensing and Geographic Information System in Bodri Watershed, Central Java, Indonesia. *Journal of International Development and Cooperation*, 18(1), 139–151.
- Mallupattu, P. K., & Reddy, J. S. R. (2013). Analysis of Land Use/Land Cover Changes Using Remote Sensing Data and GIS at an Urban Area, Tirupati, India. *The Scientific World Journal*, 2013, 1–6. <https://doi.org/10.1155/2013/268623>
- Malo, D. D., Worcester, B. K., Cassel, D. K., & Matzdorf, K. D. (1974). Soil-Landscape Relationships in a Closed Drainage System. *Soil Science Society of America Journal*, 38(5), 813–818. <https://doi.org/10.2136/sssaj1974.385813x>
- Marchi, L. (2017). Linking Debris Flows and Landslides to Large Floods. In D. Tsutsumi & J. B. Laronne (Eds.), *Gravel-bed rivers: processes and disasters* (1st ed., pp. 467–494). John Wiley & Sons Ltd.

- Marchi, L., Borga, M., Preciso, E., & Gaume, E. (2010). Characterisation of selected extreme flash floods in Europe and implications for flood risk management. *Journal of Hydrology*, 394(1–2), 118–133. <https://doi.org/10.1016/j.jhydrol.2010.07.017>
- Mas, J.-F. (1999). Monitoring land-cover changes: a comparison of change detection techniques*. *International Journal of Remote Sensing*, 20(1), 139–152. Retrieved from <http://www.tandf.co.uk/JNLS/res.htm>
- Maselli, F., Conese, C., De Filippis, T., & Romani, M. (1995). Integration of ancillary data into a maximum-likelihood classifier with nonparametric priors. *ISPRS Journal of Photogrammetry and Remote Sensing*, 50(2), 2–11. [https://doi.org/10.1016/0924-2716\(95\)98210-Q](https://doi.org/10.1016/0924-2716(95)98210-Q)
- McBratney, A. B., Mendonça Santos, M. L., & Minasny, B. (2003). On digital soil mapping. *Geoderma*, 117(1–2), 3–52. [https://doi.org/10.1016/S0016-7061\(03\)00223-4](https://doi.org/10.1016/S0016-7061(03)00223-4)
- McFeeters, S. K. (1996). The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features. *International Journal of Remote Sensing*, 17(7), 1425–1432. <https://doi.org/10.1080/01431169608948714>
- McKenzie, N. J., Grundy, M., Webster, R., & Ringrose-Voase, A. J. (2008). *Guidelines for Surveying Soil and Land Resources* (2nd ed.). Melbourne: CSIRO Publishing.
- Merritt, W. S., Letcher, R. A., & Jakeman, A. J. (2003). A review of erosion and sediment transport models. *Environmental Modelling & Software*, 18, 761–799. [https://doi.org/10.1016/S1364-8152\(03\)00078-1](https://doi.org/10.1016/S1364-8152(03)00078-1)
- Miller, B. A., & Schaetzl, R. J. (2015). Digital Classification of Hillslope Position. *Soil Science Society of America Journal*, 79(1), 132–145. <https://doi.org/10.2136/sssaj2014.07.0287>
- Möller, M., Volk, M., Friedrich, K., & Lymburner, L. (2008). Placing soil-genesis and transport processes into a landscape context: A multiscale terrain-analysis approach. *Journal of Plant Nutrition and Soil Science*, 171(3), 419–430. <https://doi.org/10.1002/jpln.200625039>
- Morgan, R. P. C., Quinton, J. N., Smith, R. E., Govers, G., Prosen, J. W. A., Auerswald, K., ... Folley, A. J. V. (1998). *The European Soil Erosion Model (EUROSEM): documentation and user guide*.
- Muñoz, C., Zagal, E., & Ovalle, C. (2007). Influence of trees on soil organic matter in Mediterranean agroforestry systems: an example from the “Espinal” of central Chile. *European Journal of Soil Science*, 58(3), 728–735. <https://doi.org/10.1111/j.1365-2389.2006.00858.x>
- Nemes, A., Rawls, W. J., & Pachepsky, Y. A. (2005). Influence of Organic Matter on the Estimation of Saturated Hydraulic Conductivity. *Soil Science Society of America Journal*, 69(4), 1330. <https://doi.org/10.2136/sssaj2004.0055>
- Nguyen, L. H., Joshi, D. R., Clay, D. E., & Henebry, G. M. (2018). Characterizing land cover/land use from multiple years of Landsat and MODIS time series: A novel approach using land surface phenology modeling and random forest classifier. *Remote Sensing of Environment*. <https://doi.org/10.1016/J.RSE.2018.12.016>
- Nyberg, G., Bargués Tobella, A., Kinyangi, J., & Ilstedt, U. (2012). Soil property changes over a 120-yr chronosequence from forest to agriculture in western Kenya. *Hydrology and Earth System Sciences*, 16(7), 2085–2094. <https://doi.org/10.5194/hess-16-2085-2012>
- Ohm, H., Hamer, U., & Marschner, B. (2007). Priming effects in soil size fractions of a podzol Bs horizon after addition of fructose and alanine. *Journal of Plant Nutrition and Soil Science*, 170(4), 551–559. <https://doi.org/10.1002/jpln.200625087>
- Ossola, A., Hahs, A. K., & Livesley, S. J. (2015). Habitat complexity influences fine scale hydrological processes and the incidence of stormwater runoff in managed urban ecosystems. *Journal of Environmental Management*, 159, 1–10. <https://doi.org/10.1016/j.jenvman.2015.05.002>
- Panahi, A., Alijani, B., & Mohammadi, H. (2010). The Effect of the Land Use/Cover Changes on the Floods of the Madarsu Basin of Northeastern Iran. *Journal of Water Resource and Protection*, 02(04), 373–379. <https://doi.org/10.4236/jwarp.2010.24043>
- Paramananthan, S., Lee, P.-X., Wong, M.-K., Van Ranst, E., Wüst, R. A. J., & Vijiandran, J. R. (2018). A Comparative Study of the Use of Organic Carbon and Loss on Ignition in Defining Tropical Organic Soil Materials. *Communications in Soil Science and Plant Analysis*, 49(5), 626–634. <https://doi.org/10.1080/00103624.2018.1435683>
- Pekel, J.-F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. *Nature*, 540(7633), 418–422. <https://doi.org/10.1038/nature20584>

- Peng, S. L., Wu, J., & You, W. H. (2012). Recovery of saturated hydraulic conductivity along a forest successional series from abandoned land to mature, evergreen broad-leaved forest in eastern China. *Soil Research*, *50*, 257–266. Retrieved from <http://tiantong.ecnu.edu.cn/ueditor/net/upload/2014-09-19/035e24eb-4005-4179-a321-de1f616db2a8.pdf>
- Pérez-Molina, E., Sliuzas, R., Flacke, J., & Jetten, V. (2017). Developing a cellular automata model of urban growth to inform spatial policy for flood mitigation: A case study in Kampala, Uganda. *Computers, Environment and Urban Systems*, *65*, 53–65. <https://doi.org/10.1016/j.compenvurbysys.2017.04.013>
- Powell, S. L., Cohen, W. B., Healey, S. P., Kennedy, R. E., Moisen, G. G., Pierce, K. B., & Ohmann, J. L. (2009). Quantification of live aboveground forest biomass dynamics with Landsat time-series and field inventory data: A comparison of empirical modeling approaches. *Remote Sensing of Environment*, *114*, 1053–1068. <https://doi.org/10.1016/j.rse.2009.12.018>
- Pramono, A. ., Fauzi, M. A., Widyani, N., Heriansyah, I., & Roshetko, J. M. (2015). *Managing smallholder teak plantations: Field guide for farmers* (1st ed.). Bogor: Center for International Forestry Research (CIFOR). <https://doi.org/10.17528/cifor/003493>
- R Core Team. (2018). A Language and Environment for Statistical Computing. Retrieved June 17, 2019, from <https://www.r-project.org/>
- Raudkivi, A. J. (1979). *Hydrology: an advanced introduction to hydrological processes and modelling* (1st ed.). Oxford: Pergamon Press.
- Riaño, D., Chuvieco, E., Salas, J., & Aguado, I. (2003). Assessment of Different Topographic Corrections in Landsat-TM Data for Mapping Vegetation Types. *IEEE Transactions on Geoscience and Remote Sensing*, *41*(5). <https://doi.org/10.1109/TGRS.2003.811693>
- Rittima, A., Piemfa, K., Uthai, N., & Jantaramana, A. (2013). Improvement of Design Rainfall Analysis of the Central Basin of Thailand. *Engineering Research and Development*, *24*(4), 28–38.
- Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, *67*(1), 93–104. <https://doi.org/10.1016/j.isprsjprs.2011.11.002>
- Ross, S. M. (1993). Organic matter in tropical soils: current conditions, concerns and prospects for conservation. *Progress in Physical Geography: Earth and Environment*, *17*(3), 265–305. <https://doi.org/10.1177/030913339301700301>
- Roy, D.P., Zhang, H. K., Ju, J., Gomez-Dans, J. L., Lewis, P. E., Schaaf, C. B., ... Kovalsky, V. (2016). A general method to normalize Landsat reflectance data to nadir BRDF adjusted reflectance. *Remote Sensing of Environment*, *176*, 255–271. <https://doi.org/10.1016/J.RSE.2016.01.023>
- Roy, David P., Li, J., Zhang, H. K., Yan, L., Huang, H., & Li, Z. (2017). Examination of Sentinel-2A multi-spectral instrument (MSI) reflectance anisotropy and the suitability of a general method to normalize MSI reflectance to nadir BRDF adjusted reflectance. *Remote Sensing of Environment*, *199*, 25–38. <https://doi.org/10.1016/J.RSE.2017.06.019>
- Rwanga, S. S., & Ndambuki, J. M. (2017). Accuracy Assessment of Land Use / Land Cover Classification Using Remote Sensing and GIS. *International Journal of Geosciences*, (8), 611–622. <https://doi.org/10.4236/ijg.2017.84033>
- Saah, D., Tenneson, K., Poortinga, A., Nguyen, Q., Chishtie, F., & San Aung, K. (2019). A novel modular free and open-source system for landcover mapping using cloud-based remote sensing and machine learning. *Remote Sensing of Environment, in review*.
- Sajikumar, N., & Remya, R. S. (2015). Impact of land cover and land use change on runoff characteristics. *Journal of Environmental Management*, *161*, 460–468. <https://doi.org/10.1016/j.jenvman.2014.12.041>
- Salomonson, V. ., & Appel, I. (2004). Estimating fractional snow cover from MODIS using the normalized difference snow index. *Remote Sensing of Environment*, *89*(3), 351–360. <https://doi.org/10.1016/J.RSE.2003.10.016>
- Sanchez-Moreno, J. F., Jetten, V., Mannaerts, C. M., & Tavares, J. P. (2014). Selecting best mapping strategies for storm runoff modeling in a mountainous semi-arid area. *Earth Surface Processes and Landforms*, *39*(8), 1030–1048. <https://doi.org/10.1002/esp.3501>
- Sanchez, P. A., Ahamed, S., Carre, F., Hartemink, A. E., Hempel, J., Huising, J., ... Zhang, G.-L. (2009). Digital Soil Map of the World. *Science*, *325*(5941), 680–681. <https://doi.org/10.1126/science.1175084>

- Saxton, K. E., & Rawls, W. J. (2006). Soil Water Characteristic Estimates by Texture and Organic Matter for Hydrologic Solutions. *Soil Science Society of America Journal*, 70, 1569–1578.
<https://doi.org/10.2136/sssaj2005.0117>
- Saxton, K. E., Rawls, W. J., Romberger, J. S., & Papendick, R. I. (1986). Estimating Generalized Soil-water Characteristics from Texture. *Soil Science Society of America Journal*, 50(4), 1031–1036.
<https://doi.org/10.2136/sssaj1986.03615995005000040039x>
- Schaetzl, R. J., & Anderson, S. (2005). *Soils Genesis and Geomorphology* (1st ed.). New York: Cambridge University Press. <https://doi.org/10.2136/sssaj2014.07.0287>
- Scharffenberg, W. (2016). Hydrologic Modeling System HEC-HMS: User’s Manual. Davis. Retrieved from http://www.hec.usace.army.mil/software/hec-hms/documentation/HEC-HMS_Users_Manual_4.2.pdf
- Schmidt, G., Jenkerson, C., Masek, J., Vermote, E., & Gao, F. (2013). *Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) Algorithm Description, Technical Report*. Reston. Retrieved from <http://www.usgs.gov/pubprod>
- Schoeneberger, P. J., Wysocki, D. A., & Benham, E. C. (2012). *Field Book for Describing and Sampling Soils* (3rd ed.). Lincoln: National Resources Conservation Service, National Soil Survey Center. Retrieved from https://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcs142p2_052523.pdf
- Schulte, E. E., & Hopkins, B. G. (1996). Estimation of Soil Organic Matter by Weight Loss-On-Ignition. In *Soil Organic Matter: Analysis and Interpretation* (pp. 21–31). Soil Science Society of America. <https://doi.org/10.2136/sssaspecpub46.c3>
- Shekhar, S., & Xiong, H. (2008). Soil and Water Assessment Tool “SWAT.” In *Encyclopedia of GIS*. Boston, MA: Springer US. https://doi.org/10.1007/978-0-387-35973-1_1231
- Shrestha, D. P. (2014). *INCREO project report - rainfall-runoff modelling in the Fella river basin*. Enschede.
- Sidle, R. C., Ziegler, A. D., Negishi, J. N., Nik, A. R., Siew, R., & Turkelboom, F. (2006). Erosion processes in steep terrain—Truths, myths, and uncertainties related to forest management in Southeast Asia. *Forest Ecology and Management*, 224(1–2), 199–225.
<https://doi.org/10.1016/J.FORECO.2005.12.019>
- Sliuzas, R., Flacke, J., & Jetten, V. (2013). Modelling Urbanization and Flooding in Kampala, Uganda. In *Network-Association of European Researchers on Urbanisation in the South (N-AERUS) XIV* (pp. 12–14). Enschede. Retrieved from <https://pdfs.semanticscholar.org/546f/2ea13e914dc89c141eff4816312ea5198186.pdf>
- Sluiter, R., & Pebesma, E. J. (2010). Comparing techniques for vegetation classification using multi- and hyperspectral images and ancillary environmental data. *International Journal of Remote Sensing*, 31(23), 6143–6161. <https://doi.org/10.1080/01431160903401379>
- Smith, J. A., Lin, T. L., & Ranson, K. J. (1980). The Lambertian Assumption and Landsat Data. *Photogrammetric Engineering and Remote Sensing*, 46(9), 1183–1189. Retrieved from https://www.asprs.org/wp-content/uploads/pers/1980journal/sep/1980_sep_1183-1189.pdf
- Sobieraj, J. A., Elsenbeer, H., Coelho, R. M., & Newton, B. (2002). Spatial variability of soil hydraulic conductivity along a tropical rainforest catena. *Geoderma*, 108(1–2), 79–90.
[https://doi.org/10.1016/S0016-7061\(02\)00122-2](https://doi.org/10.1016/S0016-7061(02)00122-2)
- Soenen, S. A., Peddle, D. R., & Coburn, C. A. (2005). SCS+C: a modified Sun-canopy-sensor topographic correction in forested terrain. *IEEE Transactions on Geoscience and Remote Sensing*, 43(9), 2148–2159.
<https://doi.org/10.1109/TGRS.2005.852480>
- Soil Survey Staff. (2014). *Soil Survey Field and Laboratory Methods Manual*.
- Steinhausen, M. J., Wagner, P. D., Narasimhan, B., & Waske, B. (2018). Combining Sentinel-1 and Sentinel-2 data for improved land use and land cover mapping of monsoon regions. *International Journal of Applied Earth Observation and Geoinformation*, 73(April), 595–604.
<https://doi.org/10.1016/j.jag.2018.08.011>
- Stepinski, T. F., & Jasiewicz, J. (2011). Geomorphons - a new approach to classification of landforms. *Geomorphometry*, 109–112.
- Tifafi, M., Guenet, B., & Hatté, C. (2018). Large Differences in Global and Regional Total Soil Carbon Stock Estimates Based on SoilGrids, HWSD, and NCSCD: Intercomparison and Evaluation Based on Field Data From USA, England, Wales, and France. *Global Biogeochemical Cycles*, 32(1), 42–56.
<https://doi.org/10.1002/2017GB005678>

- Tilaar, M., Wih, W. L., Ranti, A. S., Wasitaatmadja, S. M., Junardy, F. D., & Maily. (2008). Review of *Lansium domesticum* Corrêa and its use in cosmetics. *Boletín Latinoamericano y Caribe de Plantas Medicinales y Aromáticas*, 7(4), 183–189. Retrieved from <https://pdfs.semanticscholar.org/a844/b6f156e3db3a12ded31f65c5e63ef9c70d11.pdf>
- Tsai, Y. H., Stow, D., Chen, H. L., Lewison, R., An, L., & Shi, L. (2018). Mapping vegetation and land use types in Fanjingshan National Nature Reserve using google earth engine. *Remote Sensing*, 10(6). <https://doi.org/10.3390/rs10060927>
- USAID. (2015). *Burma Floods Fact Sheet*. Retrieved from https://www.usaid.gov/sites/default/files/documents/1866/burma_fl_fs01_08-06-2015.pdf
- Usamah, M., & Arambepola, N. (2013). Lessons Learned from the 2006 Flashfloods and Landslide in Uttaradit and Sukhothai Provinces: Implication for Effective Landslide Disaster Risk Mangement in Thailand. *Landslide Science and Practice*, 6(May 2011), 693–699. <https://doi.org/10.1007/978-3-642-31310-3>
- Van Westen, C., Jetten, V., Trigg, M., & Alkema, D. (2015). *Report on methodologies for hazard mapping in the Caribbean*. Retrieved from [http://www.charim.net/sites/default/files/handbook/otherpages/CHARIM Methodology Report.pdf](http://www.charim.net/sites/default/files/handbook/otherpages/CHARIM%20Methodology%20Report.pdf)
- Vegetation Subcommittee. (1997). *Vegetation Classification Standard*. Reston. Retrieved from <http://www.fgdc.gov/fgdc.html>
- Vermote, E., Justice, C., Claverie, M., & Franch, B. (2016). Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sensing of Environment*, 185, 46–56. <https://doi.org/10.1016/J.RSE.2016.04.008>
- Veselsky, M., Bandura, P., Burian, L., Harcinikova, T., & Bella, P. (2015). Semi-automated recognition of planation surfaces and other flat landforms: A case study from the Aggtelek Karst, Hungary. *Open Geosciences*, 7(1), 799–811. <https://doi.org/10.1515/geo-2015-0063>
- Wardlow, B. D., Egbert, S. L., & Kastens, J. H. (2007). Analysis of Time-Series MODIS 250 m Vegetation Index Data for Crop Classification in the U.S. Central Great Plains "Analysis of Time-Series MODIS 250 m Vegetation Index Data for Crop Classification in the U.S. Central Great Plains. *Remote Sensing of Environment*, 108(3), 290–310. <https://doi.org/10.1016/j.rse.2006.11.021>
- Wijaya, K., Nishimura, T., Setiawan, B. I., & Saptomo, S. K. (2009). Spatial variability of soil saturated hydraulic conductivity in paddy field in accordance to subsurface percolation. *Paddy and Water Environment*, 8(2), 113–120. <https://doi.org/10.1007/s10333-009-0190-x>
- Williams, R. D., Ahuja, L. R., & Naney, J. W. (1992). Comparison of Methods to Estimate Soil Water Characteristics from Soil Texture, Bulk Density, and Limited Data. *Soil Science*, 153(3), 172–184. Retrieved from <https://insights-ovid-com.ezproxy2.utwente.nl/crossref?an=00010694-199203000-00002>
- Wösten, J. H. M., Lilly, A., Nemes, A., & Le Bas, C. (1999). Development and use of a database of hydraulic properties of European soils. *Geoderma*, 90(3–4), 169–185. [https://doi.org/10.1016/S0016-7061\(98\)00132-3](https://doi.org/10.1016/S0016-7061(98)00132-3)
- Xu, H. (2008). A new index for delineating built-up land features in satellite imagery. *International Journal of Remote Sensing*, 29(14), 4269–4276. <https://doi.org/10.1080/01431160802039957>
- Xu, Hanqiu. (2006). Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery. *International Journal of Remote Sensing*, 27(14), 3025–3033. <https://doi.org/10.1080/01431160600589179>
- Yen, B. C., & Chow, V. T. (1980). Design Hyetographs for Small Drainage Structures. *Journal of the Hydraulic Division*, 1055–1076.
- Young, N. E., Anderson, R. S., Chignell, S. M., Vorster, A. G., Lawrence, R., & Evangelista, P. H. (2017). A survival guide to Landsat preprocessing. *Ecology*, 98(4), 920–932. <https://doi.org/10.1002/ecy.1730>
- Zha, Y., Gao, J., & Ni, S. (2003). Use of normalized difference built-up index in automatically mapping urban areas from TM imagery. *International Journal of Remote Sensing*, 24(3), 583–594. <https://doi.org/10.1080/01431160210144570>
- Zhu, Z., Gallant, A. L., Woodcock, C. E., Pengra, B., Olofsson, P., Loveland, T. R., ... Auch, R. F. (2016). Optimizing selection of training and auxiliary data for operational land cover classification for the LCMAP initiative. *ISPRS Journal of Photogrammetry and Remote Sensing*, 122, 206–221. <https://doi.org/10.1016/J.ISPRSJPRS.2016.11.004>

- Zhu, Z., & Woodcock, C. E. (2011). Object-based cloud and cloud shadow detection in Landsat imagery. *Remote Sensing of Environment*, 118, 83–94. <https://doi.org/10.1016/j.rse.2011.10.028>
- Zimmermann, B., Elsenbeer, H., & De Moraes, J. M. (2006). The influence of land-use changes on soil hydraulic properties: Implications for runoff generation. *Forest Ecology and Management*, 222(1–3), 29–38. <https://doi.org/10.1016/J.FORECO.2005.10.070>

ANNEXES

Annex 2. Pedotransfer script by Jetten and Shrestha (2018) based on the equations of Saxton et al. (2006)

```

# Model: Saxtons pedotransfer function SWAP model 2005 # #
# input data SOILGRIDS.ORG #
# Date: 05/04/2018 #
# Version: 1.1 #
# Author: V Jetten @ ITC #
#####
# $1 = soilgrids layer indication, e.g. "sl2"
# $2 is lsem layer, 1 or 2
# $3 is the degree of saturation between porosity and field capacity,
# used for the initial moisture content and initial suction head
# $4 is the bulk density you consider normal (uncompacted and not loose) in the area in kg/m3
#! ;; sl2 1 0.7
binding
S = SNDPPT_M_sl2_250m.tif; #sand %
C = CLYPPT_M_sl2_250m.tif; #clay %
# OC = oc$1.map; #organic carbon in %
Gravel = CRFVOL_M_sl2_250m.tif; #coarse fragments %, note in excel sheet it says g/cc
# but this is not correct, it is used as a volume fraction
OC = ORCDRC_M_sl2_250m.tif;
bdsg=BLDFIE_M_sl2_250m.tif; #bulkdensity kg/m3
standardBD = scalar(1470); # <= used to calibrate output 1470
fractionmoisture = scalar(0.7);
#inital moisture as fraction between porosity and field capacity
# 0 = init moist is at FC, 1.0 = init moist is at porosity
POROSITY = pore.map; #porosity (cm3/cm3)
Ksat = ksat.map; # ksat in mm/h
initmoist =thetainit.map; # inital moisture (cm3/cm3)
psi=psi1.map; # suction with init moisture in cm, used in LISEM
se = se.map; # relative moisture content between 0-1
Densityfactor = densfact.map;
BD = bulkdens.map; # ton/m3
WP = wilting.map; # wilting point moisture content
FC = fieldcap.map; # field capacity moisture content
PAW = plantAVW.map;
sand = sand.map;
clay = clay.map;
grav = graveln.map;
dem = dem30.map;
initial
# prep data
S = S/100;
C = C/100;
OC= (OC/1000)*100; # conversion OC from g/kg to percentage
OM = OC*1.72; #conversion org carbon to org matter
report om.map = OM;
Gravel = Gravel/100;
report Densityfactor = bdsg/standardBD;#scalar(1.0); # upper boundary 1.15 standardBD
# calculated as the bulk density from soilgrids divided by some standard bd
# multiple regression from excel
# wilting point stuff
M1500 =-0.024*S+0.487*C+0.006*OM+0.005*S*OM-0.013*C*OM+0.068*S*C+0.031; #W18)
# =-0.024*F18+0.487*G18+0.006*H18+0.005*F18*H18-0.013*G18*H18+0.068*F18*G18+0.031
M1500adj =M1500+0.14*M1500-0.02; #X18) =W18+0.14*W18-0.02

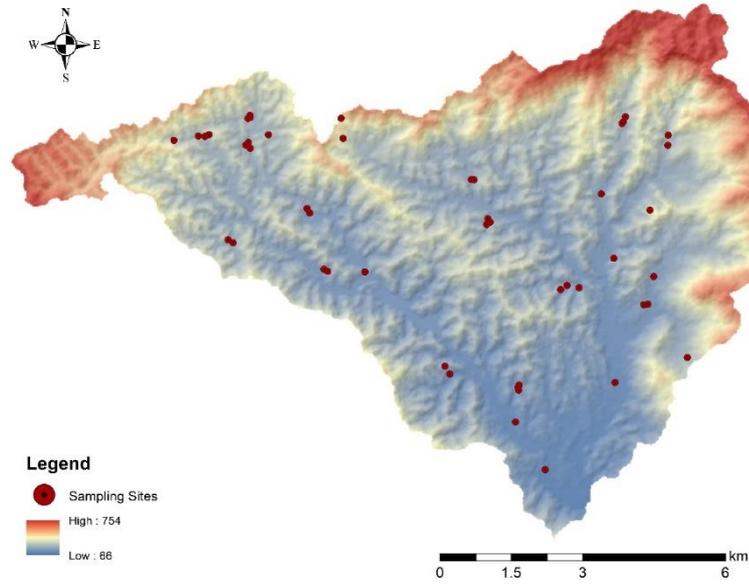
```

```

# field capacity stuff
M33 = -0.251*S+0.195*C+0.011*OM+0.006*S*OM-0.027*C*OM+0.452*S*C+0.299; #Y18)
#=-0.251*F18+0.195*G18+0.011*H18+0.006*F18*H18-0.027*G18*H18+0.452*F18*G18+0.299
M33adj = M33+(1.283*M33*M33-0.374*M33-0.015); #Z18) =Y18+(1.283*Y18*Y18-0.374*Y18-0.015)
# porosity - FC
PM33 = 0.278*S+0.034*C+0.022*OM-0.018*S*OM-0.027*C*OM-0.584*S*C+0.078; #AA18)
#=-0.278*F18+0.034*G18+0.022*H18-0.018*F18*H18-0.027*G18*H18-0.584*F18*G18+0.078
PM33adj = PM33+(0.636*PM33-0.107); #AB18) =AA18+(0.636*AA18-0.107)
# porosity
SatPM33 = M33adj + PM33adj; #AC18) =AB18+Z18
SatSadj = -0.097*S+0.043; #AD18) =-0.097*F18+0.043
SadjSat = SatPM33 + SatSadj; #AE18) =AC18+AD18
Dens_om = (1-SadjSat)*2.65; #AF18) =(1-AE18)*2.65
Dens_comp = Dens_om * Densityfactor; #AG18) =AF18*(I18)
PORE_comp = (1-Dens_om/2.65)-(1-Dens_comp/2.65); #AI18) =(1-AG18/2.65)-(1-AF18/2.65)
M33comp = M33adj + 0.2*PORE_comp; #AJ18) =Z18+0.2*AI18
#output
report POROSITY = 1-(Dens_comp/2.65); #AH18)
PoreMcomp = POROSITY-M33comp; #AK18)
LAMBDA = (ln(M33comp)-ln(M1500adj))/(ln(1500)-ln(33)); #AL18)
GravelRedKsat = (1-Gravel)/(1-Gravel*(1-1.5*(Dens_comp/2.65))); #AM18)
report Ksat = max(0.0, 1930*(PoreMcomp)**(3-LAMBDA)*GravelRedKsat); #AN18)
report BD = Gravel*2.65+(1-Gravel)*Dens_comp; #U18)
report WP = M1500adj;
report FC = M33adj;
report PAW = (M33adj - M1500adj)*(1-Gravel);
report initmoist= fractionmoisture*POROSITY+ (1-fractionmoisture)*FC;
# A = exp[ln(33) + B ln(T33)]
# B = [ln(1500) - ln(33)] / [ln(T33) - ln(T1500)]
bB = (ln(1500)-ln(33))/(ln(FC)-ln(WP));
aA = exp(ln(33)+bB*ln(FC));
report psi= aA * initmoist**-bB *100/9.8;
report se.map = initmoist/POROSITY;

```

Annex 2.1. Soil sampling locations Ban Da Na Kham watershed



Annex 2.2. Morphological properties of the soil for each site (field measurements)

Site	Longitude	Latitude	Altitude	Soil Depth (cm)	Soil Colour	Clay (%)	Sand (%)	Silt (%)	Soil Textural Class
1	619771	1967551	204	80	Reddish Grey	22	36	42	Loam
2	618856	1968672	168	90	Very Pale Brown	17	62	22	Sandy Loam
3	618942	1968680	158	-	Brown	27	40	33	Sandy Loam
4	618249	1967020	123	100	Light Brown	10	63	27	Sandy Loam
5	618225	1969653	154	100	Very Pale Brown	37	19	44	Loam
6	618984	1970672	251	80	Reddish Yellow	24	25	51	Loam
7	617962	1971016	170	100	Pinkish Grey	22	40	37	Silty Loam
8	617491	1969031	154	100	Yellow	19	40	41	Loam
9	617105	1968991	206	100	Brownish Yellow	42	14	44	Sandy Loam
10	617240	1969081	168	100	Light Grey	14	57	28	Clay Loam
11	616160	1966187	129	100	Light Yellowish Brown	2	60	38	Sandy Loam
12	614673	1967368	206	90	Light Yellowish Brown	24	27	49	Sandy Loam
13	614773	1967203	144	100	Pale Brown	23	37	40	Sandy Loam
14	612130	1969421	163	100	Brownish Yellow	31	18	51	Clay Loam
15	612205	1969377	168	100	Very Pale Brown	20	43	37	Clay Loam
16	616224	1966862	119	100	Light Grey	42	20	38	Sandy Loam
17	616212	1966923	137	100	Very Pale Brown	25	34	41	Silty Loam
18	616236	1966966	150	100	Very Pale Brown	11	50	38	Loam
19	610112	1970044	227	100	Yellow	31	20	49	Silty Loam
20	610574	1972634	228	100	Pale Brown	20	37	43	Sandy Loam
21	610571	1972681	230	100	Very Pale Brown	27	23	50	Sandy Loam
22	610531	1972616	214	80	Light Grey	20	42	38	Loam
23	609624	1972236	278	100	Brownish Yellow	28	29	43	Clay Loam
24	609712	1972273	292	100	Light Yellowish Brown	23	31	47	Loam
25	609486	1972244	224	60	Pale Brown	16	54	30	Clay Loam
26	610540	1972106	200	100	Light Yellowish Brown	19	29	52	Sandy Loam
27	610484	1972051	220	100	Very Pale Brown	27	15	59	Silty Loam
28	610577	1971984	259	100	Very Pale Brown	32	21	47	Loam
29	610219	1969983	211	62	Light Yellowish Brown	17	43	40	Loam
30	612991	1969365	139	100	Light Yellowish Brown	12	48	40	Loam
31	615578	1970494	175	100	Very Pale Brown	29	34	38	Loam
32	615625	1970421	207	100	Brownish Yellow	38	23	39	Silty Loam
33	615552	1970369	180	61	Very Pale Brown	37	23	40	Clay Loam
34	615229	1971325	169	25	Very Pale Brown	14	41	45	Loam
35	615286	1971318	177	100	Yellow	30	32	38	Loam
36	619063	1969265	162	100	Very Pale Brown	30	24	46	Clay Loam
37	619363	1972050	226	90	Greyish Brown	18	41	41	Loam
38	619368	1972261	162	100	Brown	11	59	30	Loam
39	618396	1972508	255	100	Yellow	2	61	37	Loam
40	618421	1972541	238	100	Yellow	6	54	40	Silty Clay
41	618471	1972653	217	100	Light Yellowish Brown	16	40	44	Clay Loam
42	612490	1972619	380	100	Light Yellowish Brown	5	45	50	Loam
43	612532	1972193	307	100	Pale Brown	2	58	40	Loam
44	610959	1972270	232	100	Light Yellowish Brown	2	53	45	Clay
45	611772	1970705	174	100	Light Yellowish Brown	2	49	49	Clay Loam
46	608972	1972152	233	75	Pale Brown	10	45	45	Loam
47	611828	1970614	160	100	Very Pale Brown	2	45	53	Sandy Loam
48	616782	1965176	110	100	Light Grey	4	58	38	Silty Loam

Annex 2.3. Physical and chemical properties of the soil for each site (field measurements)

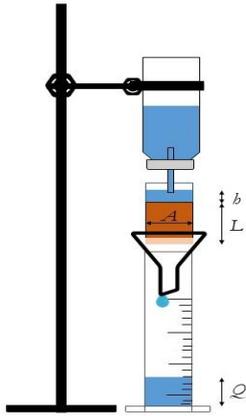
Site	Longitude	Latitude	Slope		K _s (mm h ⁻¹)	Porosity (%)	Bulk Density (g cm ⁻³)	OM (%)
			Position	Land Cover				
1	619771	1967551	Backslope	Orchard	306	50	1.32	3.7
2	618856	1968672	Summit	Cropland	2444	47	1.39	2.8
3	618942	1968680	Valley	Cropland	87	35	1.72	2.4
4	618249	1967020	Valley	Cropland	41	38	1.64	2.9
5	618225	1969653	Backslope	Orchard	81	51	1.29	2.2
6	618984	1970672	Backslope	Cropland	382	54	1.21	2.8
7	617962	1971016	Valley	Cropland	68	46	1.42	2.0
8	617491	1969031	Valley	Mixed Forest	417	55	1.20	2.8
9	617105	1968991	Summit	Mixed Forest	3703	70	0.80	6.2
10	617240	1969081	Valley	Orchard	978	45	1.45	2.9
11	616160	1966187	Valley	Teak Plantation	468	46	1.43	3.2
12	614673	1967368	Summit	Mixed Forest	191	60	1.07	4.8
13	614773	1967203	Backslope	Orchard	917	63	0.99	5.9
14	612130	1969421	Backslope	Mixed Forest	2657	55	1.20	5.6
15	612205	1969377	Valley	Cropland	560	54	1.22	4.2
16	616224	1966862	Valley	Mixed Forest	815	60	1.06	3.3
17	616212	1966923	Backslope	Mixed Forest	2078	63	0.98	4.6
18	616236	1966966	Backslope	Mixed Forest	204	50	1.32	5.1
19	610112	1970044	Backslope	Orchard	1069	59	1.09	3.5
20	610574	1972634	Backslope	Orchard	59	51	1.30	4.0
21	610571	1972681	Backslope	Orchard	905	60	1.07	3.7
22	610531	1972616	Valley	Orchard	1782	59	1.08	5.3
23	609624	1972236	Backslope	Orchard	436	56	1.18	3.8
24	609712	1972273	Summit	Orchard	1461	53	1.24	6.1
25	609486	1972244	Valley	Orchard	417	51	1.29	5.7
26	610540	1972106	Valley	Orchard	12	49	1.36	5.8
27	610484	1972051	Backslope	Orchard	322	59	1.09	6.7
28	610577	1971984	Summit	Orchard	382	61	1.03	3.9
29	610219	1969983	Valley	Orchard	7	52	1.27	4.7
30	612991	1969365	Valley	Orchard	726	54	1.22	3.6
31	615578	1970494	Backslope	Orchard	92	58	1.12	5.1
32	615625	1970421	Backslope	Mixed Forest	1986	76	0.64	8.8
33	615552	1970369	Valley	Orchard	62	59	1.07	4.9
34	615229	1971325	Valley	Orchard	229	48	1.39	3.6
35	615286	1971318	Backslope	Orchard	664	52	1.27	5.2
36	619063	1969265	Backslope	Orchard	5	41	1.58	2.7
37	619363	1972050	Valley	Orchard	153	48	1.37	6.0
38	619368	1972261	Backslope	Orchard	10784	64	0.95	6.5
39	618396	1972508	Summit	Orchard	244	57	1.15	4.7
40	618421	1972541	Backslope	Orchard	643	61	1.03	4.1
41	618471	1972653	Valley	Orchard	323	61	1.04	4.8
42	612490	1972619	Backslope	Teak Plantation	719	55	1.19	6.1
43	612532	1972193	Backslope	Orchard	0.1	54	1.21	4.6
44	610959	1972270	Backslope	Orchard	85	55	1.20	6.9
45	611772	1970705	Backslope	Orchard	218	63	0.97	6.4
46	608972	1972152	Backslope	Orchard	4	56	1.17	6.9
47	611828	1970614	Valley	Orchard	17	50	1.32	3.8
48	616782	1965176	Backslope	Orchard	16	43	1.50	3.9

K_s measurements framed in red were excluded from the soil analysis

Annex 2.4. Description of laboratory measurements for soil property determination

Saturated Hydraulic Conductivity

K_s was measured using the Constant Head method as described in the operating instructions for a laboratory-permeameter (Eijkelkamp, 2013). Since the laboratory equipment did not include a permeameter, an alternative apparatus had to be constructed. Before the measurements the samples were first soaked in water for 24 hours to reach the saturated state. To prevent disturbance the samples were covered on one site with a thin nylon membrane. Further, a plastic tape was attached to the upper half of the sampling ring so that around 3 – 4 cm of the tape protruded. The sample was then placed in the funnel and the protruding tape filled with water. During the measurement the water level above the sample was kept constant in order to have no changes in pressure. The measurement was considered as final when the recorded amount of water flowing through the sample reached a constant state. Finally, K_s was computed with equation 15.



$$K_s \text{ (mm h}^{-1}\text{)} = \frac{Q * L}{A * t * h} \quad (15)$$

Where Q = volume of water flowing through the sample, L = the length of soil sample, A = the cross-section surface of sample, t = the length of time lapse and h = the constant column height of the water above the sample.

Bulk Density and Porosity

Subsequent to the K_s measurement D_b and porosity were measured following the method introduced by Soil Survey Staff (2014). For D_b the core samples dried in an oven for 24 h at 105 °C. After removing from the oven and cooling for approximately 30 min the core samples were weighted. Equation 16 and equation 17 were used to calculate D_b and porosity, respectively.

$$D_b \text{ (g cm}^{-3}\text{)} = \frac{W_d}{V} \quad (16)$$

Where W_d = weight of dry soil (g) and V = volume of the sampling ring (cm³).

$$\text{Porosity (\%)} = 100 - \left(\frac{D_b}{P_d} * 100 \right) \quad (17)$$

Where P_d = Particle Density; an average value of 2.65 g cm⁻³ was used.

Particle Size Distribution (PSD)

PSD Analysis was used to assess the size distribution of the individual soil particles in each disturbed soil sample, in general terms, the percentage of silt, sand and clay particles. Using the Hydrometer Method as published by the Soil Science Society of America and outlined in the Soil Survey and Laboratory Manual (Soil Survey Staff, 2014), the PSD was assessed. The air-dried and grinded disturbed samples were passed through a 2 mm sieve. 40 g of each sample was mixed in a glass beaker with 100 ml of distilled water and 100 ml of a 5 % sodium hexametaphosphate solution (HMP). The HMP solution is applied to attain complete dispersal of the soil particles. The samples were left over night to soak and then stirred with an industrial mixer for 5 min. In a sedimentation cylinder which was filled up to the 1 L mark, the samples could equilibrate thermally. Parallel a sedimentation cylinder with a blank HMP solution was prepared for hydrometer reading correction computation. In a manual up-and-down motion, the samples were stirred again for 30 sec. Immediately, after stirring, time was recorded, temperature measured and a hydrometer (ASTM 152H) inserted. Hydrometer and thermometer readings were taken at 30 sec, 60 sec, 3 min, 10 min, 30 min, 60 min, 90 min, 120 min and 24 h for the soil and blank solution. Percentages of sand silt and clay were subsequently computed as suggested in the Soil Survey and Laboratory Manual (Soil Survey Staff, 2014).

Soil Organic Matter

For the determination of the SOM content of the disturbed surface samples the Loss-on-Ignition (LOI) method as set out by Schulte and Hopkins (1996) was applied. LOI was chosen as it represents an even more effective and simpler way to determine SOM compared to other conventional methods such as the Walkley and Black Carbon (WB-C) method which requires additional chemicals and laboratory facilities (Paramanathan et al., 2018). As pre-treatment the soil samples (5 g) were dried by 105 °C for 24 h in an oven to remove all available water. In a desiccator the samples cooled down and were weighted before going to the furnace. The furnace was pre-heated to have constant temperature of 400 °C. For 8 h the samples were exposed to the furnace before cooling down to room temperature in a desiccator again. The percentage of SOM was then determined by calculating the weight loss between oven-dry soil state and furnace ignition state.

Annex 2.5. Physical and chemical properties of SoilGrids for each site

Site	Slope Position	Land Cover	Clay (%)	Sand (%)	Silt (%)	K _s (mm h ⁻¹)	Porosity (%)	Bulk Density (g cm ⁻³)	OM (%)	Soil Textural Class
1	Backslope	Orchard	24	41	34	36	59	1.5	5.8	Loam
2	Summit	Cropland	26	40	33	32	59	1.5	6.2	Loam
3	Valley	Cropland	29	39	32	32	59	1.5	6.2	Clay Loam
4	Valley	Cropland	29	39	32	31	58	1.6	5.8	Clay Loam
5	Backslope	Orchard	25	41	34	23	55	1.5	4.6	Loam
6	Backslope	Cropland	22	43	35	57	65	1.5	8.3	Loam
7	Valley	Cropland	25	41	34	41	58	1.4	5.3	Loam
8	Valley	Mixed Forest	27	42	31	26	56	1.5	4.8	Loam
9	Summit	Mixed Forest	24	43	33	41	61	1.3	7.1	Loam
10	Valley	Orchard	30	38	32	28	57	1.4	5.7	Clay Loam
11	Valley	Teak Plantation	26	41	33	22	56	1.5	5.0	Loam
12	Summit	Mixed Forest	26	41	34	34	57	1.6	4.8	Loam
13	Backslope	Mixed Forest	28	42	31	29	57	1.5	5.2	Clay Loam
14	Backslope	Mixed Forest	28	42	30	36	58	1.5	5.2	Clay Loam
15	Valley	Cropland	28	42	31	36	56	1.6	4.6	Clay Loam
16	Valley	Mixed Forest	30	39	32	23	56	1.5	4.5	Clay Loam
17	Backslope	Mixed Forest	28	42	30	23	56	1.5	4.5	Clay Loam
18	Backslope	Mixed Forest	29	38	33	23	56	1.5	4.5	Clay Loam
19	Backslope	Orchard	25	46	30	37	59	1.6	6.2	Loam
20	Backslope	Orchard	24	43	33	36	60	1.6	6.7	Loam
21	Backslope	Orchard	24	43	33	45	60	1.4	6.7	Loam
22	Valley	Orchard	24	44	32	36	60	1.6	6.7	Loam
23	Backslope	Orchard	26	42	33	36	61	1.6	7.1	Loam
24	Summit	Orchard	26	42	32	41	61	1.4	7.2	Loam
25	Valley	Orchard	26	42	33	36	61	1.6	7.1	Loam
26	Valley	Orchard	27	41	32	34	59	1.5	6.5	Loam
27	Backslope	Orchard	27	41	32	34	59	1.5	6.5	Loam
28	Summit	Orchard	24	43	33	32	59	1.5	6.7	Loam
29	Valley	Orchard	26	44	30	34	58	1.5	5.7	Loam
30	Valley	Orchard	24	41	34	35	58	1.5	5.5	Loam
31	Backslope	Orchard	25	40	34	27	55	1.5	4.5	Loam
32	Backslope	Orchard	25	43	32	27	55	1.5	4.5	Loam
33	Valley	Orchard	25	40	34	37	59	1.5	5.7	Loam
34	Valley	Orchard	27	40	34	42	60	1.4	6.5	Loam
35	Backslope	Orchard	25	42	32	42	60	1.4	6.5	Loam
36	Backslope	Orchard	25	42	32	37	61	1.4	6.9	Loam
37	Valley	Orchard	28	39	33	40	62	1.4	7.4	Clay Loam
38	Backslope	Orchard	26	40	34	41	62	1.5	7.1	Loam
39	Summit	Orchard	27	39	34	41	61	1.5	6.7	Loam
40	Backslope	Orchard	23	44	34	41	61	1.5	6.7	Loam
41	Valley	Orchard	23	44	34	41	61	1.5	6.7	Loam
42	Backslope	Teak Plantation	30	37	33	22	56	1.5	4.6	Clay Loam
43	Backslope	Orchard	25	41	34	28	56	1.6	5.0	Loam
44	Backslope	Orchard	28	41	30	31	59	1.6	6.4	Clay Loam
45	Backslope	Orchard	27	41	32	30	57	1.5	5.0	Loam
46	Backslope	Orchard	25	43	31	35	60	1.4	6.5	Loam
47	Valley	Orchard	26	41	33	39	57	1.4	4.8	Loam
48	Backslope	Orchard	29	39	32	17	53	1.6	3.6	Clay Loam

Annex 3. Description of Landsat images used for yearly and seasonal composites

Median composite 2005/06			
ID	Date	Satellite	
LANDSAT/LT05/C01/T1_SR/LT05_130048_20050131	31/01/2005	Landsat 5	
LANDSAT/LT05/C01/T1_SR/LT05_130048_20050216	16/02/2005		
LANDSAT/LT05/C01/T1_SR/LT05_130048_20050421	21/04/2005		
LANDSAT/LT05/C01/T1_SR/LT05_130048_20060203	03/02/2006		
LANDSAT/LT05/C01/T1_SR/LT05_130048_20060307	07/03/2006		

Median composite dry cool season			
Year	ID	Date	Satellite
2018	LANDSAT/LC08/C01/T1_SR/LC08_130048_20180119	19/01/2018	Landsat 8
	LANDSAT/LC08/C01/T1_SR/LC08_130048_20180204	04/02/2018	
	LANDSAT/LC08/C01/T1_SR/LC08_130048_20180220	20/02/2018	
	LANDSAT/LC08/C01/T1_SR/LC08_130048_20181103	03/11/2018	
	LANDSAT/LC08/C01/T1_SR/LC08_130048_20181119	19/11/2018	
	LANDSAT/LC08/C01/T1_SR/LC08_130048_20181205	05/12/2018	
	LANDSAT/LC08/C01/T1_SR/LC08_130048_20181221	21/12/2018	
2017	LANDSAT/LC08/C01/T1_SR/LC08_130048_20190207	07/02/2019	Landsat 8
	LANDSAT/LC08/C01/T1_SR/LC08_130048_20170116	16/01/2017	
	LANDSAT/LC08/C01/T1_SR/LC08_130048_20170201	01/02/2017	
	LANDSAT/LC08/C01/T1_SR/LC08_130048_20170217	17/02/2017	
	LANDSAT/LC08/C01/T1_SR/LC08_130048_20171202	02/12/2017	
	LANDSAT/LC08/C01/T1_SR/LC08_130048_20171218	18/12/2017	
	LANDSAT/LC08/C01/T1_SR/LC08_130048_20180119	19/01/2018	
	LANDSAT/LC08/C01/T1_SR/LC08_130048_20180204	04/02/2018	
	LANDSAT/LC08/C01/T1_SR/LC08_130048_20180220	20/02/2018	
	LANDSAT/LC08/C01/T1_SR/LC08_130048_20181103	03/11/2018	
	LANDSAT/LC08/C01/T1_SR/LC08_130048_20181119	19/11/2018	
LANDSAT/LC08/C01/T1_SR/LC08_130048_20181205	05/12/2018		
LANDSAT/LC08/C01/T1_SR/LC08_130048_20181221	21/12/2018		

Median composite rainy season			
Year	ID	Date	Satellite
2018	LANDSAT/LC08/C01/T1_SR/LC08_130048_20180916	16/09/2018	Landsat 8
	LANDSAT/LC08/C01/T1_SR/LC08_130048_20181018	18/10/2018	
2017	LANDSAT/LC08/C01/T1_SR/LC08_130048_20170913	13/09/2017	Landsat 8
	LANDSAT/LC08/C01/T1_SR/LC08_130048_20171031	31/10/2017	

Median composite dry hot season

Year	ID	Date	Satellite
2018	LANDSAT/LC08/C01/T1_SR/LC08_130048_20180308	08/03/2018	Landsat 8
	LANDSAT/LC08/C01/T1_SR/LC08_130048_20180409	09/04/2018	
	LANDSAT/LC08/C01/T1_SR/LC08_130048_20180425	25/04/2018	
2017	LANDSAT/LC08/C01/T1_SR/LC08_130048_20170305	05/03/2017	Landsat 8
	LANDSAT/LC08/C01/T1_SR/LC08_130048_20170321	21/03/2017	
	LANDSAT/LC08/C01/T1_SR/LC08_130048_20170406	06/04/2017	
	LANDSAT/LC08/C01/T1_SR/LC08_130048_20170422	22/04/2017	
	LANDSAT/LC08/C01/T1_SR/LC08_130048_20170508	08/05/2017	
	LANDSAT/LC08/C01/T1_SR/LC08_130048_20170625	25/06/2017	

Annex 3.1. Google Earth Engine code for cloud removal by Saah et al. (2019)

```

// Compute a cloud score and adds a band that represents the cloud mask.
// This expects the input image to have the common band names:
// ["red", "blue", etc], so it can work across sensors.
exports.landsatCloudScore = function(ls8,cloudScoreThresh,cloudScorePctl,contractPixels,dilatePixels) {
  function getCloudScore (img) {
    // Compute several indicators of cloudiness and take the minimum of them.
    var score = ee.Image(1.0);
    // Clouds are reasonably bright in the blue band.
    score = score.min(rescale(img, 'img.blue', [0.1, 0.3]));
    // Clouds are reasonably bright in all visible bands.
    score = score.min(rescale(img, 'img.red + img.green + img.blue', [0.2, 0.8]));
    // Clouds are reasonably bright in all infrared bands.
    score = score.min(
      rescale(img, 'img.nir + img.swir1 + img.swir2', [0.3, 0.8]));
    // Clouds are reasonably cool in temperature.
    score = score.min(rescale(img, 'img.temp', [300, 290]));
    // However, clouds are not snow.
    var ndsi = img.normalizedDifference(['green', 'swir1']);
    score = score.min(rescale(ndsi, 'img', [0.8, 0.6]));
    score = score.multiply(100).byte();
    score = score.clamp(0,100);
    return img.addBands(score.rename(['cloudScore']));
  }
  function maskScore(img) {
    var cloudMask =
img.select(['cloudScore']).lt(cloudScoreThresh).focal_max(contractPixels).focal_min(dilatePixels).rename('cloudMask');
    return img.updateMask(cloudMask).addBands(cloudMask);
  }
  ls8 = ls8.map(getCloudScore);
  // Find low cloud score pctl for each pixel to avoid comission errors
  var minCloudScore = ls8.select(['cloudScore']).reduce(ee.Reducer.percentile(['cloudScorePctl']));
  ls8 = ls8.map(maskScore);
return ls8
};
exports.QAMaskCloud = function(ls8) {
// Functions for applying fmask to SR data with QA band
var fmaskBitDict = {'cloud': 32, 'shadow': 8, 'snow':16};
function cFmask(img, fmaskClass) {
  var m = img.select('pixel_qa').bitwiseAnd(fmaskBitDict[fmaskClass]).neq(0);
  return img.updateMask(m.not());
}
function cFmaskCloud(img) {
  return cFmask(img, 'cloud');
}
function cFmaskCloudShadow(img) {
  return cFmask(img, 'shadow');
}
ls8 = ls8.map(cFmaskCloud).map(cFmaskCloudShadow)
return ls8
};

```

Annex 3.2. Google Earth Engine code for cloud shadow removal by Saah et al. (2019)

```
// Function for finding dark outliers in time series.
// Original concept written by Carson Stam and adapted by Ian Housman.
// Adds a band that is a mask of pixels that are dark, and dark outliers.
exports.shadowMask =
function(collection,studyArea,zScoreThresh,shadowSumThresh,contractPixels,dilatePixels) {
  var shadowSumBands = ['nir','swir1'];
  var allCollection = collection.filterBounds(studyArea).select(shadowSumBands);
  // Get some pixel-wise stats for the time series
  var irStdDev = allCollection.select(shadowSumBands).reduce(ee.Reducer.stdDev());
  var irMean = allCollection.select(shadowSumBands).mean();
  var maskDarkOutliers = function(img) {
    var zScore = img.select(shadowSumBands).subtract(irMean).divide(irStdDev);
    var irSum = img.select(shadowSumBands).reduce(ee.Reducer.sum());
    var TDOMMask =
zScore.lt(zScoreThresh).reduce(ee.Reducer.sum()).eq(2).and(irSum.lt(shadowSumThresh));
    TDOMMask =
TDOMMask.focal_min(contractPixels).focal_max(dilatePixels).rename("TDOMMask");
    return img.updateMask(TDOMMask.not()).addBands(TDOMMask);
  };
  // Mask out dark dark outliers
  collection = collection.map(maskDarkOutliers)
  return collection;
};
```

Annex 3.3. Google Earth Engine code for BRDF correction by Saah et al. (2019)

```

var PI = ee.Number(3.14159265359);
var MAX_SATELLITE_ZENITH = 7.5;
var MAX_DISTANCE = 1000000;
var UPPER_LEFT = 0;
var LOWER_LEFT = 1;
var LOWER_RIGHT = 2;
var UPPER_RIGHT = 3;
exports.brdfS2 = function(collection) {
  collection = collection.map(applyBRDF);
  return collection;
  function applyBRDF(image) {
    var date = image.date();
    var footprint = ee.List(image.geometry().bounds().bounds().coordinates().get(0));
    var angles = getsunAngles(date, footprint);
    var sunAz = angles[0];
    var sunZen = angles[1];
    var viewAz = azimuth(footprint);
    var viewZen = zenith(footprint);
    var kval = _kvol(sunAz, sunZen, viewAz, viewZen);
    var kval0 = kval[0];
    var kval1 = kval[1];
    var result = _apply(image, kval.multiply(PI), kval0.multiply(PI));
    return result;
  }
  /* Get sunAngles from the map given the data. */
  function getsunAngles(date, footprint) {
    var jdp = date.getFraction('year');
    var seconds_in_hour = 3600;
    var hourGMT = ee.Number(date.getRelative('second', 'day')).divide(seconds_in_hour);
    var latRad = ee.Image.pixelLonLat().select('latitude').multiply(PI.divide(180));
    var longDeg = ee.Image.pixelLonLat().select('longitude');
    // Julian day proportion in radians
    var jdpr = jdp.multiply(PI).multiply(2);
    var a = ee.List([0.000075, 0.001868, 0.032077, 0.014615, 0.040849]);
    var meanSolarTime = longDeg.divide(15.0).add(ee.Number(hourGMT));
    var localSolarDiff1 = value(a, 0)
      .add(value(a, 1).multiply(jdpr.cos()))
      .subtract(value(a, 2).multiply(jdpr.sin()))
      .subtract(value(a, 3).multiply(jdpr.multiply(2).cos()))
      .subtract(value(a, 4).multiply(jdpr.multiply(2).sin()));
    var localSolarDiff2 = localSolarDiff1.multiply(12 * 60);
    var localSolarDiff = localSolarDiff2.divide(PI);
    var trueSolarTime = meanSolarTime
      .add(localSolarDiff.divide(60)) .subtract(12.0);
  }
}

```

```

// Hour as an angle;
var ah = trueSolarTime.multiply(ee.Number(MAX_SATELLITE_ZENITH *
2).multiply(PI.divide(180))) ;
var b = ee.List([0.006918, 0.399912, 0.070257, 0.006758, 0.000907, 0.002697, 0.001480]);
var delta = value(b, 0)
    .subtract(value(b, 1).multiply(jdpr.cos()))
    .add(value(b, 2).multiply(jdpr.sin()))
    .subtract(value(b, 3).multiply(jdpr.multiply(2).cos()))
    .add(value(b, 4).multiply(jdpr.multiply(2).sin()))
    .subtract(value(b, 5).multiply(jdpr.multiply(3).cos()))
    .add(value(b, 6).multiply(jdpr.multiply(3).sin()));
var cosSunZen = latRad.sin().multiply(delta.sin())
    .add(latRad.cos().multiply(ah.cos()).multiply(delta.cos()));
var sunZen = cosSunZen.acos();
// sun azimuth from south, turning west
var sinSunAzSW = ah.sin().multiply(delta.cos()).divide(sunZen.sin());
sinSunAzSW = sinSunAzSW.clamp(-1.0, 1.0);
var cosSunAzSW = (latRad.cos().multiply(-1).multiply(delta.sin())
    .add(latRad.sin().multiply(delta.cos()).multiply(ah.cos())))
    .divide(sunZen.sin());
var sunAzSW = sinSunAzSW.asin();
sunAzSW = where(cosSunAzSW.lte(0), sunAzSW.multiply(-1).add(PI), sunAzSW);
sunAzSW = where(cosSunAzSW.gt(0).and(sinSunAzSW.lte(0)), sunAzSW.add(PI.multiply(2)),
sunAzSW);
var sunAz = sunAzSW.add(PI);
// # Keep within [0, 2pi] range
sunAz = where(sunAz.gt(PI.multiply(2)), sunAz.subtract(PI.multiply(2)), sunAz);
var footprint_polygon = ee.Geometry.Polygon(footprint);
sunAz = sunAz.clip(footprint_polygon);
sunAz = sunAz.rename(['sunAz']);
sunZen = sunZen.clip(footprint_polygon).rename(['sunZen']);
return [sunAz, sunZen];
}
/* Get azimuth. */
function azimuth(footprint) {
    function x(point) {return ee.Number(ee.List(point).get(0))}
    function y(point) {return ee.Number(ee.List(point).get(1))}
    var upperCenter = line_from_coords(footprint, UPPER_LEFT,
UPPER_RIGHT).centroid().coordinates();
    var lowerCenter = line_from_coords(footprint, LOWER_LEFT,
LOWER_RIGHT).centroid().coordinates();
    var slope =
((y(lowerCenter)).subtract(y(upperCenter))).divide((x(lowerCenter)).subtract(x(upperCenter)));
    var slopePerp = ee.Number(-1).divide(slope);
    var azimuthLeft = ee.Image(PI.divide(2).subtract((slopePerp).atan()));

```

```

    return azimuthLeft.rename(['viewAz']);
}
/* Get zenith. */
function zenith(footprint) {
    var leftLine = line_from_coords(footprint, UPPER_LEFT, LOWER_LEFT);
    var rightLine = line_from_coords(footprint, UPPER_RIGHT, LOWER_RIGHT);
    var leftDistance = ee.FeatureCollection(leftLine).distance(MAX_DISTANCE);
    var rightDistance = ee.FeatureCollection(rightLine).distance(MAX_DISTANCE);
    var viewZenith = rightDistance.multiply(ee.Number(MAX_SATELLITE_ZENITH * 2))
        .divide(rightDistance.add(leftDistance))
        .subtract(ee.Number(MAX_SATELLITE_ZENITH))
        .clip(ee.Geometry.Polygon(footprint))
        .rename(['viewZen']);
    return viewZenith.multiply(PI.divide(180));
}
/* apply function to all bands
function _apply(image, kvol, kvol0) {
    var f_iso = 0;
    var f_geo = 0;
    var f_vol = 0;
    var blue = _correct_band(image, 'blue', kvol, kvol0, f_iso=0.0774, f_geo=0.0079, f_vol=0.0372);
    var green = _correct_band(image, 'green', kvol, kvol0, f_iso=0.1306, f_geo=0.0178, f_vol=0.0580);
    var red = _correct_band(image, 'red', kvol, kvol0, f_iso=0.1690, f_geo=0.0227, f_vol=0.0574);
    var re1 = _correct_band(image, 're1', kvol, kvol0, f_iso=0.2085, f_geo=0.0256, f_vol=0.0845);
    var re2 = _correct_band(image, 're2', kvol, kvol0, f_iso=0.2316, f_geo=0.0273, f_vol=0.1003);
    var re3 = _correct_band(image, 're3', kvol, kvol0, f_iso=0.2599, f_geo=0.0294, f_vol=0.1197);
    var nir = _correct_band(image, 'nir', kvol, kvol0, f_iso=0.3093, f_geo=0.0330, f_vol=0.1535);
    var re4 = _correct_band(image, 're4', kvol, kvol0, f_iso=0.2907, f_geo=0.0410, f_vol=0.1611);
    var swir1 = _correct_band(image, 'swir1', kvol, kvol0, f_iso=0.3430, f_geo=0.0453, f_vol=0.1154);
    var swir2 = _correct_band(image, 'swir2', kvol, kvol0, f_iso=0.2658, f_geo=0.0387, f_vol=0.0639);
    var temp = image.select('temp');
    return image.select([]).addBands([blue, green, red, nir, re1, re2, re3, nir, re4, swir1, swir2, temp]);
}
/* correct band function
function _correct_band(image, band_name, kvol, kvol0, f_iso, f_geo, f_vol) {
//""fiso + fvol * kvol + fgeo * kgeo""
    var iso = ee.Image(f_iso);
    var geo = ee.Image(f_geo);
    var vol = ee.Image(f_vol);
    var pred = vol.multiply(kvol).add(geo.multiply(kvol)).add(iso).rename(['pred']);
    var pred0 = vol.multiply(kvol0).add(geo.multiply(kvol0)).add(iso).rename(['pred0']);
    var cfac = pred0.divide(pred).rename(['cfac']);
    var corr = image.select(band_name).multiply(cfac).rename([band_name]);
    return corr;
}

```

```

/* calculate kvol and kvol0 */
function _kvol(sunAz, sunZen, viewAz, viewZen) {
var relative_azimuth = sunAz.subtract(viewAz).rename(['relAz']);
var pa1 = viewZen.cos().multiply(sunZen.cos());
var pa2 = viewZen.sin().multiply(sunZen.sin()).multiply(relative_azimuth.cos());
var phase_angle1 = pa1.add(pa2);
var phase_angle = phase_angle1.acos();
var p1 = ee.Image(PI.divide(2)).subtract(phase_angle);
var p2 = p1.multiply(phase_angle1);
var p3 = p2.add(phase_angle.sin());
var p4 = sunZen.cos().add(viewZen.cos());
var p5 = ee.Image(PI.divide(4));
var kvol = p3.divide(p4).subtract(p5).rename(['kvol']);
var viewZen0 = ee.Image(0);
var pa10 = viewZen0.cos().multiply(sunZen.cos());
var pa20 = viewZen0.sin().multiply(sunZen.sin()).multiply(relative_azimuth.cos());
var phase_angle10 = pa10.add(pa20);
var phase_angle0 = phase_angle10.acos();
var p10 = ee.Image(PI.divide(2)).subtract(phase_angle0);
var p20 = p10.multiply(phase_angle10);
var p30 = p20.add(phase_angle0.sin());
var p40 = sunZen.cos().add(viewZen0.cos());
var p50 = ee.Image(PI.divide(4));
var kvol0 = p30.divide(p40).subtract(p50).rename(['kvol0']);
return [kvol, kvol0]}

/* helper function */
function line_from_coords(coordinates, fromIndex, toIndex) {
return ee.Geometry.LineString(ee.List([
coordinates.get(fromIndex),
coordinates.get(toIndex)]));
}
function where(condition, trueValue, falseValue) {
var trueMasked = trueValue.mask(condition);
var falseMasked = falseValue.mask(invertMask(condition));
return trueMasked.unmask(falseMasked);
}
function invertMask(mask) {
return mask.multiply(-1).add(1);
}
function value(list, index) {
return ee.Number(list.get(index));
}
};

```

Annex 3.4. Google Earth Engine code for topographic correction by Saah et al. (2019)

```

ar scale = 300;
var toaOrSR = 'SR';
// get terrain layers
var dem = ee.Image("USGS/SRTMGL1_003")
var degree2radian = 0.01745;
exports.terrainCorrection = function(collection) {
  function getTopo(img) {
    /* function to filter for areas with terrain and areas without */
    dem = dem.unmask(0);
    var geom = ee.Geometry(img.get('system:footprint')).bounds();
    print(geom)
    var slp_rad = ee.Terrain.slope(dem).clip(geom);
    var slope = slp_rad.reduceRegion({reducer: ee.Reducer.percentile([80]),geometry: geom,scale: 100});
    return img.set('slope',slope.get('slope'));
  }
  function pixelArea(img){
    /* check if there is data in the image */
    var geom = ee.Geometry(img.get('system:footprint')).bounds();
    var area = img.select(['red']).gt(0).reduceRegion({reducer:ee.Reducer.sum(),geometry: geom,scale:100});
    return img.set("pixelArea",area.get("red"))}
  collection = collection.map(pixelArea);
  collection = collection.filter(ee.Filter.gt("pixelArea",100));
  //collection = collection.map(getTopo);
  var correction = collection.filter(ee.Filter.gte("slope",10));
    var notcorrection = collection.filter(ee.Filter.lt("slope",10));
  collection = collection.map(illuminationCondition);
  collection = collection.map(illuminationCorrection);
  return(collection);
  // Extract image metadata about solar position
  var SZ_rad =
ee.Image.constant(ee.Number(img.get("SOLAR_ZENITH_ANGLE"))).multiply(3.14159265359).divide(18
0).clip(img.geometry().buffer(10000));
  var SA_rad =
ee.Image.constant(ee.Number(img.get("SOLAR_AZIMUTH_ANGLE"))).multiply(3.14159265359).divide(1
80).clip(img.geometry().buffer(10000));
  // Creat terrain layers
  var slp = ee.Terrain.slope(dem).clip(img.geometry().buffer(10000));
  var slp_rad =
ee.Terrain.slope(dem).multiply(3.14159265359).divide(180).clip(img.geometry().buffer(10000));
  var asp_rad =
ee.Terrain.aspect(dem).multiply(3.14159265359).divide(180).clip(img.geometry().buffer(10000));
  // Calculate the Illumination Condition (IC) slope part of the illumination condition
  var cosZ = SZ_rad.cos();

```

```

var cosS = slp_rad.cos();
var slope_illumination = cosS.expression("cosZ * cosS",
    {'cosZ': cosZ,
     'cosS': cosS.select('slope')});
// aspect part of the illumination condition
var sinZ = SZ_rad.sin();
var sinS = slp_rad.sin();
var cosAziDiff = (SA_rad.subtract(asp_rad)).cos();
var aspect_illumination = sinZ.expression("sinZ * sinS * cosAziDiff",
    {'sinZ': sinZ,
     'sinS': sinS,
     'cosAziDiff': cosAziDiff});
// full illumination condition (IC)
var ic = slope_illumination.add(aspect_illumination);
// Add IC to original image
var img_plus_ic =
ee.Image(img.addBands(ic.rename('IC')).addBands(cosZ.rename('cosZ')).addBands(cosS.rename('cosS')).ad
dBands(slp.rename('slope')));
return img_plus_ic;
}
// Function to apply the Sun-Canopy-Sensor + C (SCSc) correction method to each image
function illuminationCorrection(img) {
    var props = img.toDictionary();
    var st = img.get('system:time_start');
    var img_plus_ic = img;
    var mask1 = img_plus_ic.select('nir').gt(-0.1);
    var mask2 = img_plus_ic.select('slope').gte(5)
        .and(img_plus_ic.select('IC').gte(0))
        .and(img_plus_ic.select('nir').gt(-0.1));
    var img_plus_ic_mask2 = ee.Image(img_plus_ic.updateMask(mask2));
    // Specify Bands to topographically correct
    var bandList = ['blue','green','red','nir','swir1','swir2'];
    var compositeBands = img.bandNames();
    var nonCorrectBands = img.select(compositeBands.removeAll(bandList));
    var geom = ee.Geometry(img.get('system:footprint').bounds().buffer(10000));
    function apply_SCSccorr(band) {
        var method = 'SCSc';
        var out = img_plus_ic_mask2.select('IC', band).reduceRegion({
            reducer: ee.Reducer.linearFit(),
// Compute coefficients:
a(slope), b(offset), c(b/a) geometry: ee.Geometry(img.geometry()).buffer(-5000)),
        scale: 300,
        maxPixels: 1000000000
        });
    if (out === null || out === undefined) {

```

```
    return img_plus_ic_mask2.select(band);
  }
else{
  var out_a = ee.Number(out.get('scale'));
  var out_b = ee.Number(out.get('offset'));
  var out_c = out_b.divide(out_a);
  // Apply the SCS correction
  var SCS_output = img_plus_ic_mask2.expression(
    "((image * (cosB * cosZ + cvalue)) / (ic + cvalue))", {
    'image': img_plus_ic_mask2.select(band),
    'ic': img_plus_ic_mask2.select('IC'),
    'cosB': img_plus_ic_mask2.select('cosS'),
    'cosZ': img_plus_ic_mask2.select('cosZ'),
    'cvalue': out_c
  });
  return SCS_output;
}
}
}
var img_SCScorr = ee.Image(bandList.map(apply_SCScorr)).addBands(img_plus_ic.select('IC'));
var bandList_IC = ee.List([bandList, 'IC']).flatten();
img_SCScorr = img_SCScorr.unmask(img_plus_ic.select(bandList_IC)).select(bandList);
return img_SCScorr.addBands(nonCorrectBands)
  .setMulti(props)
  .set('system:time_start',st);
}
}
```

Annex 3.5. Google Earth Engine code to generate covariate layer (seasonal) adapted from Saah et al. (2019)

```

var elevation = ee.Image("USGS/SRTMGL1_003");
var jrcImage = ee.Image("JRC/GSW1_0/GlobalSurfaceWater");
var ndCovariatesList = [
  ['blue', 'green'],
  ['blue', 'red'],
  ['blue', 'nir'],
  ['blue', 'swir1'],
  ['blue', 'swir2'],
  ['green', 'red'],
  ['green', 'nir'],
  ['green', 'swir1'],
  ['green', 'swir2'],
  ['red', 'swir1'],
  ['red', 'swir2'],
  ['nir', 'red'],
  ['nir', 'swir1'],
  ['nir', 'swir2'],
  ['swir1', 'swir2']
];
var rCovariatesList = [
  ['swir1', 'nir'],
  ['red', 'swir1']
];
var ComputeNDCovariatesList = function (season) {
  var list = [];
  for (var index in ndCovariatesList) {
    var list_ = [season + '_' + ndCovariatesList[index][0], season + '_' + ndCovariatesList[index][1]];
    list.push(list_);
  }
  return list;
};
var addNDCovariates = function (season, image) {
  var list = ComputeNDCovariatesList(season);
  for (var index in list) {
    image = image.addBands(image.normalizedDifference(list[index]).rename(season + '_ND_' + ndCovariatesList[index][0] + '_' + ndCovariatesList[index][1]));
  }
  return image;
};
var ComputeRCovariatesList = function (season) {
  var list = [];
  for (var index in rCovariatesList) {
    var list_ = [season + '_' + rCovariatesList[index][0], season + '_' + rCovariatesList[index][1]];
    list.push(list_);
  }
  return list;
};
var addRCovariates = function (season, image) {
  var list = ComputeRCovariatesList(season);
  for (var index in list) {
    image = image.addBands(image.select(list[index][0]).divide(image.select(list[index][1]))
      .rename(season + '_R_' + rCovariatesList[index][0] + '_' + rCovariatesList[index][1]));
  }
}

```

```

return image;
};
//Enhanced Vegetation Index (EVI)//
var addEVI = function (season, image) {
  var evi = image.expression('2.5 * ((NIR - RED) / (NIR + 6 * RED - 7.5 * BLUE + 1))', {
    'NIR': image.select(season + '_nir'),
    'RED': image.select(season + '_red'),
    'BLUE': image.select(season + '_blue')
  }).float();
  return image.addBands(evi.rename(season + '_EVI'));
};
//Enhanced Built-Up and Bareness Index (EBBI)//
var addEBBI = function (season, image) {
  print(image);
  var ebbi= image.expression('(swir-nir) / 10*sqrt(swir+temp)',{
    'swir': image.select(season + '_swir1'),
    'nir' : image.select(season + '_nir'),
    'temp': image.select(season + '_temp'),
  });
  return image.addBands(ebbi.rename(season + '_EBBI'));
};
//Normalized Burn Index (NBR)//
var addNBR = function (season, image) {
  var nbr = image.normalizedDifference([season + '_nir', season + '_swir2']).float();
  return image.addBands(nbr.rename(season + '_NBR'));
};
//Normalized Difference Vegetation Index//
var addNDVI = function(season, image) {
  var NDVI =image.normalizedDifference([season + '_red', season + '_nir']).float();
  return image.addBands(NDVI.rename(season + '_NDVI'));
};
//Soil Adjusted Vegetation Index (SAVI)//
var addSAVI = function (season, image) {
  // Add Soil Adjust Vegetation Index (SAVI)
  // using L = 0.5;
  var savi = image.expression('(NIR - RED) * (1 + 0.5)/(NIR + RED + 0.5)', {
    'NIR': image.select(season + '_nir'),
    'RED': image.select(season + '_red')
  }).float();
  return image.addBands(savi.rename(season + '_SAVI'));
};
// Index-based Built-up Index//
var addIBI = function (season, image) {
  var ibiA = image.expression('2 * SWIR1 / (SWIR1 + NIR)', {
    'SWIR1': image.select(season + '_swir1'),
    'NIR' : image.select(season + '_nir')
  }).rename(['IBI_A']);
  var ibiB = image.expression('(NIR / (NIR + RED)) + (GREEN / (GREEN + SWIR1))', {
    'NIR' : image.select(season + '_nir'),
    'RED' : image.select(season + '_red'),
    'GREEN': image.select(season + '_green'),
    'SWIR1': image.select(season + '_swir1')
  }).rename(['IBI_B']);
  var ibiAB = ibiA.addBands(ibiB);
  var ibi = ibiAB.normalizedDifference(['IBI_A', 'IBI_B']);

```

```

return image.addBands(ibi.rename([season + '_IBI']));
};
// Function to compute the Tasseled Cap transformation//
var getTassledCapComponents = function (season, image) {
  var coefficients = ee.Array([
    [0.3037, 0.2793, 0.4743, 0.5585, 0.5082, 0.1863],
    [-0.2848, -0.2435, -0.5436, 0.7243, 0.0840, -0.1800],
    [0.1509, 0.1973, 0.3279, 0.3406, -0.7112, -0.4572],
    [-0.8242, 0.0849, 0.4392, -0.0580, 0.2012, -0.2768],
    [-0.3280, 0.0549, 0.1075, 0.1855, -0.4357, 0.8085],
    [0.1084, -0.9022, 0.4120, 0.0573, -0.0251, 0.0238]
  ]);
  var bands = ee.List([season + '_blue', season + '_green', season + '_red', season + '_nir', season +
'_swir1', season + '_swir2']);
  var arrayImage1D = image.select(bands).toArray();
  var arrayImage2D = arrayImage1D.toArray(1);
  var componentsImage = ee.Image(coefficients).matrixMultiply(arrayImage2D).arrayProject([0])
    .arrayFlatten([[season + '_brightness', season + '_greenness', season + '_wetness', season +
'_fourth', season + '_fifth', season + '_sixth']]).float();
  return image.addBands(componentsImage);
};
var getTassledCapAngleAndDistance = function (season, image) {
  var brightness = image.select(season + '_brightness');
  var greenness = image.select(season + '_greenness');
  var wetness = image.select(season + '_wetness');
  var tcAngleBG = brightness.atan2(greenness).divide(Math.PI).rename([season + '_tcAngleBG']);
  var tcAngleGW = greenness.atan2(wetness).divide(Math.PI).rename([season + '_tcAngleGW']);
  var tcAngleBW = brightness.atan2(wetness).divide(Math.PI).rename([season + '_tcAngleBW']);
  var tcDistanceBG = brightness.hypot(greenness).rename([season + '_tcDistanceBG']);
  var tcDistanceGW = greenness.hypot(wetness).rename([season + '_tcDistanceGW']);
  var tcDistanceBW = brightness.hypot(wetness).rename([season + '_tcDistanceBW']);
  image =
image.addBands(tcAngleBG).addBands(tcAngleGW).addBands(tcAngleBW).addBands(tcDistanceBG).ad
dBands(tcDistanceGW).addBands(tcDistanceBW);
  return image;
};
var computeTassledCap = function (season, image) {
  image = getTassledCapComponents(season, image);
  image = getTassledCapAngleAndDistance(season, image);
  return image;
};
var addTopography = function (image) {
  // Calculate slope, aspect and hillshade
  var topo = ee.Algorithms.Terrain(elevation);
  // From aspect (a), calculate eastness (sin a), northness (cos a)
  var deg2rad = ee.Number(Math.PI).divide(180);
  var aspect = topo.select(['aspect']);
  var aspect_rad = aspect.multiply(deg2rad);
  var eastness = aspect_rad.sin().rename(['eastness']).float();
  var northness = aspect_rad.cos().rename(['northness']).float();
  topo = topo.select(['elevation','slope','aspect']).addBands(eastness).addBands(northness);
  image = image.addBands(topo);
  return image;
};
//JRC dataset//

```

```
var addJRCDataset = function (image) {
  // Update the mask.
  jrcImage = jrcImage.unmask(0);
  image = image.addBands(jrcImage.select(['occurrence']).rename(['occurrence']));
  image = image.addBands(jrcImage.select(['change_abs']).rename(['change_abs']));
  image = image.addBands(jrcImage.select(['change_norm']).rename(['change_norm']));
  image = image.addBands(jrcImage.select(['seasonality']).rename(['seasonality']));
  image = image.addBands(jrcImage.select(['transition']).rename(['transition']));
  image = image.addBands(jrcImage.select(['max_extent']).rename(['max_extent']));
  return image;
};
//Add covariates to season images//
var addCovariates = function (season, image) {
  image = addNDCovariates(season, image);
  image = addRCovariates(season, image);
  image = addNDVI(season, image);
  image = addEVI(season, image);
  image = addSAVI(season, image);
  image = addIBI(season, image);
  image = addEBBI(season, image);
  image = addNBR(season, image);
  image = computeTassledCap(season, image);
  return image;
};
var addJRCAndTopo = function (image) {
  image = addTopography(image);
  image = addJRCDataset(image);
  return image;
};
exports.addCovariates = addCovariates;
exports.addJRCAndTopo = addJRCAndTopo;
```

Annex 3.6. Google Earth Engine code to generate covariate layer (yearly) adapted from Saah et al. (2019)

```

ar elevation = ee.Image("USGS/SRTMGL1_003");
var jrcImage = ee.Image("JRC/GSW1_0/GlobalSurfaceWater");

var ndCovariatesList = [
  ['blue', 'green'],
  ['blue', 'red'],
  ['blue', 'nir'],
  ['blue', 'swir1'],
  ['blue', 'swir2'],
  ['green', 'red'],
  ['green', 'nir'],
  ['green', 'swir1'],
  ['green', 'swir2'],
  ['red', 'swir1'],
  ['red', 'swir2'],
  ['nir', 'red'],
  ['nir', 'swir1'],
  ['nir', 'swir2'],
  ['swir1', 'swir2']
];
var rCovariatesList = [
  ['swir1', 'nir'],
  ['red', 'swir1']
];
var ComputeNDCovariatesList = function () {
  var list = [];
  for (var index in ndCovariatesList) {
    var list_ = [ndCovariatesList[index][0], ndCovariatesList[index][1]];
    list.push(list_);
  }
  return list;
};
var addNDCovariates = function (image){
  var list = ComputeNDCovariatesList();
  print(list);
  for (var index in list) {
    image = image.addBands(image.normalizedDifference(list[index]).rename('ND_'+
    ndCovariatesList[index][0] + '_' + ndCovariatesList[index][1]));
  }
  return image;
};
var ComputeRCovariatesList = function () {
  var list = [];
  for (var index in rCovariatesList) {
    var list_ = [rCovariatesList[index][0], rCovariatesList[index][1]];
    list.push(list_);
  }
  return list;
};
var addRCovariates = function (image) {
  var list = ComputeRCovariatesList();
  for (var index in list) {
    image = image.addBands(image.select(list[index][0]).divide(image.select(list[index][1])))
  }
}

```

```

.rename('_R_' + rCovariatesList[index][0] + '_' + rCovariatesList[index][1]));
}
return image;
};
var addEVI = function (image) {
var evi = image.expression('2.5 * ((NIR - RED) / (NIR + 6 * RED - 7.5 * BLUE + 1))', {
'NIR' : image.select('nir'),
'RED' : image.select('red'),
'BLUE': image.select('blue')
}).float();
return image.addBands(evi.rename('EVI'));
}
var addEBBI = function (image) {
print(image);
var ebbi= image.expression("(swir-nir) / 10*sqrt(swir+temp)",
{'swir' : image.select("swir1"),
'nir' : image.select("nir"),
'temp' : image.select("temp")
});
return image.addBands(ebbi.rename('EBBI'));
};
var addNBR = function (image) {
var nbr = image.normalizedDifference(['nir','swir2']).float();
return image.addBands(nbr.rename('NBR'));
};
var addNDVI = function(image) {
var NDVI =image.normalizedDifference(['red','nir']).float();
return image.addBands(NDVI.rename('NDVI'));
};
var addSAVI = function (image) {
// Add Soil Adjust Vegetation Index (SAVI)
// using L = 0.5;
var savi = image.expression('(NIR - RED) * (1 + 0.5)/(NIR + RED + 0.5)', {
'NIR': image.select('nir'),
'RED': image.select('red')
}).float();
return image.addBands(savi.rename('SAVI'));
};
var addIBI = function (image) {
// Add Index-Based Built-Up Index (IBI)
var ibiA = image.expression('2 * SWIR1 / (SWIR1 + NIR)', {
'SWIR1': image.select('swir1'),
'NIR' : image.select('nir')
}).rename(['IBI_A']);
var ibiB = image.expression('(NIR / (NIR + RED)) + (GREEN / (GREEN + SWIR1))', {
'NIR' : image.select('nir'),
'RED' : image.select('red'),
'GREEN': image.select('green'),
'SWIR1': image.select('swir1')
}).rename(['IBI_B']);
var ibiAB = ibiA.addBands(ibiB);
var ibi = ibiAB.normalizedDifference(['IBI_A', 'IBI_B']);
return image.addBands(ibi.rename(['IBI']));
};

```

```

// Function to compute the Tasseled Cap transformation and return an image
var getTassledCapComponents = function (image) {
var coefficients = ee.Array([
[0.3037, 0.2793, 0.4743, 0.5585, 0.5082, 0.1863],
[-0.2848, -0.2435, -0.5436, 0.7243, 0.0840, -0.1800],
[0.1509, 0.1973, 0.3279, 0.3406, -0.7112, -0.4572],
[-0.8242, 0.0849, 0.4392, -0.0580, 0.2012, -0.2768],
[-0.3280, 0.0549, 0.1075, 0.1855, -0.4357, 0.8085],
[0.1084, -0.9022, 0.4120, 0.0573, -0.0251, 0.0238]
]);
var bands = ee.List(['blue', 'green', 'red', 'nir', 'swir1', 'swir2']);
// Make an Array Image, with a 1-D Array per pixel.
var arrayImage1D = image.select(bands).toArray();
// Make an Array Image with a 2-D Array per pixel, 6 x 1
var arrayImage2D = arrayImage1D.toArray(1);
var componentsImage = ee.Image(coefficients).matrixMultiply(arrayImage2D).arrayProject([0])
.arrayFlatten(['brightness', 'greenness', 'wetness', 'fourth', 'fifth', 'sixth']).float();
// Get a multi-band image with TC-named bands
return image.addBands(componentsImage);
};
// Function to add Tasseled Cap angles and distances to an image. Assumes image has bands: 'brightness',
'greenness', and 'wetness'.
var getTassledCapAngleAndDistance = function (image) {
var brightness = image.select('brightness');
var greenness = image.select('greenness');
var wetness = image.select('wetness');
// Calculate tassled cap angles and distances
var tcAngleBG = brightness.atan2(greenness).divide(Math.PI).rename(['tcAngleBG']);
var tcAngleGW = greenness.atan2(wetness).divide(Math.PI).rename(['tcAngleGW']);
var tcAngleBW = brightness.atan2(wetness).divide(Math.PI).rename(['tcAngleBW']);
var tcDistanceBG = brightness.hypot(greenness).rename(['tcDistanceBG']);
var tcDistanceGW = greenness.hypot(wetness).rename(['tcDistanceGW']);
var tcDistanceBW = brightness.hypot(wetness).rename(['tcDistanceBW']);
image =
image.addBands(tcAngleBG).addBands(tcAngleGW).addBands(tcAngleBW).addBands(tcDistanceBG).ad
dBands(tcDistanceGW).addBands(tcDistanceBW);
return image;
};

var computeTassledCap = function (image) {
image = getTassledCapComponents(image);
image = getTassledCapAngleAndDistance(image);
return image;
};
var addTopography = function (image) {
// Calculate slope, aspect and hillshade
var topo = ee.Algorithms.Terrain(elevation);
// From aspect (a), calculate eastness (sin a), northness (cos a)
var deg2rad = ee.Number(Math.PI).divide(180);
var aspect = topo.select(['aspect']);
var aspect_rad = aspect.multiply(deg2rad);
var eastness = aspect_rad.sin().rename(['eastness']).float();
var northness = aspect_rad.cos().rename(['northness']).float();
// Add topography bands to image
topo = topo.select(['elevation','slope','aspect']).addBands(eastness).addBands(northness);

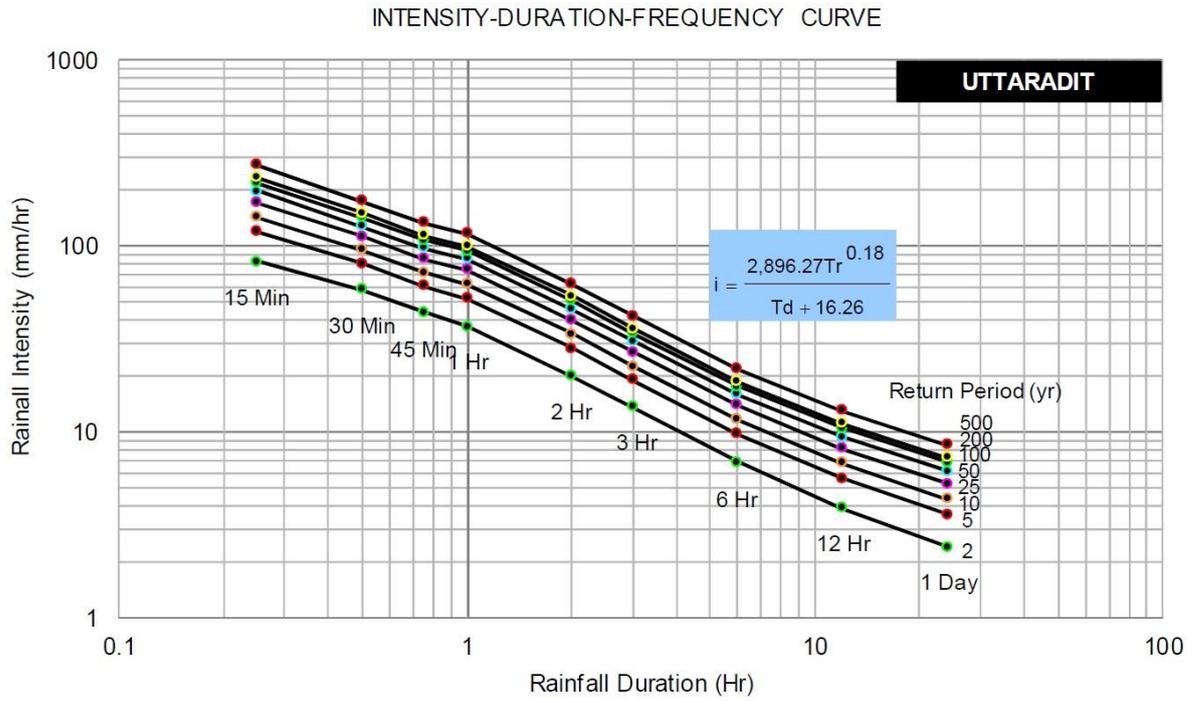
```

```
image = image.addBands(topo);
return image;
};
var addJRCDataset = function (image) {
// Update the mask.
jrcImage = jrcImage.unmask(0);
image = image.addBands(jrcImage.select(['occurrence']).rename(['occurrence']));
image = image.addBands(jrcImage.select(['change_abs']).rename(['change_abs']));
image = image.addBands(jrcImage.select(['change_norm']).rename(['change_norm']));
image = image.addBands(jrcImage.select(['seasonality']).rename(['seasonality']));
image = image.addBands(jrcImage.select(['transition']).rename(['transition']));
image = image.addBands(jrcImage.select(['max_extent']).rename(['max_extent']));
return image;
};
exports.addCovariates = function (image) {
image = addNDCovariates(image);
image = addNDVI(image);
image = addEVI(image);
image = addNBR(image);
image = addSAVI(image);
image = addIBI(image);
image = addEBBI(image);
image = computeTassledCap(image);
return image;
};
var addJRCAndTopo = function (image) {
image = addTopography(image);
image = addJRCDataset(image);
return image;
};
exports.addJRCAndTopo = addJRCAndTopo;
```

Annex 3.7. Selected covariates layers for the classifications of 2005 and 2018

Covariate layers 2005 (single composite)	Covariate layers 2018 (seasonal composites)
NBR ND_blue_green ND_blue_nir ND_blue_swir1 ND_green_nir ND_green_swir1 ND_green_swir2 ND_nir_swir2 ND_red_swir1 ND_swir1_swir2 brightness fifth fourth greenness nir swir1 tcAngleGW tcDistanceBG tcDistanceGW NBR ND_blue_green ND_blue_nir ND_blue_swir1	dryCool_ND_green_swir1 dryCool_ND_red_swir2 dryCool_brightness dryCool_cosS dryCool_green dryCool_nir dryCool_red dryCool_sixth dryCool_slope dryCool_tcDistanceBW dryHot_ND_blue_swir1 dryHot_ND_green_swir1 dryHot_ND_red_swir1 dryHot_ND_red_swir2", dryHot_brightness dryHot_nir dryHot_sixth dryHot_swir1 dryHot_tcDistanceBG dryHot_tcDistanceBW dryHot_tcDistanceGW dryHot_wetness elevation rainy_EVI rainy_ND_blue_green rainy_ND_blue_nir rainy_ND_blue_red rainy_ND_blue_swir1 rainy_ND_green_nir rainy_ND_green_red rainy_ND_green_swir1 rainy_ND_green_swir2 rainy_ND_nir_red rainy_ND_nir_swir2 rainy_ND_red_swir1 rainy_ND_red_swir2 rainy_ND_swir1_swir2 rainy_R_red_swir1 rainy_SAVI rainy_cosS rainy_fifth rainy_fourth rainy_greenness rainy_red rainy_sixth rainy_swir1 rainy_tcAngleBG rainy_temp

Annex 4. IDF curves for Uttaradit province (Rittima et al., 2013)

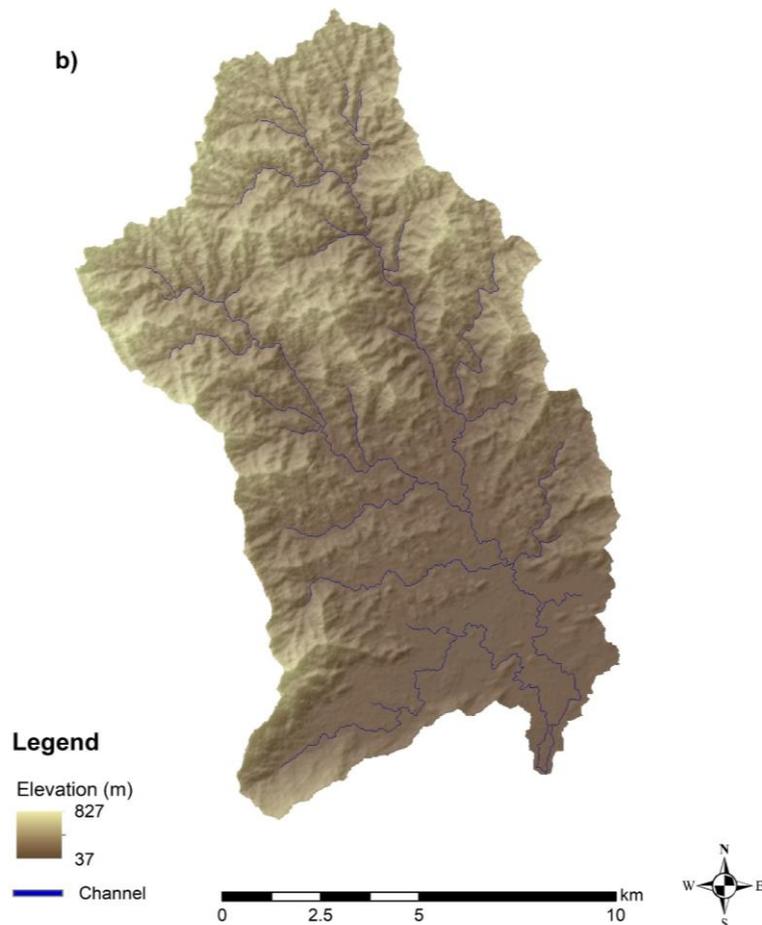
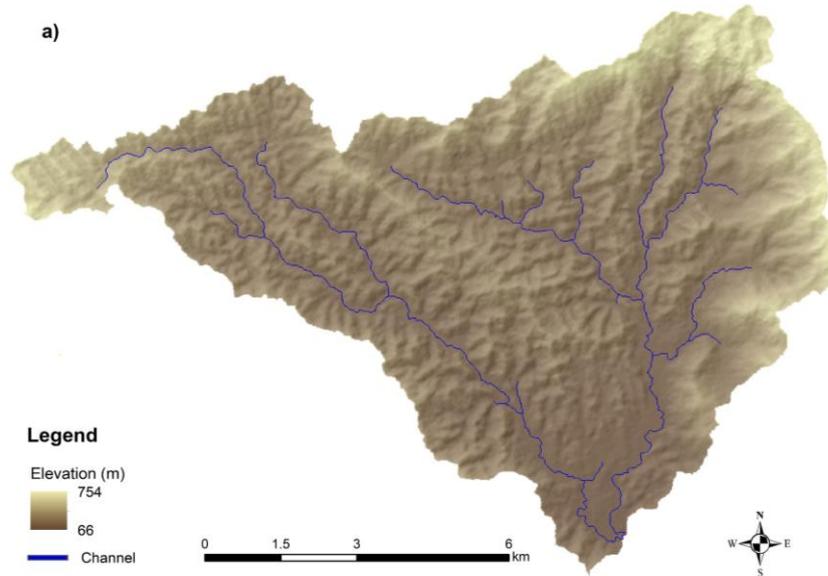


Annex 5. Stream dimension measurements, locations and average values for validation

Number	X-UTM	Y-UTM	Depth	Width
4.1	618205	1970317	2.8	14.76
4.2	619022	1968612	4.06	17.9
4.3	613011	1969383	0.6	2.2
5.1	617007	1965044	1.66	4.4
5.2	615851	1966393	1.67	5
5.3	617997	1969297	4.3	17.9
5.4	618126	1969100	3.8	27.2
6.1	617562	1965041	3.17	13.88
6.2	618275	1967013	3.15	16.85

Stream Order	Depth average (m)	Width average (m)	Shape
4th	2.5	11.6	U
5th	2.9	13.6	U
6th	3.2	15.4	U

Annex 5.1. Channel maps Ban Da Na Kham a) and Laplae b) watershed



Annex 6. PCRaster script for soil depth map generation based on Kuriakose et al. (2009)

```

#! --lddout --unittrue

binding
DEM = dem.map;
soildepth = soildepth20m.map;
LDD = ldd.map;
soild = soild.map;
initial
mask=DEM/DEM;
aspect = scalar(aspect(DEM))*mask;
grad = max(0.001,slope(DEM))*100*mask;
curv = profcurv(DEM);
# LDD = lddcreate(DEM,1e20,1e20,1e20,1e20);
stream = nominal(accuflux(LDD,celllength()) gt 2000);
# used or accumulation of material, so choose larger rivers, not incising rivers
Distance = spread(stream,0,1)*mask;
wetness = ln(accuflux(LDD,cellarea()))/(grad))*mask;
demmax = mapmaximum(DEM);
distmax = mapmaximum(Distance);
report soild = cover(
-0.1*DEM/demmax # lower altitudes give deeper soils
-0.06*grad # steeper slopes giver undeeep soils
+0.01*wetness # higher wetness accumulates material, deeper soils
-0*aspect # no aspect effect
-0*curv # profile curv is - when convex, + when concave. Convex has deeper soils
-0.2*Distance/distmax # perpendicular distance to river, closer gives deeper soils
,0)*mask;
maxd = mapminimum(soild);
soildb = 20*(soild-maxd)**1.1;
# m to mm for lisem, higher power emphasizes deep, updeep
report soildepth = windowaverage(soildb,5*celllength());
# smooth
asp = cover(scalar( aspect(DEM)),0);
# sine gradient (-), make sure slope > 0.001
shade = cos(15)*sin(grad)*cos(asp+45) + sin(15)*cos(grad);
report shade.map = (shade-mapminimum(shade))/(mapmaximum(shade)-mapminimum(shade));
##### not used in lisem

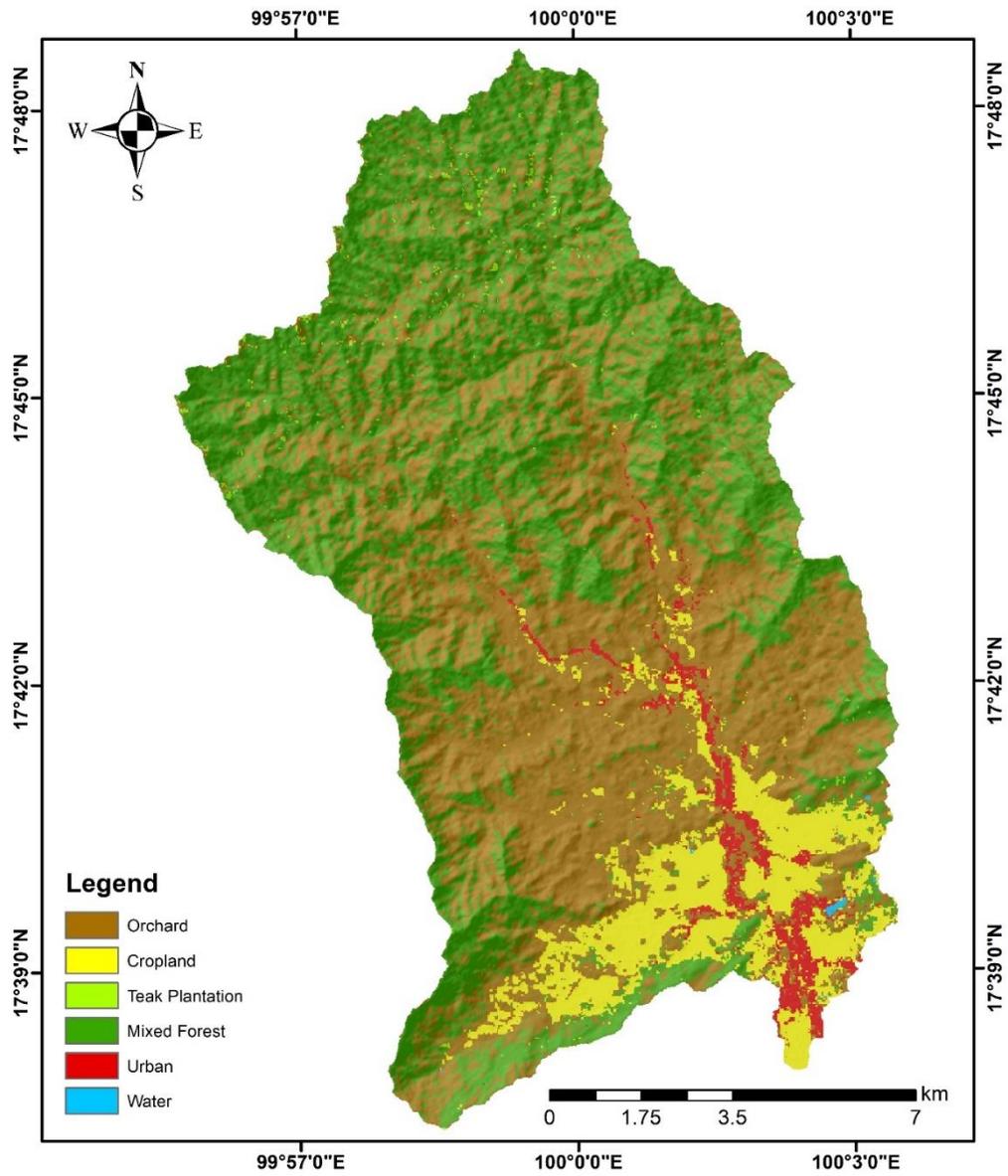
```

Annex 7. Soil sampling site specific Cosine Similarity results

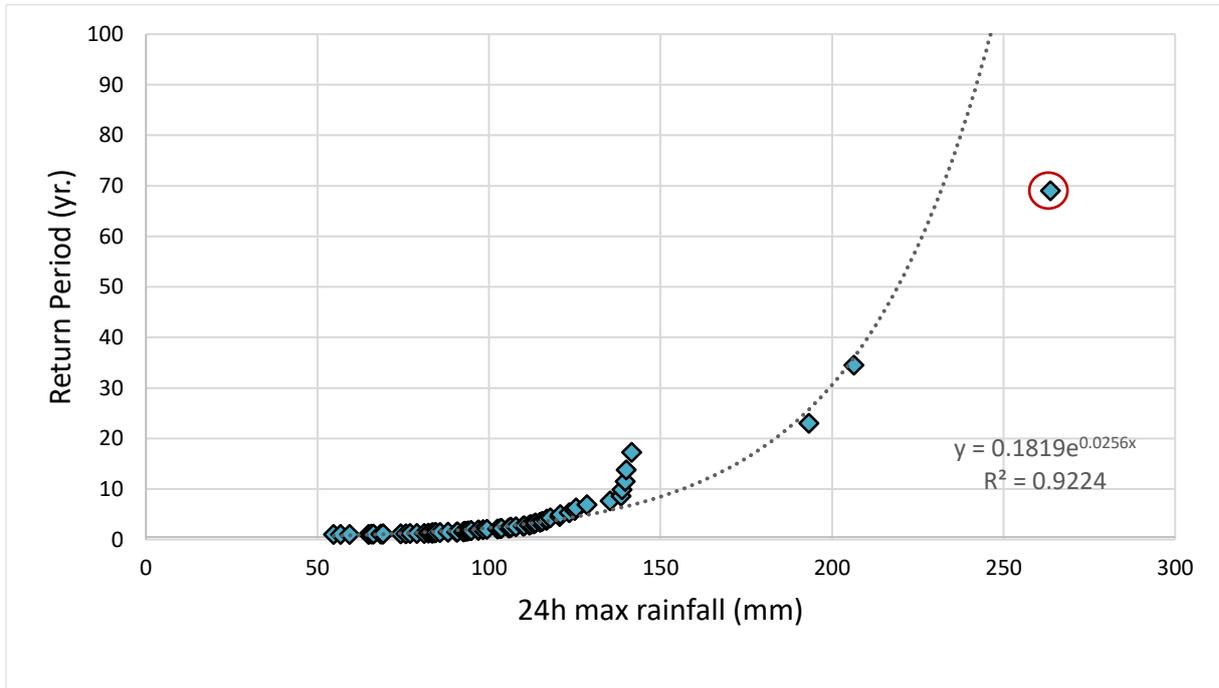
Site	Properties	Properties without	
		PSD	PSD
1	0.7	0.73	0.88
2	0.83	0.9	0.73
3	0.71	0.45	0.79
4	0.9	0.99	0.75
5	0.62	0.9	0.48
6	0.71	0.55	0.57
7	0.65	0.9	0.75
8	0.59	0.59	0.58
9	0.78	0.99	0.62
10	0.89	0.79	0.88
11	0.89	0.59	0.91
12	0.55	0.73	0.68
13	0.75	0.91	0.75
14	0.82	0.8	0.89
15	0.61	0.78	0.88
16	0.8	0.46	0.82
17	0.79	0.92	0.7
18	0.93	0.92	0.94
19	0.82	0.36	0.86
20	0.74	0.95	0.58
21	0.84	0.79	0.86
22	0.92	0.95	0.88
23	0.88	0.95	0.8
24	0.9	0.92	0.91
25	0.93	0.88	0.89
26	0.9	0.86	0.82
27	0.83	0.78	0.68
28	0.82	0.72	0.68
29	0.87	0.79	0.93
30	0.76	0.91	0.61
31	0.8	0.86	0.74
32	0.91	0.81	0.87
33	0.87	0.93	0.83
34	0.7	0.91	0.68
35	0.75	0.83	0.81
36	0.7	0.88	0.49
37	0.79	0.63	0.93
38	0.68	0.5	0.68
39	0.86	0.97	0.9
40	0.89	0.96	0.78
41	0.6	0.78	0.87
42	0.73	0.41	0.62
43	0.47	0.7	0.65
44	0.82	0.86	0.82
45	0.74	0.84	0.62
46	0.63	0.44	0.84
47	0.76	0.73	0.81
48	0.54	0.48	0.97

PSD= Particle Size Distribution, Properties included: Soil Organic Matter, Saturated Hydraulic Conductivity, Porosity, Bulk Density

Annex 8. Land cover map Laplae watershed (2005)



Annex 9. Gumble plot of daily long-term precipitation measurements (1951-2018). Red circle represents the flash flood event in 2006.

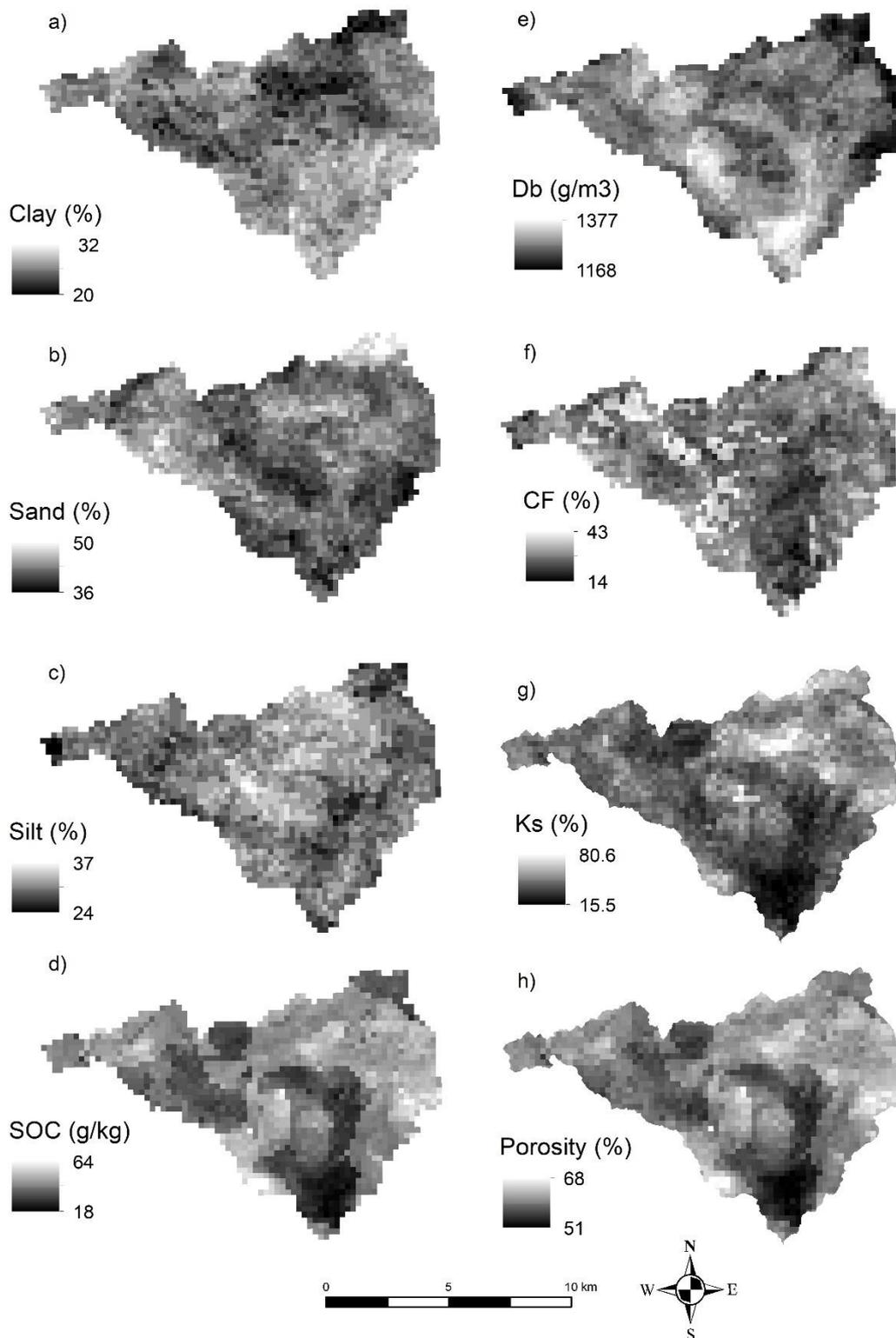


Annex 10. Simulated flood height measurements in comparison to measured flood height points

Point	X-UTM	Y-UTM	Height Measured (m)	Height D _o (m)	Height D _v (m)	Height D _m (m)	Height SRTM (m)
1	609784	1952841	0.704	0.0	0.0	0.0	0.0
2	609766	1952888	0.875	0.0	0.0	0.0	0.0
3	609828	1952893	1.165	1.8	0.1	2.3	0.3
4	609954	1952879	1.08	0.0	0.0	0.0	0.0
5	609819	1952776	0.505	0.0	0.0	0.0	0.1
6	610087	1952820	1.99	0.0	0.0	0.0	0.0
7	610203	1952754	2.35	0.0	1.0	0.0	0.0
8	610261	1952699	2.045	0.6	2.1	0.6	0.2
9	609924	1952528	0.545	0.0	0.0	0.0	0.2
10	609861	1952672	0.585	0.0	0.0	0.0	0.0
11	610251	1952598	1.99	0.0	0.2	0.0	0.0
12	610261	1952462	1.585	0.3	0.5	0.3	0.0
13	610114	1952239	0.6	0.3	0.3	0.3	0.1
14	610060	1952288	0.725	0.1	0.1	0.1	0.0
15	610009	1952344	0.635	0.5	1.3	0.5	0.2
16	609960	1952424	0.965	0.2	1.0	0.2	1.0
17	610278	1952376	1.52	0.0	0.0	0.0	0.0
18	610321	1952256	1.64	0.0	0.5	0.0	0.1
19	610638	1951968	1.91	0.3	0.0	0.5	0.1
20	610545	1951975	2.025	0.1	0.2	0.1	0.0
21	610465	1952010	1.7	0.0	0.0	0.0	0.0
22	610377	1952034	1.51	0.4	1.1	0.4	0.0
23	610334	1952056	1	0.0	0.5	0.0	0.1
24	610298	1952065	0.66	0.0	0.0	0.0	0.6
25	610283	1952101	0.86	0.2	0.2	0.2	1.6
26	610248	1952146	0.66	0.0	0.0	0.0	1.1
27	610163	1952196	0.67	0.5	0.5	0.5	0.0
28	610358	1952150	1.44	0.4	1.1	0.4	0.1
29	610365	1952075	1.375	0.4	1.1	0.4	0.0
30	610295	1952032	0.54	0.0	0.0	0.0	0.0
31	610738	1951944	1.85	0.3	2.2	0.5	0.0
32	610687	1951955	2.01	1.5	3.4	1.8	0.0
33	610272	1951929	0.525	0.0	0.0	0.0	0.1
34	610258	1951778	0.925	0.5	0.0	0.5	0.0
35	610536	1951933	1.575	0.1	0.1	0.1	0.1
36	610509	1951864	1.73	0.0	0.1	0.0	2.4
37	610487	1951768	1.305	0.3	0.4	0.3	0.0
38	610777	1951942	1.35	0.0	1.8	0.1	0.1
39	610255	1951668	0.88	1.8	1.4	1.8	3.1
40	610271	1951554	1.03	2.1	1.7	2.1	1.1
41	610486	1951667	1.46	0.9	1.2	0.9	1.0
42	610478	1951554	1.61	0.0	0.0	0.0	0.9
43	610277	1951446	1.1	0.8	0.4	0.8	0.0
44	610496	1951456	1.275	0.9	1.2	0.9	0.1
45	610519	1951379	1.255	0.0	0.0	0.0	1.3

Height measured = historical flood height, Height D_o = simulated water height with original DEM, Height D_v = simulated water height with DEM without vegetation, Height D_m = simulated water height with manipulated DEM, Height SRTM = simulated water height with SRTM DEM

Annex 11. Physical property layers of SoilGrids for soil layer 1 (5 cm)



D_b = Bulk Density, CF = Coarse Fragments, K_s = Saturated Hydraulic Conductivity, SOC = Soil Organic Carbon