

Final Report

Improvement of Input Spatial Information to Support Meteorological Modeling

โครงการปรับปรุงข้อมูลเชิงพื้นที่นำเข้าเพื่อสนับสนุนการจำลองอุตุนิยมวิทยา

Jointly Conducted by



Researchers

Kasemsan Manomaiphiboon (JGSEE)

Kanoksri Sarinnapakorn (HAIL)

Nishit Aman (JGSEE)

Thippawan Thodsan (HAIL)

Surajate Boonya-Aroonnet (HAIL)

Nosha Assareh (JGSEE)

Peraya Tantanuparp (HAIL)

Aungkana Pratumthong (HAIL)

Funded by

Hydro and Agro Informatics Institute (Public Organization)

March 2017

Final Report

Improvement of Input Spatial Information to Support Meteorological Modeling

Jointly Conducted by

Hydro and Agro Informatics Institute (HAI) and
The Joint Graduate School of Energy and Environment (JGSEE)

Researchers

Kasemsan Manomaiphiboon (JGSEE, PI)	Surajate Boonya-Aroonnet (HAI, PI)
Kanoksri Sarinnapakorn (HAI)	Nosha Assareh (JGSEE)
Nishit Aman (JGSEE)	Peraya Tantanuparp (HAI)
Thippawan Thodsan (HAI)	Aungkana Pratumthong (HAI)

Funded by

Hydro and Agro Informatics Institute (Public Organization)

March 2017

รายงานวิจัยฉบับสมบูรณ์

โครงการปรับปรุงข้อมูลเชิงพื้นที่นำเข้าเพื่อสนับสนุน

การจำลองอุทกนิยามวิทยา

วิจัยร่วมโดย

สถาบันสารสนเทศทรัพยากรน้ำและการเกษตร (องค์การมหาชน)

และ บัณฑิตวิทยาลัยร่วมด้านพลังงานและสิ่งแวดล้อม

คณะผู้วิจัย

นายเกษมสันต์ มโนมัยพิบูลย์ (หัวหน้าโครงการ บัณฑิตวิทยาลัยร่วมด้านพลังงานและสิ่งแวดล้อม)

นายสุรเจตส์ บุญญาอรุณเนตร (หัวหน้าโครงการ สถาบันสารสนเทศทรัพยากรน้ำและการเกษตร)

นางสาวกนกศรี ศรีนินภากร (สถาบันสารสนเทศทรัพยากรน้ำและการเกษตร)

นายโนชา อาษาเรห์ (บัณฑิตวิทยาลัยร่วมด้านพลังงานและสิ่งแวดล้อม)

นายนิชิต อามาน (บัณฑิตวิทยาลัยร่วมด้านพลังงานและสิ่งแวดล้อม)

นางสาวพีรญา ตันตือนภาพ (สถาบันสารสนเทศทรัพยากรน้ำและการเกษตร)

นางสาวทิพย์วรรณ ทอดแสน (สถาบันสารสนเทศทรัพยากรน้ำและการเกษตร)

นางสาวอังคณา ประทุมทอง (สถาบันสารสนเทศทรัพยากรน้ำและการเกษตร)

สนับสนุนทุนวิจัยโดย

สถาบันสารสนเทศทรัพยากรน้ำและการเกษตร (องค์การมหาชน)

มีนาคม 2560

Abstract

This study has completed the development of a suite of computer programs, called GUPROC (GEOGRID Updating Processor), which is capable of updating input surface data for WRF modeling to reduce uncertainty in the modeling due to unrealistic or unrepresentative surface data. GUPROC incorporates new surface data compiled from various sources into GEOGRID (a main WRF preprocessor) using an offline approach. This approach does not interfere with any internal source codes, scripts, and input control files in GEOGRID. The GUPROC development relied mainly on open-source and/or freeware software and tools. Currently, GUPROC is able to update the following variables: terrain height, USGS land use, monthly green fraction, monthly leaf area index, monthly surface albedo, top-layer soil texture, bottom-layer soil texture, and land mask. It can handle both Mercator or Lambert conformal conic map projections and support modeling studies over areas in Thailand and in the Lower Mekong River region at resolutions of 1 km or coarser. GUPROC was demonstrated using the WRF modeling domains used by the HAIL forecasting operation. The updated results given by GUPROC generally show significant differences from the default results for all variables considered. For terrain height, both default and updated results are generally comparable but their discrepancy tends to be amplified over mountainous areas. The updating satisfactorily yields more realistic results with improved spatial details for the land use, green fraction, leaf area index, and albedo variables. For soil texture (both top-layer and bottom-layer), the default results are shown to be superior because the new soil texture data selected for the updating are unfortunately somewhat outdated for Thailand and its neighboring countries. Hence, the default soil texture data should still be maintained for use. Finally, certain practical recommendations as well as perspectives regarding GUPROC application and future enhancement are also given.

บทคัดย่อ

การศึกษานี้ได้พัฒนาชุดโปรแกรมคอมพิวเตอร์ โดยให้ชื่อว่า GUPROC ซึ่งย่อมาจาก GEOGRID Updating Processor โดยมีความสามารถในการปรับปรุงข้อมูลผิวพื้นนำเข้าแบบจำลองสภาพ (Weather Research and Forecasting Model หรือ WRF) เพื่อช่วยลดความไม่แน่นอนในการจำลองในกรณีที่ข้อมูลนำเข้ามีความไม่สอดคล้องกับสภาพจริงหรือไม่มีความเป็นตัวแทนที่เหมาะสมต่อการนำมาใช้ ชุดโปรแกรมคอมพิวเตอร์ดังกล่าวได้พิจารณาและเตรียมข้อมูลผิวพื้นชุดใหม่ขึ้นมา ซึ่งมาจากการรวบรวมจากแหล่งต่างๆ โดยนำมาใช้กับ GEOGRID ซึ่งเป็นหน่วยประมวลขั้นต้นของแบบจำลองสภาพ โดยการปรับปรุงมีลักษณะเป็นแบบออฟไลน์ (Offline) ซึ่งจะไม่เกี่ยวข้องแก้ไขหรือรบกวนโปรแกรมคอมพิวเตอร์ชุดคำสั่ง และไฟล์ควบคุม ซึ่งอยู่ภายใน GEOGRID แต่อย่างใด การพัฒนาชุดโปรแกรมคอมพิวเตอร์ได้ใช้ภาษาคอมพิวเตอร์มาตรฐานและอาศัยซอฟต์แวร์โอเพนซอร์สเป็นหลักในการทำงาน โดย ณ ปัจจุบันสามารถปรับปรุงตัวแปรต่างๆ ดังนี้: ความสูงภูมิประเทศ (Terrain Height) การใช้ประโยชน์ที่ดินและสิ่งปกคลุมดิน (Land Use/Land Cover) สัดส่วนพืชเป็นรายเดือน (Monthly Green Fraction) ดัชนีพื้นที่ใบเป็นรายเดือน (Leaf Area Index) อัลบิโดผิวพื้นหรือค่าสะท้อนแสงของผิวพื้นเป็นรายเดือน (Surface Albedo) เนื้อดินชั้นบน (Top-Layer Soil Texture) เนื้อดินชั้นล่าง (Bottom-Layer Soil Texture) และมาสก์พื้นดิน (Land Mask) (1: พื้นดิน and 0: น้ำ) นอกจากนี้ สามารถรองรับเส้นโครงแผนที่ทั้งแบบเมอร์เคเตอร์ (Mercator) และแบบแลมเบิร์ต (Lambert Conformal Conic) และรองรับการจำลองสำหรับพื้นที่ในประเทศไทยและพื้นที่ลุ่มน้ำโขงตอนล่างได้ด้วยความละเอียดกริด 1 กิโลเมตร หรือหยาบกว่า การศึกษานี้ยังได้ทำการสาธิตการปรับปรุงข้อมูลผิวพื้นด้วยชุดโปรแกรมคอมพิวเตอร์สำหรับโดเมนจำลองสภาพที่ใช้ในงานพยากรณ์ของสถาบันสารสนเทศทรัพยากรน้ำและการเกษตร และพบว่า โดยทั่วไป ผลที่ได้จากการปรับปรุงมีความแตกต่างอย่างมีนัยสำคัญเมื่อเปรียบเทียบกับผลเดิมก่อนการปรับปรุง สำหรับตัวแปรความสูงภูมิประเทศ พบว่า ผลเดิมกับผลที่ปรับปรุงมีความสอดคล้องกันในภาพรวม อย่างไรก็ตาม ผลทั้งสองจะแตกต่างกันมากขึ้นและเห็นได้ชัดสำหรับพื้นที่หรือบริเวณภูเขา สำหรับตัวแปรการใช้ประโยชน์ที่ดินและสิ่งปกคลุมดิน สัดส่วนพืช ดัชนีพื้นที่ใบ และอัลบิโดผิวพื้น การปรับปรุงให้ผลเป็นที่น่าพอใจ มีความสอดคล้องกับสภาพจริงมากขึ้นและให้รายละเอียดเชิงพื้นที่ได้ดีขึ้นสำหรับตัวแปรเนื้อดินชั้นบนและชั้นล่าง พบว่า ผลเดิมมีคุณภาพสูงกว่าผลที่ปรับปรุง ทั้งนี้ เนื่องจากข้อมูลเนื้อดินที่ได้ลองเลือกเพื่อใช้ปรับปรุงนั้น เป็นข้อมูลที่ค่อนข้างเก่าไม่ทันสมัยสำหรับพื้นที่ในประเทศไทยและประเทศเพื่อนบ้าน ดังนั้น จึงควรใช้ผลเดิมเพื่อการจำลองสำหรับตัวแปรนี้ และท้ายสุด ได้ให้มุมมองและข้อเสนอแนะสำหรับการประยุกต์ใช้ชุดโปรแกรมคอมพิวเตอร์ที่ได้พัฒนาขึ้นมาและการพัฒนาเสริมในอนาคต

Acknowledgements

The researchers would like to thank the following organizations and people:

- Hydro and Agro Informatics Institute for financial support and technical assistance
- The Joint Graduate School of Energy and Environment and King Mongkut's University of Technology Thonburi for general support
- Land Development Department for the land use/land cover data for Thailand
- Royal Irrigation Department for the irrigated-area data for Thailand
- Dr. Sugunyanee Yavinchan, Thai Meteorological Department, as consultant to this study
- Development and supporting people of the following software and data products: WRF, FEDORA CORE, G95/GFORTRAN, QGIS, R, GrADS, and MODIS (land use/land cover, vegetation indices, vegetation continuous fraction, and surface albedo), ASTER-GDEM, MERRA Reanalysis, and HWSO and
- People at the JGSEE Computational Laboratory (KMUTT Bang Khun Thian Campus) for general assistance.

Table of Contents

	Page
Abstract	i
บทคัดย่อ (Abstract in Thai)	ii
Acknowledgements	iv
Table of Contents	vi
List of Tables	viii
List of Figures	ix
Chapter 1: Introduction	1
1.1 Rationale	1
1.2 Objective	3
1.3 Scope of work	3
Chapter 2: Methods	5
2.1 Offline updating	5
2.2 Reference domain and tiles	5
2.3 Thailand tiles	6
2.4 Cell assignment by bin filling	11
2.5 Fixing cells with a missing value	12
2.6 Software, data formats, and data manipulation	12
2.7 Variables to be updated	15
2.8 Terrain height data	20
2.9 Land use data	20
2.10 Monthly green fraction data	32
2.11 Monthly leaf area index data	33
2.12 Monthly surface albedo data	33
2.13 Soil texture	35
Chapter 3: GUPROC and Illustrated Results	39
3.1 GUPROC	39
3.2 Folder domain/	39
3.3 Folder gis/	40

Table of Contents

(Continued)

	Page
3.4 Folder io_default/	41
3.5 Folder io_update/	41
3.6 Folder proc/	42
3.7 Illustrated results	49
Chapter 4: Conclusions and Recommendations	81
4.1 Conclusions	81
4.2 Recommendations	82
References	85
Appendix	89
Researcher Information	111

List of Tables

	Page
Table 2.1 Land use matching from IGBP to USGS classes	26
Table 2.2 Land use matching from LDD to USGS classes	27
Table 2.3 Soil texture matching from USDA to WRF classes	37

List of Figures

	Page
Figure 2.1 Spatial extents of the reference, HAIL and TMD grids	7
Figure 2.2 Modeling domains in the HAIL forecasting operation	8
Figure 2.3 Reference tiles	9
Figure 2.4 Thailand tiles	10
Figure 2.5 Bin filling with the 5:1 ratio	13
Figure 2.6 Fixing a missing value using neighboring cells	14
Figure 2.7 Soil texture classifications by the USDA	19
Figure 2.8 Sinusoidal-projected MODIS tiles	24
Figure 2.9 Representative years of the LDD land use data	25
Figure 3.1 Regions of Thailand according to the TMD	52
Figure 3.2 Terrain height (m) in Northern region	53
Figure 3.3 Terrain height (m) in Northeastern region	54
Figure 3.4 Terrain height (m) in Central & Eastern regions	55
Figure 3.5 Terrain height (m) in Southern region	56
Figure 3.6 Dominant land use class in Northern region	57
Figure 3.7 Dominant land use class in Northeastern region	58
Figure 3.8 Dominant land use class in Central & Eastern regions	59
Figure 3.9 Dominant land use class in Southern region	60
Figure 3.10 Green fraction in Northern region	61
Figure 3.11 Green fraction in Northeastern region	62
Figure 3.12 Green fraction in Central & Eastern regions	63
Figure 3.13 Green fraction in Southern region	64
Figure 3.14 Leaf area index ($\text{m}^2 \text{m}^{-2}$) in Northern region	65
Figure 3.15 Leaf area index ($\text{m}^2 \text{m}^{-2}$) in Northeastern region	66
Figure 3.16 Leaf area index ($\text{m}^2 \text{m}^{-2}$) in Central & Eastern regions	67
Figure 3.17 Leaf area index ($\text{m}^2 \text{m}^{-2}$) in Southern region	68

List of Figures

(Continued)

	Page
Figure 3.18 Surface albedo (%) in Northern region	69
Figure 3.19 Surface albedo (%) in Northeastern region	70
Figure 3.20 Surface albedo (%) in Central & Eastern regions	71
Figure 3.21 Surface albedo (%) in Southern region	72
Figure 3.22 Dominant top-layer soil texture in Northern region	73
Figure 3.23 Dominant top-layer soil texture in Northeastern region	74
Figure 3.24 Dominant top-layer soil texture in Central & Eastern regions	75
Figure 3.25 Dominant top-layer soil texture in Southern region	76
Figure 3.26 Dominant bottom-layer soil texture in Northern region	77
Figure 3.27 Dominant bottom-layer soil texture in Northeastern region	78
Figure 3.28 Dominant bottom-layer soil texture in Central & Eastern regions	79
Figure 3.29 Dominant bottom-layer soil texture in Southern region	80
Figure 4.1 Hardware platform used in GUPROC development	84
Figure A1 Terrain height (m) in Domain 1	89
Figure A2 Terrain height (m) in Domain 2	90
Figure A3 Terrain height (m) in Domain 3	91
Figure A4 Dominant land use class in Domain 1	92
Figure A5 Dominant land use class in Domain 2	93

List of Figures

(Continued)

	Page
Figure A6	94
Figure A7	95
Figure A8	96
Figure A9	97
Figure A10	98
Figure A11	99
Figure A12	100
Figure A13	101
Figure A14	102
Figure A15	103
Figure A16	104
Figure A17	105
Figure A18	106
Figure A19	107
Figure A20	108
Figure A21	109

Chapter 1

Introduction

1.1. Rationale

Numerical Weather Prediction (NWP) is a technical procedure by which current and future weather or meteorological conditions at a location or over an area of interest is predicted numerically. The central component of the NWP is meteorological modeling that includes a mathematical model or models, representing various atmospheric (and oceanic) processes. There is a broad range of spatial scales over which meteorological modeling can be performed – mainly, global, continental, synoptic, regional/meso, and microscale/local. At a regional scale or mesoscale, various physical processes are generally involved, and they can interact with each other, e.g., land surface, radiation, atmospheric boundary layer and turbulence, and cloud. Several mesoscale meteorological models have been developed, e.g., WRF (Weather Research and Forecasting), MM5 (5th Generation Penn State/NCAR Mesoscale Model, which is WRF's predecessor), RAMS (Regional Atmospheric Modeling System), COSMO (Consortium for Small-Scale Modeling), and HIRLAM (High-Resolution Limited Area Model). They are used in weather forecasting operation, weather/climate research, environmental management & warning (e.g., water resources and air pollution), and also renewable energy (e.g., solar and wind forecasting and resource assessment).

WRF (Skamarock et al. 2008) is a mesoscale meteorological model widely used in weather and climate by research and operational communities. Its development has been mainly supported by various organizations in the US, which include the National Center for Atmospheric Research (NCAR), the National Oceanic and Atmospheric Administration, the National Centers for Environmental Prediction (NCEP) and others. The model is open-source, community-based, and free of charge in use, with continuous development and enhancement. It has two dynamical cores, ARW (Advanced Research WRF) and NMM

(Non-Hydrostatic Mesoscale Model). The former is the core, generally adopted by model users and often used as a conventional platform for coupling with other related systems, e.g. WRFDA (WRF Data Assimilation System), WRF-Chem (WRF Atmospheric Chemistry Model), WRF-Hydro (WRF Hydrological Modeling System). More WRF-related information is referred to <http://www2.mmm.ucar.edu/wrf/users>. In Thailand, there are active WRF (specifically, WRF-ARW) users in governmental agencies (notably, the Thai Meteorological Department or TMD and the Hydro and Agro Informatics Institute or HAIL), universities, and research centers. At the HAIL, the forecasting operation (Torsri et al. 2014) has been based mainly on WRF, but extended to coupling with ROMS (Regional Ocean Modeling System) (Shchepetkin and McWilliams 2005), within the framework of COAWST (Coupled-Ocean-Atmosphere-Wave-Sediment Transport Modeling System (Warner et al. 2010).

To implement a mesoscale meteorological model, such as WRF, requires a large numbers of input datasets, and a group of which is concerned with static spatial surface data primarily for characterizing user-specified modeling domains through GEOGRID. GEOGRID is essentially a WRF preprocessor, part of WRF's Preprocessing System (WPS). Its main function is to define modeling domains and fill their cells with surface (or terrestrial) data. By default, static spatial surface data are given inside the WRF package. However, the quality of some (or most) of the default surface data is technically inadequate or not high to support WRF application for Thailand and its vicinity, e.g., coarse spatial resolution, datedness, and misrepresentation. These may potentially impact the modeling at later stages and affect modeled results as a result. In other words, they potentially produce flawed or incorrect feedbacks from surface-related processes included in the model, causing a) reduced scientific reliability for the modeling, b) increased uncertainty in interpreting modeled results, and c) increased difficulty in judging the model's prediction performance, especially due to different physics options. Accordingly, good-quality and realistic surface data should be incorporated into the modeling. Such data can be available from local and

international sources. Satellite-derived data are alternatively considered one such data source. They generally cover a relatively large spatial extent with a moderate-to-fine resolution, which is applicable to mesoscale modeling, and have today become more available and accessible to the public. There are modeling studies with updated input surface data have been conducted. A large number of international studies were found, e.g., Foy et al. (2006), Li et al. (2014), Ran et al. (2016), Sugimoto et al. (2015), Vahmani and Hogue (2014), and Wang et al. (2014). However, much fewer studies were found for Thailand, e.g., Manomaphiboon et al. (2016), Octaviani and Manomaiphiboon (2011), and Paton and Manomaiphiboon (2013). Motivated by these, this study aims to incorporate new surface data in support of WRF application to Thailand.

1.2 Objective

To update input surface data at GEOGRID to improve the suitability of WRF modeling for areas in Thailand and its vicinity.

1.3 Scope of work

This study has implemented the following tasks:

1. Study the attributes of GEOGRID default input surface data
2. Identify GEOGRID output variables to be updated
3. Determine an overall spatial extent to be covered by the study
4. Survey, compile, inspect, and rearrange new surface data
5. Develop a suite of computer programs, to be referred to as “GEOGRID Updating Processor” (shortly, GUPROC), which is to incorporate the compiled new surface data into WRF modeling domains of interest
6. Demonstration GUPROC and illustrate results generated using the default and new surface data and
7. Transfer the developed GUPROC to appropriate governmental agencies.

Chapter 2

Methods

In this chapter, the key technical components of GUPROC and their concepts will be introduced and described.

2.1 Offline updating

Incorporating new surface data into WRF at GEOGRID can be done by either an online or offline approach. The former is to prepare new input data as layers for GEOGRID. Modification of some internal source codes and input control files of GEOGRID may be needed. The latter is to run GEOGRID using default surface data first and then replace or update any desired variables of GEOGRID output with new surface data and does not interfere with any internal source codes, input control files, and default surface data. Due to the latter's relative simplicity, the offline approach was adopted as the updating basis for GUPROC in this study.

2.2 Reference domain and tiles

One of the very first technical steps to take in the implementation of this study is to determine the maximum spatial extent for GUPROC to handle, which will be referred to as “the reference domain (or grid)”. Here, the largest WRF modeling domain used by the HAIL forecasting operation was examined, and so was that by the TMD forecasting operation (Figure 2.1). It is noted that, at the time of study, the HAIL forecast operation uses three modeling domains with grid resolutions (shortly, resolutions) of 27 km, 9 km, and 3 km (Figure 2.2), whereas the TMD uses two modeling domains with resolutions of 30 km and 10 km. Based on the examination, the size of the reference domain was set to be larger than both, which is from -15° to 45° in latitude (width = 60°) and from 70° to 140° in longitude (length = 70°).

For convenience in data handling and processing by GUPROC, the reference domain is divided into 168 $5^\circ \times 5^\circ$ tiles (Figure 2.3). Each individual tile has 2,500 cells along the tile width, thus containing 6.25-million $0.002^\circ \times 0.002^\circ$ cells in total. These tiles will be referred to as “the reference tiles”. In view of distance, 1° is approximately equivalent to 108 km near the equator. The primary purpose of the reference tiles is to store data in a flat binary format, which is the main data format used by GUPROC to store intermediately generated data before they are finally processed to update GEOGRID output variables (see Section 2.7). Since GUPROC aims to be technically capable of supporting any WRF modeling domains with resolutions of 1 km or coarser, the resolution of the reference tiles (0.002° or 216 m) is considered sufficiently fine to fill values into such modeling domains using the bin-filling method (see Section 2.4).

2.3 Thailand tiles

Besides the reference tiles, another set of tiles was used in this study, which is 76 $1^\circ \times 1^\circ$ tiles covering all land parts of Thailand (Figure 2.4). They will be referred to as “the Thailand tiles”, each of which has 2,000 cells along its width and thus contains 4-million $0.0005^\circ \times 0.0005^\circ$ cells in total. The purpose of these tiles is to store data with a very fine resolution (e.g., 30 m), which are rasterized from high-quality vector data acquired from local sources, before being later filled into the reference tiles. Since the resolution of the reference tiles is about 4 times coarser than that of the Thailand tiles, the bin-filling method (see Section 2.4) is considered suitable to fill data stored in the Thailand tiles into cells of the reference tiles.

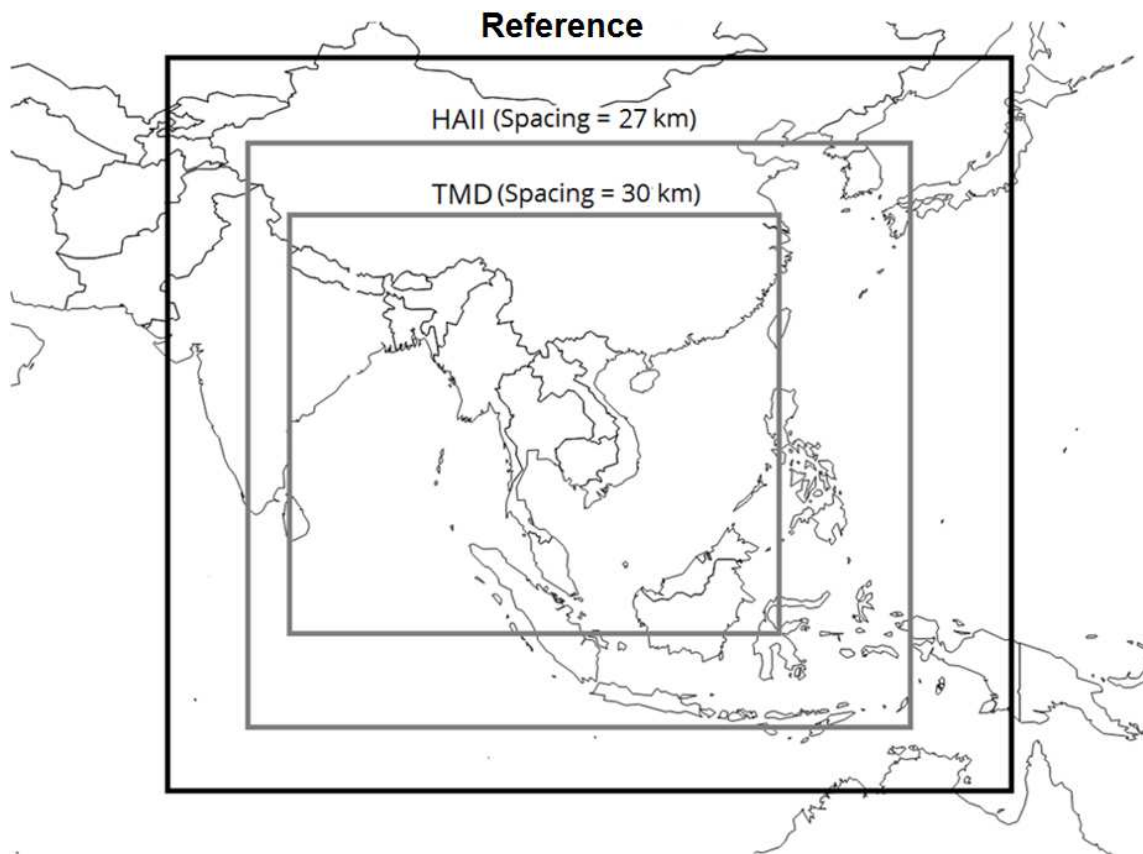
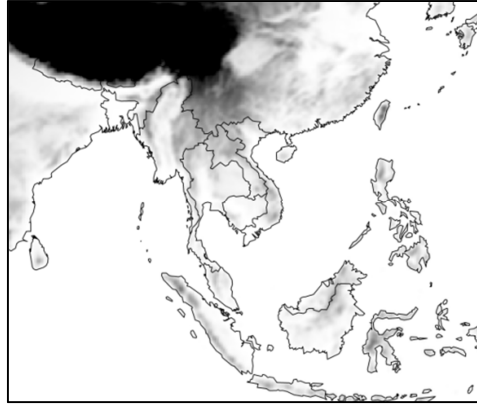
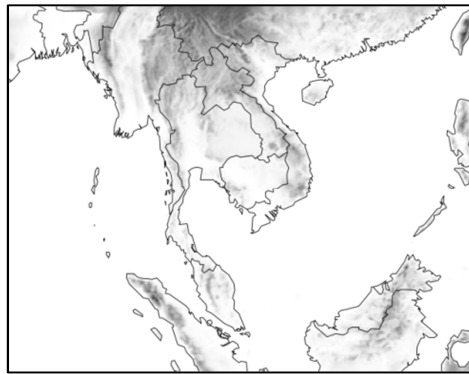


Figure 2.1 Spatial extents of the reference, HAIL and TMD grids.

a) Domain 1



b) Domain 2



c) Domain 3

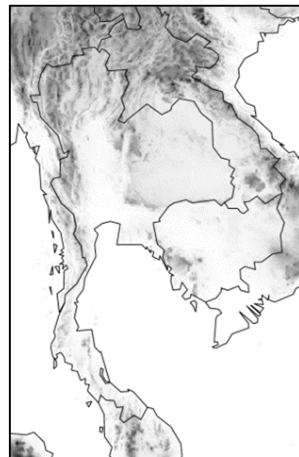


Figure 2.2 Modeling domains in the HAIL forecasting operation.

Gray shading denotes terrain height above mean sea level.

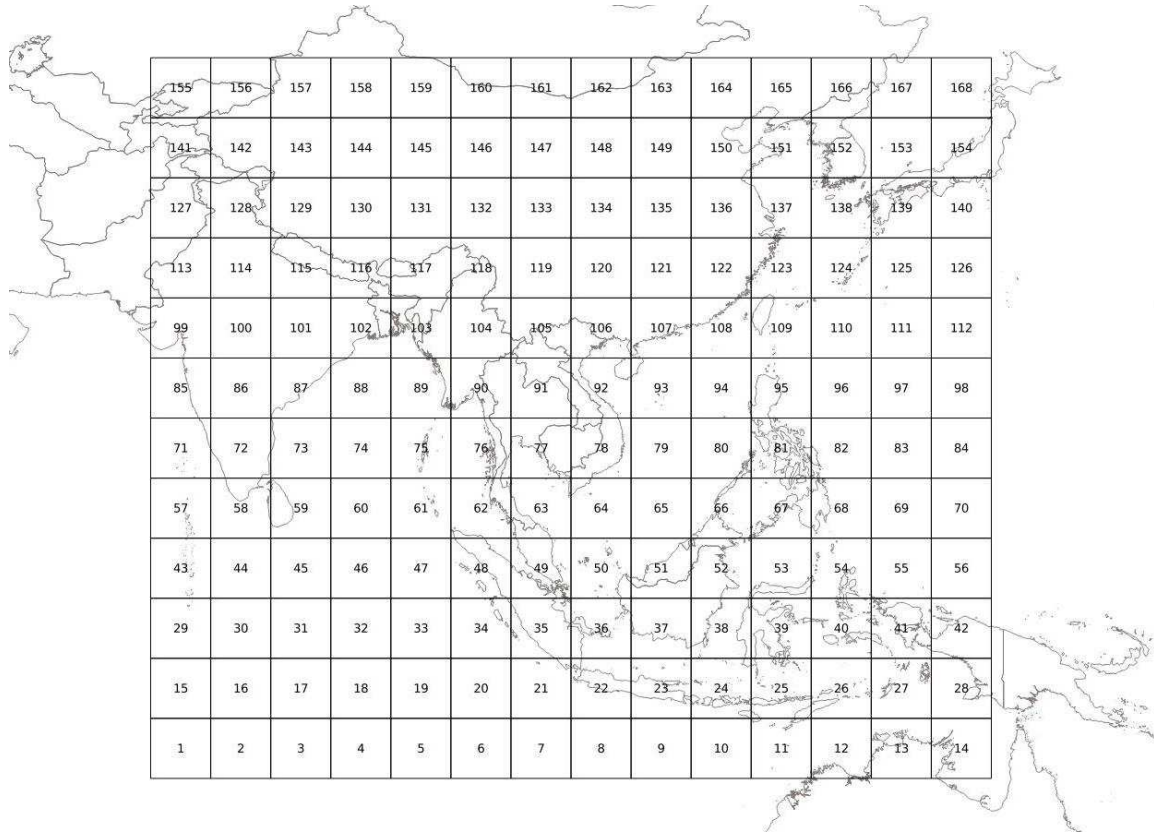


Figure 2.3 Reference tiles.

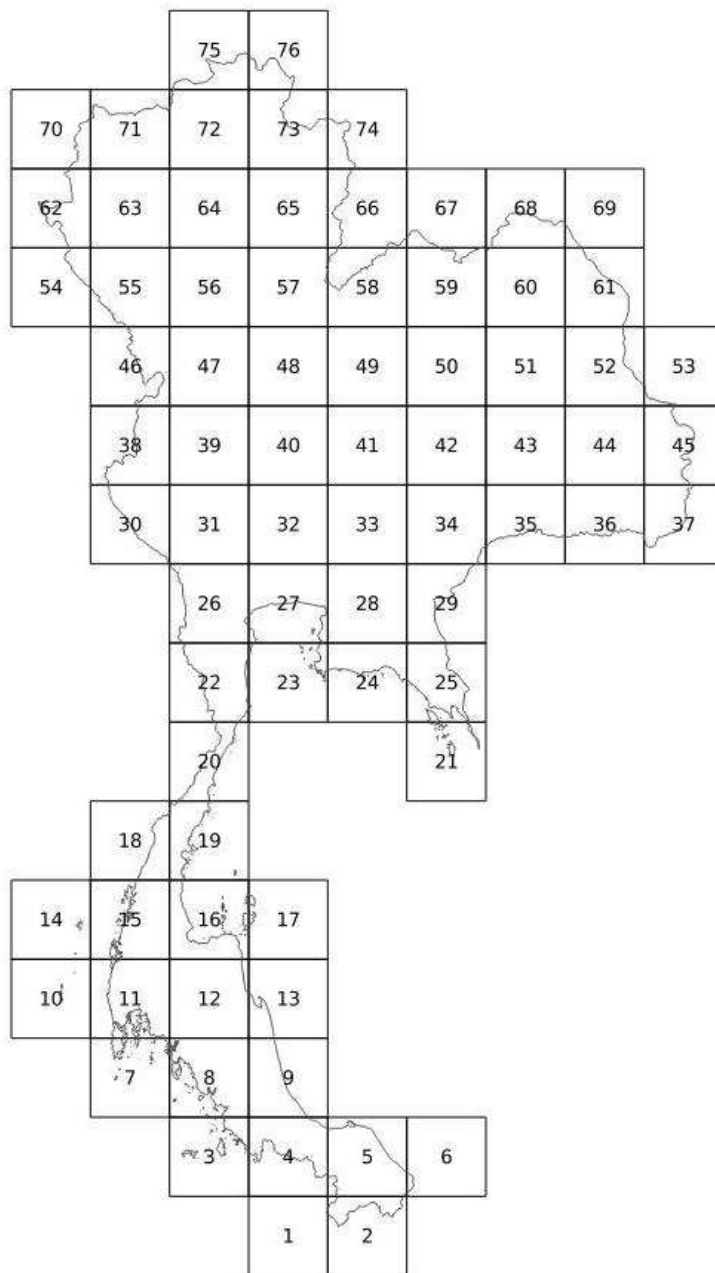


Figure 2.4 Thailand tiles.

2.4 Cell assignment by bin filling

Several datasets are involved in this study, different grid configurations and data formats. In general, a grid configuration is determined by map projection and cell arrangement (shape, number, size, and alignment). To assign or compute the value of a cell in a target grid using cell values from an initial or filling grid, the map projections and datums of both grids are first needed to be compatible. Next, both grids are overlaid upon each other. The value of a target cell is then determined using those of the filling cells that intersect with the target cell or are in its proximity. Here, a simple bin-filling method was employed for cell assignment, considering only the filling cells whose centers fall within the target cell (Figure 2.5). For a categorical variable, one can assign the mode (i.e., majority) of the values pooled from the filling cells to the target cell:

$$q_{\text{target}} = \text{Mode}(q_1, q_2, q_3, \dots, q_n), \quad (2.1)$$

where q_{target} is the value of the target cell, and q_i is the value of the i^{th} filling cell. For a numeric variable, the average value can be used:

$$q_{\text{target}} = \frac{\sum_i^n q_i}{n}. \quad (2.2)$$

In practice, the bin-filling method has been used and viewed acceptable in earth-science modeling. The quality of its cell assignment depends directly on the ratio of the resolution of the target grid to that of the filling grid (the larger ratio, the better quality). The 3:1 ratio is typically recommended as the minimum in modeling practice. As for margin of error, the 3:1, 4:1, and 5:1 ratios approximately yield <26%, <20%, and <17%, respectively. An example shown in Figure 2.5 has the target grid coarser than the filling grid. The cell assignment can be thought of as cell aggregation from finer (filling) to coarser (target) resolutions. Contrary to that, when the target grid is finer than the filling grid, the value of the target cell can be directly assigned as that of the filling cell upon which the target cell's center falls. The

nature of this cell assignment is resampling from coarser (filling) to finer (target) resolutions. In case of the resolutions of the target and filling grids do not differ much or are the same, each target cell may be further divided into smaller cells to achieve improved accuracy in the resampling.

2.5 Fixing cells with a missing value

Gridded data may contain missing (or undefined) values. Here, a simple method was employed to fix them, by which a cell (with the missing value) is assigned with a representative value computed using the values pooled from its neighboring cells (average or median for a numeric variable, but mode for a categorical variable). The number of values to be pooled depends on the size of neighborhood for the cell to be fixed. More description of this method is arranged in the caption of Figure 2.6.

2.6 Software, data formats, and data manipulation

The development of GUPROC has mainly relied on open-source or freeware software and tools. All computer scripts and codes were written in standard computer languages (here, Csh, FORTRAN, and R), which are thus convenient to modify and revise in the future. Since several datasets are involved, as mentioned above, and they may have different data formats. The following are the important data formats handled in this study:

- NetCDF (Network Common Data Form)
(<https://www.unidata.ucar.edu/software/netcdf>)
- HDF (Hierarchical Data Format) & HDF-EOS (HDF-Earth Observing System)
(<https://support.hdfgroup.org/HDF5> and
<https://earthdata.nasa.gov/standards/hdf-eos5>)
- GeoTiff (Georeferenced Tagged Image File Format)
(<https://trac.osgeo.org/geotiff> and
http://www.gdal.org/frmt_gtiff.html)

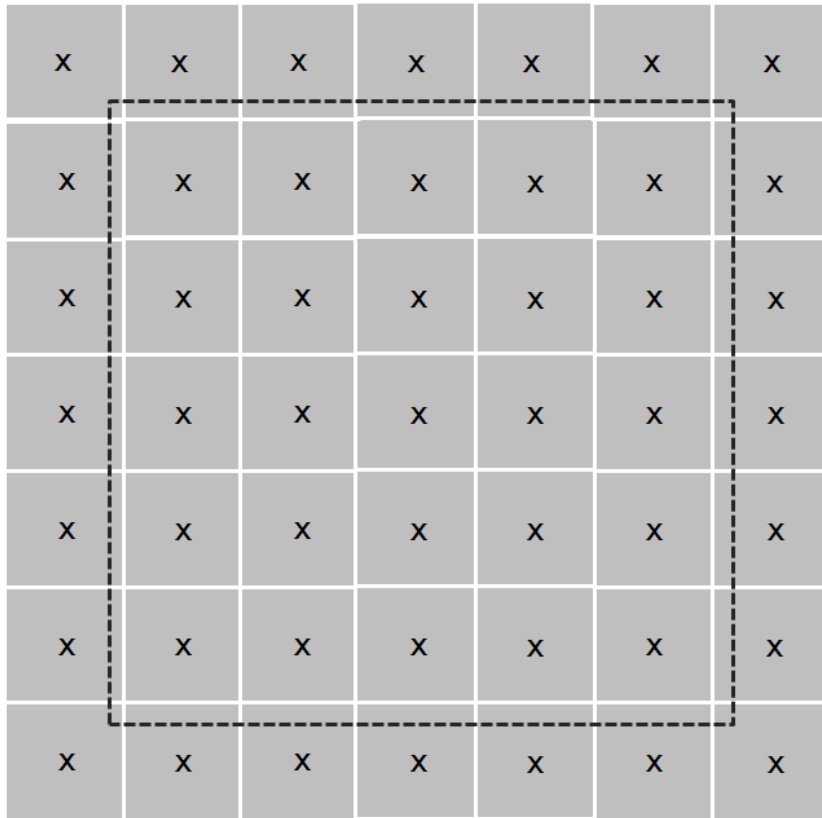


Figure 2.5 Bin filling with the 5:1 ratio.

In the above figure, the value of the target cell (dashed, coarse) is determined by those of the filling cells (gray-shaded, fine) whose centers fall within the target cell.

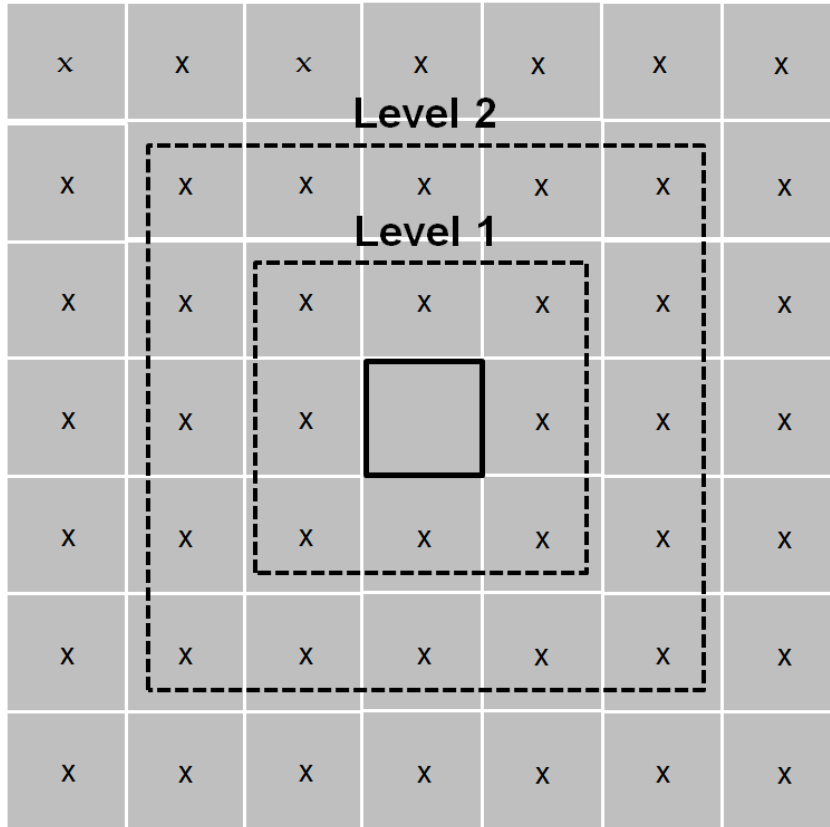


Figure 2.6 Fixing a missing value using neighboring cells.

In the above figure, the value of the target cell (black solid lines) is assumed missing, which is to be fixed or replaced with a value computed from the non-missing values pooled from its neighboring cells. The concept of pooling used in this study is as follows: Assume only two neighborhood levels are considered, the non-missing values of the eight closet cells (at Level 1) are pooled together. Then, the average or median can be assigned to the target cell for a numeric variable, or the mode can be used for a categorical variable. However, if the number of the pooled values from Level 1 is not enough (based on a user-specified threshold or cut-off), the non-missing values from Level 2 are additionally pooled and combined. If the number of all pooled values is still not found enough, the target cell is then assigned as missing.

- ESRI ArcInfo ASCII
(http://resources.esri.com/help/9.3/arcgisengine/java/GP_ToolRef/spatial_analyst_tools/esri_ascii_raster_format.htm) and
- Shapefile
(<http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf>).

All of the above data formats store raster data, except for the last one which stores vector data. Here, data manipulation (i.e., extraction, subsetting, format conversion, rasterization, and map re-projection) was done using standard libraries and utilities associated with these data formats, GDAL (Geospatial Data Abstraction Library) tools (<http://www.gdal.org>), MODIS HDF-EOS tools (<https://modis.gsfc.nasa.gov/tools>), and certain computer codes written specifically. Another data format, essential to GUPROC, is flat binary (FB), which stores data in a very simple manner. In FB, a matrix of numeric data is written out as one single record by properly setting the length of a data record and then dumping the data matrix to an output file, here as binary little-endian 4-byte (i.e., single-precision) real numbers. Additional matrices of the same matrix dimension can be sequentially written out as next records. Due to simplicity, reading and writing a FB file can be easily done by a code in a standard computer language (e.g., FORTRAN and C). Moreover, such FB data can be geo-referenced and graphically viewed, if desired, using GrADS (Grid Analysis and Display System) (<http://cola.gmu.edu/grads>).

2.7 Variables to be updated

GEOGRID generates many output variables, but only those related to surface properties or conditions are of interest and relevance here. Manomaiphiboon et al. (2016) updated certain GEOGRID output variables to enhance the quality of wind resource forecasting over a site on a hill top in Nakhon Ratchasima province. This study used Manomaiphiboon et al. as an initial guideline but extended to include more variables for updating. The

following is the GEOGRID output variables selected and considered in this study:

- Terrain height above mean sea level (HGT_M in GEOGRID, m), as a single data layer
- Dominant land use/land cover class (shortly, land use) (LU_INDEX in GEOGRID, categorical), as a single data layer. The land use classification of current interest is of the US Geological Survey (USGS) because many land surface models in WRF support or use this classification, at the time of study. It has a total of 24 classes as follows:

1. Urban and built-up land
2. Dryland cropland and pasture
3. Irrigated cropland and pasture
4. Mixed dryland/irrigated cropland and pasture
5. Cropland/grassland mosaic
6. Cropland/woodland mosaic
7. Grassland
8. Shrubland
9. Mixed shrubland/grassland
10. Savanna
11. Deciduous broadleaf forest
12. Deciduous needleleaf forest
13. Evergreen broadleaf forest
14. Evergreen needleleaf forest
15. Mixed forest
16. Water bodies
17. Herbaceous wetland
18. Wooden wetland
19. Barren or sparsely vegetated
20. Herbaceous tundra
21. Wooded tundra
22. Mixed tundra

23. Bare ground tundra and

24. Snow or ice

- Fraction of land use class (LANDUSEF in GEOGRID, fraction), as 24 single layers of data corresponding the above individual 24 land use classes
- Monthly leaf area index (LAI12M in GEOGRID, $\text{m}^2 \text{m}^{-2}$), as 12 single data layers corresponding to the individual months of year. By definition, leaf area index (LAI) is the ratio of the total one-sided green leaf area to the surface area, which is an essential parameter in land surface modeling to partition energy and water vapor in plant canopy.
- Monthly surface albedo (shortly, albedo) (ALBED012M in geogrid, %), as 12 single data layers corresponding to the individual months of year. By definition, it is the fraction of shortwave solar radiation reflected from the Earth's surface back to space, which is an essential parameter in land surface modeling, particularly partitioning energy at the surface.
- Dominant top-layer soil texture (SCT_DOM in GEOGRID, categorical), as a single data layer. Soil texture defined in WRF has a total of 16 types as follows:
 1. Sand
 2. Loamy sand
 3. Sandy loam
 4. Silt loam
 5. Silt
 6. Loam
 7. Sandy clay loam
 8. Silty clay loam
 9. Clay loam
 10. Sandy clay
 11. Silty clay
 12. Clay

- 13. Organic material
- 14. Water
- 15. Bedrock and
- 16. Others (land ice)

The first 13 types in the above list follow those in the US Department of Agriculture (USDA) classification (Figure 2.7). Soil physical parameters (mainly, wilting point, saturated soil moisture, field capacity, Campbell's porosity index, saturated soil water potential, and saturated soil moisture conductivity) are dependent on the type of soil texture. In land surface modeling, they affect soil thermal and hydrological processes, which in turn impact heat and moisture partitioning in soil and at the surface.

- Dominant bottom-layer soil texture (SCB_DOM in GEOGRID, categorical), as a single data layer, classified as the 16 soil texture types above
- Fraction of top-layer soil texture (SOILCTOP in GEOGRID, fraction), as 16 single data layers corresponding to the individual 16 soil texture types above
- Fraction of bottom-layer soil texture (SOILCBOT in GEOGRID, fraction), as 16 single data layers corresponding to the individual 16 soil texture types above and
- Land mask (LANDMASK in GEOGRID, 1: land and 0: water), as a single data layer.

(Intentionally left blank)

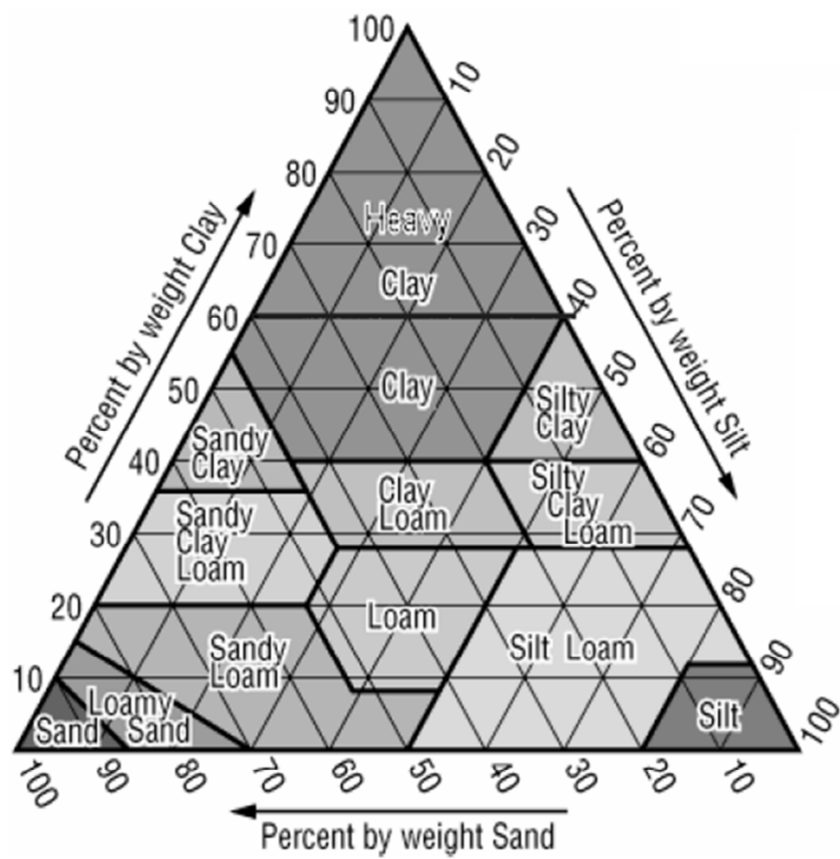


Figure 2.7 Soil texture classification by the USDA.

Source: FAO/IIASA/ISRIC/ISS-CAS/JRC (2012)

2.8 Terrain height data

ASTER GDEM (ASTER Global Digital Elevation Model) version 2 data (ASTER GDEM Validation Team 2011, <https://doi.org/10.5067/ASTER/ASTGTM.002>) were used as the sole source of the updating data. The data was developed jointly by US National Aeronautics and Space Administration (NASA) and the Ministry of Economy, Trade, and Industry (METI) of Japan. ASTER (Advanced Space-Borne Thermal Emission and Reflection Radiometer) is a sensor on board of Terra, polar-orbiting NASA satellite launched in 1999. ASTER GDEM version 2 represents all acquisitions since the beginning year 2000 till late 2010 (as opposed to version 1 representing only the years 2000–2008). The data has a 2.4-sec. (72 m) resolution but distributed or posted as a 1-sec. (30 m) resolution. The data were downloaded and aggregated on to the reference tiles. The default terrain height data provided in GEOGRID is GMTED2010 (Global Multi-Resolution Terrain Elevation Data 2010), jointly developed by the USGS and the National Geospatial-Intelligence Agency (NGA) (Danielson and Gesch 2011, <https://lta.cr.usgs.gov/GMTED2010>). Three resolutions (30 sec., 15 sec., and 7.5 sec.) are available in GMTED2010 but only the 30-sec. resolution is stored in GEOGRID. GMTED2010 is a major improvement over its predecessor GTOPO30 (Global 30 Arc-Second Elevation) (<https://lta.cr.usgs.gov/GTOPO30>) of the USGS, once stored as GEOGRID default prior to GMTED2010. Since ASTER-GDEM and GMTED2010 are more recent than GTOPO30, both are expected to have more compatibility in terms of accuracy and consistency, i.e., superior to GTOPO30. For the data processed and filled into the reference grid here, no cells with a missing value were found. However, if such a cell is present, one can fix it using the values of its neighboring cells.

2.9 Land use data

MODIS-derived data, developed by the NASA, were used as the primary source for land use outside Thailand. MODIS (Moderate-Resolution Imaging Spectroradiometer) is a sensor on board of each of Terra and Aqua, polar-orbiting NASA satellites launched in 1999 and 2002, respectively. The land use

product used here is MCD12Q1.051 MODIS/Terra and Aqua Land Cover Type Yearly L3 Global 500m SIN Grid Version 051 (Friedl et al. 2010, https://lpdaac.usgs.gov/dataset_discovery/modis/modis_products_table/mcd12q1).

Various MODIS land and atmosphere products are distributed as tiles in the sinusoidal map projection (Figure 2.8). The data have a 500-m resolution, are given as yearly, spanning 2001–2015. Here, the data for the year 2015 were used as the most recent data available at the time of study. The data were downloaded and resampled onto the reference tiles. Multiple land use classifications are provided inside the data, one of which is IGBP (International Geosphere–Biosphere Program). However, the USGS classification is not present. The IGBP classification was used since it has 17 classes (i.e., more than any other classifications given in the data) as follows:

1. Water
2. Evergreen needleleaf forest
3. Evergreen broadleaf forest
4. Deciduous needleleaf forest
5. Deciduous broadleaf forest
6. Mixed forest
7. Closed shrubland
8. Open shrubland
9. Woody savannas
10. Savannas
11. Grasslands
12. Permanent wetlands
13. Croplands
14. Urban and built-up
15. Cropland and natural vegetation mosaic
16. Snow and ice and
17. Barren or sparsely vegetated

To map the 17 IGBP to 24 USGS classes, a matching scheme was needed and developed (Table 2.1). As seen from the table, tree fraction is a parameter

additionally required to help partition “permanents wetlands” in IGBP into two wetland classes in USGS (wooded and herbaceous). Here, it was obtained from MOD44B MODIS/Terra Vegetation Continuous Fields (VCF) Yearly L3 Global 250m SIN Grid V006 (DiMiceli et al. 2017, <https://doi.org/10.5067/MODIS/MOD44B.006>). This data product contains surface cover percentages per pixel for the following three types: tree, non-tree, and non-vegetated (or bare). The data have a 250-m resolution in the sinusoidal projection and are given as yearly since 2000. The data for the year 2013 were used as the most recent data available at the time of study. The data were downloaded and resampled onto the reference tiles. To map “croplands” in IGBP “dryland, mixed dryland/irrigated, and irrigated cropland and pasture” in USGS, irrigated area data are needed. Global Map of Irrigation Areas (GMIA) version 5 data (Siebert et al. 2013, <http://www.fao.org/nr/water/aquastat/irrigationmap/>) were used for areas outside Thailand, whose posting resolution is 5 min. (9 km). For such areas within Thailand, a shapefile (based on a map scale of 1:50,000) developed by the Royal Irrigation Department was used here. The representative year of the data is 2013. The GMIA data were resampled to the reference tiles while the RID shapefile was first rasterized to the Thailand tiles and then rasterized data were later aggregated to the reference tiles.

For Thailand, land use data developed by the Land Development Department (LDD) were obtained and used as the primary source. The LDD land use data is of their recognized reliable or acceptable quality. The data obtained are in the shapefile format, based on a map scale of 1:25,000. It is noted that the LDD does not develop or revise the land use data for each province every year. The representative year of the data obtained for all 77 provinces of Thailand varies among 2009 (30 provinces), 2012 (18 provinces), 2013 (21 provinces), and 2015 (8 provinces) (Figure 2.9). The land use classification by the LDD has three levels (Levels 1-3), with Level 1 being the most aggregate and Level 3 being the most detailed. In the classification version 2014, Level 1 has five aggregate classes (urban and built-up land,

agricultural land, forest land, water body, and miscellaneous land), Level 2 has 33 classes, and Level 3 includes more than 200 classes. Mapping the Level-3 classes to the 24 USGS classes was performed here. A couple matching schemes were developed for different classification versions. As an example, Table 2.2 shows the matching scheme for the classification version 2014. After mapping, the shapefile data was rasterized to the Thailand tiles, and the rasterized data were later aggregated to the reference tiles. The MODIS and LDD land use data were merged to form the new land use data used for the updating. To fix any cell with a missing value present in the merged data, the mode was used, which was computed from the values pooled from its neighboring cells.

(Intentionally left blank)

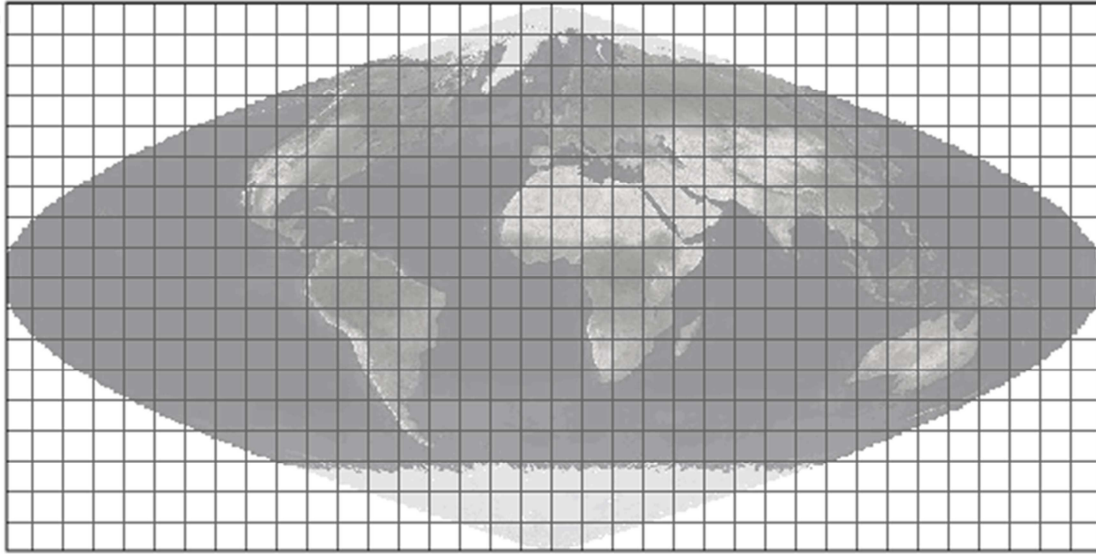


Figure 2.8 Sinusoidal-projected MODIS tiles.

Source: https://modis-land.gsfc.nasa.gov/MODLAND_grid.html

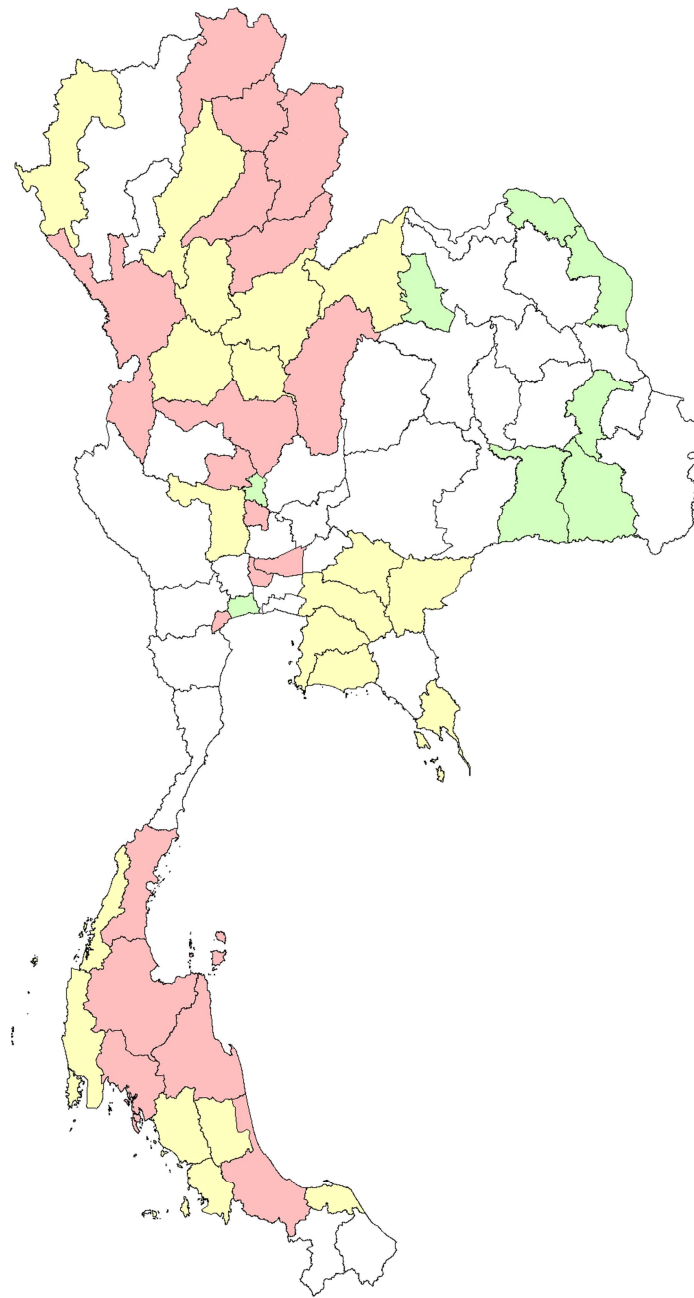


Figure 2.9 Representative years of the LDD land use data.

The years 2009, 2012, 2013, and 2015 are denoted by
white, red, yellow, green, respectively.

Table 2.1 Land use matching from IGBP to USGS classes.

IGBP	USGS
Water	Water bodies
Evergreen needleleaf forest	Evergreen needleleaf forest
Evergreen broadleaf forest	Evergreen broadleaf forest
Deciduous needleleaf forest	Deciduous needleleaf forest
Deciduous broadleaf forest	Deciduous broadleaf forest
Mixed forest	Mixed forest
Closed shrubland	Shrubland
Open shrubland	Mixed shrubland/grassland
Woody savannas	Savannas
Savannas	Savannas
Grasslands	Grassland
Permanents wetlands ^a	a. Wooded wetland b. Herbaceous wetland
Croplands ^b	a. Dryland cropland and pasture b. Mixed dryland/irrigated cropland and pasture c. Irrigated cropland and pasture
Urban and built-up	Urban and built-up land
Cropland and natural vegetation mosaic	Cropland/grassland mosaic
Snow and ice	Snow or ice
Barren or sparsely vegetated	Barren or sparsely vegetated

- a. To map permanent wetlands from IGBP to USGS, the tree fraction for a grid cell is estimated and. If fraction ≥ 0.5 “wooded wetland” is assigned to that grid cell. Otherwise (fraction < 0.5), “herbaceous wetland” is assigned.
- b. To map cropland from IGBP to USGS, the irrigated area fraction for a grid cell is inspected. When fraction < 30.0 “Dryland cropland and pasture” is assigned. If $30 \geq \text{fraction} < 70$, “Mixed dryland/irrigated cropland and pasture” is assigned. Otherwise (fraction ≥ 70), “Irrigated cropland and pasture” is assigned.

Table 2.2 Land use matching from LDD to USGS classes.

LDD	Description	USGS	LDD	Description	USGS
A000	Agricultural land	4	A221	Millet	2
A001	Integrated farm diversified farm	4	A222	Ginger	2
A100	Abandoned paddy field	2	A223	Cabbage	5
A101	Active paddy field	3	A224	Tomato	2
A200	Abandoned field crop	2	A225	Aloe vera	2
A201	Mixed field crop	2	A226	Agave	2
A202	Corn	2	A227	Paper mulberry	2
A203	Sugarcane	2	A228	Sunflower	2
A204	Cassava	2	A229	Chili	2
A205	Pineapple	2	A230	Wheat	5
A206	Tobacco	2	A231	Barley	5
A207	Cotton	2	A232	Rye	5
A208	Mungbean	2	A233	Opium	2
A209	Soybean	2	A234	Marijuana hemp	2
A210	Peanut	2	A235	Roselle	2
A211	Kenaf jute	2	A236	Taro	2
A212	Black bean red bean	2	A300	Abandoned perennial	15
A213	Sorghum	2	A301	Mixed perennial	15
A214	Castor bean	2	A302	Para rubber	6
A215	Sesame	2	A303	Oil palm	6
A216	Upland rice	5	A304	Eucalyptus	6
A217	Potato	2	A305	Teak	11
A218	Jicama	2	A306	Magosa	6
A219	Sweet potato	2	A307	Casuarina	14
A220	Watermelon	2	A308	Acacia	6

(1/5, continued on next page)

Table 2.2 (Continued)

LDD	Description	USGS	LDD	Description	USGS
A309	Padauk	11	A413	Longan	6
A310	Gmelina sp	11	A414	Guava	6
A311	Mangrove	18	A415	Papaya	6
A312	Coffee	9	A416	Jackfruit	6
A313	Tea	9	A417	Santol	6
A314	Mulberry	9	A418	Rose apple	6
A315	Bamboo	9	A419	Mangosteen	6
A316	Kapok	11	A420	Langsat	6
A317	Betel palm	6	A421	Rakum sala	6
A318	Rain tree	6	A422	Lime	6
A319	White cheesewood	6	A423	Subtropical fruit	6
A320	Croton sp	11	A424	Manila tamarind	6
A321	Indian mahogany	6	A425	Elaeocarpaceae	6
A322	Eagle wood	6	A426	Dragon fruit	6
A323	Bur flower tree	6	A427	Pomelo	6
A400	Abandoned orchard	2	A428	Sapodilla	6
A401	Mixed orchard	6	A429	Plummango	6
A402	Orange	6	A430	Burmese grape	6
A403	Durian	6	A431	Pomegranate	6
A404	Rambutan	6	A500	Abandoned horticulture	6
A405	Coconut	6	A501	Mixed horticulture	5
A406	Litchi	6	A502	Truck crop	5
A407	Mango	6		Floricultural	
A408	Cashew	6	A503	ornamental plant	5
A409	Jujube	6	A504	Grape	5
A410	Custard apple	6	A505	Pepper	5
A411	Banana	6	A506	Strawberry	5
A412	Tamarind	6	A507	Passion fruit	5
			A508	Raspberry	5

(2/5, continued on next page)

Table 2.2 (Continued)

LDD	Description	USGSG	LDD	Description	USGS
A509	Herbs	5	A615	Sesame shifting cultivation	2
A510	Grass plantation	7	A616	Upland rice shifting cultivation	5
A511	Rattan	5	A617	Potato shifting cultivation	2
A512	Cantaloupe	5	A618	Jicama shifting cultivation	2
A513	Okra	5	A619	Sweet potato shifting cultivation	2
A514	Asparagus	5	A620	Watermelon shifting cultivation	2
A515	Mushroom	5	A621	Millet shifting cultivation	2
A600	Swidden cultivation bush fallow	19	A622	Ginger shifting cultivation	2
A601	Mixed field crop shifting cultivation	2	A623	Cabbage shifting cultivation	5
A602	Corn shifting cultivation	2	A624	Tomato shifting cultivation	2
A603	Sugarcane shifting cultivation	2	A625	Aloe vera shifting cultivation	2
A604	Cassava shifting cultivation	2	A626	Agave shifting cultivation	2
A605	Pineapple shifting cultivation	2	A627	Paper mulberry shifting cultivation	2
A606	Tobacco shifting cultivation	2	A628	Sunflower shifting cultivation	2
A607	Cotton shifting cultivation	2	A629	Chilli shifting cultivation	2
A608	Mungbean shifting cultivation	2	A630	Wheat shifting cultivation	5
A609	Soybean shifting cultivation	2	A631	Barley shifting cultivation	5
A610	Peanut shifting cultivation	2	A632	Rye shifting cultivation	5
A611	Kenaf jute shifting cultivation	2	A633	Opium shifting cultivation	2
A612	Black bean red bean shifting cultivation	2	A634	Marijuana hemp shifting cultivation	2
A613	Sorghum shifting cultivation	2	A635	Roselle shifting cultivation	2
A614	Castor bean shifting cultivation	2	A636	Taro shifting cultivation	2

(3/5, continued on next page)

Table 2.2 (Continued)

LDD	Description	USGS	LDD	Description	USGS
A700	Abandoned farm house	1	F301	Dense mangrove forest	18
A701	Pasture	2	F400	Disturbed swamp forest	17
A702	Cattle farm house	1	F401	Dense swamp forest	17
A703	Poultry farm house	1	F500	Disturbed forest plantation	15
A704	Swine farm house	1	F501	Dense forest plantation	15
A800	Aquatic plant	17	F600	Disturbed agroforestry	15
A801	Mixed aquatic plant	17	F601	Dense agroforestry	15
A802	Reed	17	F700	Disturbed beach forest	15
A803	Lotus	17	F701	Dense beach forest	15
A804	Water caltrop	17	M000	Miscellaneous land	4
A805	Water chestnut	17	M100	Rangeland	9
A806	Water spinach	17	M101	Grass	7
A807	Water mimosa	17	M102	Shrub	8
A900	Abandoned aquacultural land	16	M103	Giant thorny bamboo	9
A901	Mixed aquacultural land	16	M200	Marsh swamp	17
A902	Fish farm	16	M201	Marsh swamp	17
A903	Shrimp farm	16	M300	Abandoned mine pit	19
A904	Crab shellfish farm	16	M301	Mine	19
A905	Crocodile farm	16	M302	Laterite pit	19
F000	Forest land	15	M303	Sand pit	19
F100	Disturbed evergreen forest	13	M304	Soil pit	19
F101	Dense evergreen forest	13	M305	Oil field	19
F200	Disturbed deciduous forest	11	M400	Other miscellaneous land	4
F201	Dense deciduous forest	11	M401	Material dump	19
F300	Disturbed mangrove forest	18	M402	Landslide	19

(4/5, continued on next page)

Table 2.2 (Continued)

LDD	Description	USGG	LDD	Description	USGS
M403	Rock out crop	19	U500	Abandoned industrial land	1
M404	Dummy	19	U501	Industrial estate	1
M405	Landfill	19	U502	Factory	1
M500	Abandoned salt flat	19	U503	Agricultural product trading center	1
M501	Salt flat	19	U600	Abandoned area	1
M600	Beach	19	U601	Recreation area	1
M601	Beach	19	U602	Resort hotel guesthouse	1
M700	Garbage dump	19	U603	Cemetery	1
M701	Garbage dump	19	U604	Refugee camp	1
U000	Urban built up	1	U605	Gasoline station	1
U100	City town commercial	1	U700	Abandoned golf course	7
U101	City town commercial	1	U701	Golf course	7
U200	Abandoned village	1	W000	Water body	16
U201	Village	1	W100	Natural water body	16
U202	Hill tribe village	1	W101	River canal	16
U203	Moken village	1	W102	Lake lagoon	16
U300	Abandoned institutional land	1	W103	Ocean	16
U301	Institutional land	1	W200	Artificial water body	16
U400	Abandoned communication utility	1	W201	Reservoir	16
U401	Airport	1	W202	Farm pond	16
U402	Railway station	1	W203	Irrigation canal	16
U403	Bus station	1			
U404	Harbor	1			
U405	Road	1			
U406	Railway	1			

(5/5)

2.10 Monthly green fraction data

Green fraction represents how much vegetated area is present. Values of this parameter vary with time, monthly or seasonally. It is not possible to utilize the MODIS VCF data (see Section 2.9) since the data are available only on a yearly basis. Nevertheless, satellite-derived NDVI (normalized difference vegetation index) data can be used to compute green fraction. An NDVI-based estimation method proposed by Gutman and Ignatov (1997 and 1998) was employed here, as in Manomaiphiboon et al. (2016) and Paton and Manomaiphiboon (2013). Details of the method are not given here but those interested are referred to the original work. By definition,

$$\text{NDVI} = \frac{r_{\text{NIR}} - r_{\text{red}}}{r_{\text{NIR}} + r_{\text{red}}}, \quad (2.3)$$

where r_{NIR} and r_{red} are the reflectances in the near-infrared and red parts of the light spectrum, respectively. Chlorophyll in green vegetation generally absorbs visible light (0.4–0.7 μm) for photosynthesis while cell walls reflect near-infrared light (0.7–1.1 μm). Hence, the more leaves a live plant has, the more absorption and reflectance occur over these wavelengths, respectively. Values for NDVI range from -1 to 1. The larger positive values, the higher live green vegetation densities. For non-vegetated areas (e.g., bare soil, snow, and ice), NDVI typically has values of -0.1 to 0. Here, NDVI data was from MOD13A1 MODIS/Terra vegetation indices 16-day L3 global 500m SIN Grid V006 (Didan 2015, <https://doi.org/10.5067/MODIS/MOD13A1.006>). The data product contains a number of vegetation indices (including NDVI). The data have a 500-m resolution in the sinusoidal projection and a temporal interval of 16 days (non-overlapping), and span from the year 2000. The data for the year 2015 were used as the most recent data available at the time of this study. The data were downloaded, resampled onto the reference tiles, and arranged into individual months as weighted-average NDVI values, and converted to monthly green fraction using the method of Gutman and Ignatov. In GEOGRID, two default green fraction datasets are available. The first one has a 0.144° resolution (15.6 km). With no meta-data found, its representative period is

guessed to be about 1990s. The other one is MODIS-based with a 30-sec. resolution. As with the previous dataset, no meta-data was found. Nevertheless, its representative period is assured to be recent (e.g., 2000s or later). For the data processed and filled into the reference grid here, no cells with a missing value were found. However, if such a cell is present, one can fix it by determining the average green fraction value by land use class and by month, and the missing value can be replaced with the average green value corresponding to the land use class of that cell in the month considered.

2.11 Monthly leaf area index data

MODIS-derived LAI data were used, which is MCD15A2H MODIS/Terra+Aqua leaf area index/FPAR 8-day L4 Global 500m Sin Grid V006 (Myneni and Park 2015, <https://doi.org/10.5067/MODIS/MCD15A2H.006>). The data have a 500-m resolution in the sinusoidal projection and a temporal interval of 8 days (non-overlapping), and span from the year 2002. The data for the year 2015 were used as the most recent data available at the time of study. The data were downloaded, resampled onto the reference tiles, and arranged into individual months as weighted-average LAI values. In GEOGRID, one default dataset is given, which is MODIS-based, with a 30-sec. resolution with no meta-data found. Nevertheless, its representative period is assured to be recent (e.g., 2000s or later). For the data processed and filled into the reference grid here, a number of cells with a missing value were found and fixed. The concept of fixing such a cell is first to determine the average value by land use class and by month. The missing value is then replaced with the average value corresponding to the land use class of that cell in the month considered.

2.12 Monthly surface albedo data

MODIS-derived data is the primary source of the new albedo data used in this study, which is MCD43A1_v006 MODIS/Terra and Aqua Albedo daily L3 Global

500m (Schaaf 2015, <https://doi.org/10.5067/MODIS/MCD43A3.006>). This data product is generated on a daily basis but processed from satellite data retrieved over 16 days. It includes directional-hemispherical reflectance (i.e., black-sky albedo or $\alpha_{\text{black-sky}}$) and bi-hemispherical (diffuse) reflectance (i.e., white-sky albedo or $\alpha_{\text{white-sky}}$) at local solar noon in seven spectral and three broad bands. The data have a 500-m resolution in the sinusoidal projection and span from the year 2000. The data for the year 2015 were used as the most recent data available at the time of study. To calculate monthly albedo (i.e., blue-sky albedo or $\alpha_{\text{blue-sky}}$), the 8th, 12th, 19th, and 22nd days of the month were representatively used, and the data corresponding to these four days were downloaded. The reason of selecting only four days was simply disk-space saving. To find monthly black-sky or white-sky albedo, values on the selected days were pooled together and directly averaged as monthly. Monthly blue-sky albedo was computed as the weighted sum of the black-sky and white-sky albedo (Lucht et al. 2000):

$$\alpha_{\text{blue-sky}}(\theta, \lambda) = (1 - S(\theta, \lambda)) \times \alpha_{\text{black-sky}}(\theta, \lambda) + S(\theta, \lambda) \times \alpha_{\text{white-sky}} , \quad (2.4)$$

where S is the fraction of diffuse incident radiation and is a function of the solar zenith angle (θ) and wavelength (λ). To determine S , monthly solar diffuse radiation data from the NASA's MERRA-2 (Modern Era Retrospective-Analysis for Research and Analysis, version 2) (Gelaro et al. 2017) were used to provide both diffuse and direct components of incident radiation in the photosynthetically active radiation (PAR), and near-infrared (NIR) spectral ranges. The formula to compute S is as follows:

$$S = \frac{PAR_{\text{df}} + NIR_{\text{df}}}{PAR_{\text{df}} + PAR_{\text{dr}} + NIR_{\text{df}} + NIR_{\text{dr}}} , \quad (2.5)$$

where PAR_{df} and PAR_{dr} are the diffuse and direct radiation in the PAR spectral range, respectively, while NIR_{df} and NIR_{dr} are the diffuse and direct radiation

in the NIR spectral range, respectively. Since MODIS albedo data is quantified only for land surface, any cells designated as surface water was here assigned a WRF-default value of 0.08. For the data processed and filled into the reference grid here, a number of cells with a missing value were found and fixed. The concept of fixing such a cell is first to determine the average value by land use class and by month. The missing value is then replaced with the average value corresponding to the land use class of that cell in the month considered.

2.13 Soil texture

Soil texture data was obtained from the Harmonized World Soil Database version 1.2 (HWSD) (FAO/IIASA/ISRIC/ISSCAS/JRC 2012, <http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/harmonized-world-soil-database-v12/en/>). The HWSD was synthesized by harmonizing and combining existing regional and national soil datasets worldwide, and data quality in different regions or countries may vary. The data are available in a raster format, with a 30-sec. resolution. For South Asia and Southeast Asia, the main data sources are FAO (1977) and FAO (1979), respectively, data from which may potentially be old or not up-to-date. Specifically, for Thailand, the soil information integrated into the HWSD is from the Soil Survey Division of LDD and pertains to the years 1972–1973. In the HWSD, a total of 13 types are designated as listed below:

1. Clay (Heavy)
2. Silty clay
3. Clay
4. Silty clay loam
5. Clay loam
6. Silt
7. Silty loam
8. Sandy clay
9. Loam

10. Sandy clay loam
11. Sandy loam
12. Loamy sand and
13. Sand

To map the 13 USDA classes to 17 classes used in WRF, a matching scheme was developed and applied (Table 2.3). To fix a cell with a missing value, we first tried to assign the mode value from its neighboring cells. If there was still any cell with a missing value present, we assigned the default GEGORID soil texture code to that cell. Like land use, to fix any cell with a missing value, the mode of the values pooled from its neighboring cells was used to replace the missing value. Nevertheless, it was found that there are too many missing values over large areas over the land part of the reference domain, which was not possible to fix all of them with the neighborhood method. Therefore, any remaining missing cells were directly gap-filled using GEOGRID default output values.

Table 2.3 Soil texture matching from USDA to WRF classes.

USDA	WRF
Clay (Heavy)	Clay
Silty clay	Silty clay
Clay	Clay
Silty clay loam	Silty clay loam
Clay loam	Clay loam
Silt	Silt
Silty loam	Silty loam
Sandy clay	Sandy clay
Loam	Loam
Sandy clay loam	Sandy clay loam
Sandy loam	Sandy loam
Loamy sand	Loamy sand
Sand	Sand

Chapter 3

GUPROC and Illustrated Results

This chapter provides the technical description of GUPROC. An illustrated example of GUPROC implementation is also given, in which the WRF modeling domains of the HAI forecasting operation (Figure 2.2) are updated and their GEOGRID output results using the default and new surface data are compared.

3.1 GUPROC

GUPROC represents a suite of computer programs developed in order to update GEOGRID output variables with new surface data. The basis of updating is offline (see Section 2.1). The file structure of GUPROC begins with *geogrid_update/* as the root folder. Inside it, there are five main folders where all source codes, scripts, and input/output data and files of GUPROC are stored:

```
geogrid_update/  
├── domain/  
├── gis/  
├── io_default/  
├── io_update/  
└── proc/
```

3.2 Folder *domain/*

This folder contains the basic information of all WRF modeling domains considered in this study. Its file structure is as follows:

```
domain/  
├── domain_eastth_lcc  
└── domain_haii_merc
```


/— domain_tmd_merc
/— eastth_lcc/
/— haii_merc /
/— tmd_merc/

Inside *domain/*, three text files are present, which are *domain_eastth_lcc*, *domain_haii_merc*, and *domain_tmd_merc*. *domain_haii_merc* contains the configuration information of the three WRF modeling domains used in the HAI forecast operation. They have 27-km, 9-km, and 3-km resolutions, respectively, in a Mercator projection centered on Thailand (Figure 2.2). *domain_tmd_merc* contains the same information but for two modeling domains used in the TMD forecasting operation, whose resolutions are 30 km and 10 km, respectively, in a Mercator projection centered on Thailand. *domain_eastth_lcc* is for three modeling domains with resolutions of 18 km, 6 km, and 2 km, respectively, in a Lambert conformal conic projection centered on Eastern Thailand. In fact, they are not used in any forecasting operations but given here for the purpose of example only. It is noted that, for now, GUPROC supports these two map projections only on an assumed spherical earth with a 6,370-km radius, which is the value that WRF assumes in modeling. In addition to the three text files, three folders (*haii_merc/*, *tmd_merc/*, and *eastth_lcc/*) are present, each of which contains the GEOGRID output files in NETCDF, generated by running GEOGRID with the default surface data. These NETCDF files are to be used as input to GUPROC when updating with the new surface data.

3.3 Folder *gis/*

This folder stores various shapefiles and images. Two important text files are *tileinfo_ref168t* and *tileinfo_th76t*, which contains the configuration information of the reference tiles and the Thailand tiles, respectively. A number of shapefiles as well as the two text files are also present and used as input to certain processing tasks in GUPROC.

3.4 Folder *io_default/*

This folder contains GEOGRID default surface data, available at http://www2.mmm.ucar.edu/wrf/users/download/get_sources_wps_geog.html. Of note, the data are not used by GUPROC but are given only for reference purposes.

3.5 Folder *io_update/*

This folder contains all new surface data compiled for updating and also the final destination of GEOGRID output files after the completion of updating. Its file structure is as follows (only 2 levels down are shown):

io_update/

└─ *init/*

└─ *albedo_modis/*

└─ *irrig_gmia/*

└─ *irrig_rid/*

└─ *lai_modis/*

└─ *lulc_ldd/*

└─ *lulc_modis/*

└─ *rad_merra/*

└─ *stxt_hwsd/*

└─ *ter_aster/*

└─ *vcf_modis/*

└─ *veg_modis/*

└─ *refbin/*

└─ *albedo/*

└─ *lai/*

└─ *lulc/*

└─ *soil/*

└─ *ter/*

└── *vf/*

The folder *init/* contains all downloaded data required for processing, which are arranged into the following folders: *albedo_modis/* for the MODIS albedo data, *irrig_gmia/* for the GMIA irrigated-area data, *irrig_rid/* for the RID irrigated-area data, *lai_modis/* for the MODIS LAI data, *lulc_ldd/* for the LDD land use data of Thailand, *lulc_modis/* for the MODIS land use data, *rad_merra/* for the MERRA solar diffuse radiation data, *stxt_hwsd/* for the HWSO soil texture data, *ter_aster/* for the ASTER-GDEM terrain height data, *vcf_modis/* for the MODIS VCF data, and *veg_modis/* for the MODIS vegetation indices data.

As for *refbin/*, it contains data files in the FB format, as the final output generated from *io_update/s01.prep_data/*. The files are arranged onto the reference tiles. Like *init/*, the data are arranged by variable as follows: *albedo/* for albedo (monthly), *lai/* for LAI (monthly), *lulc/* for land use, *soil/* for soil texture, *ter/* for terrain height, and *vf/* for green fraction (monthly).

3.6 Folder *proc/*

This folder represents the core of GUPROC, where almost all data manipulation and processing, and GEOGRID updating take place. Its file structure is as follows:

proc/

└── *s01.prep_data/*

└── *s02.fill_domain/*

└── *s03.update2wrf_nc/*

Through *s01.prep_data/*, the data stored in *io_update/init/* are processed and arranged onto the reference tiles, and finally placed in

geogrid_update/io_update/refbin/. The file structure of this folder is as follows:

```
s01.prep_data/
├── albedo/
├── lai/
├── land_ref168t/
├── land_ref76t/
├── lulc/
├── soil/
├── ter/
├── vcf/
└── vf/
```

Each of the above folders pertains to each of the variables considered for processing, except for *land_ref168t/* and *land_th76t/* which contain land mask data for individual reference tiles and individual Thailand tiles, respectively.

Terrain height is processed in *ter/*. The file structure of *ter/* is as follows:

```
ter/
├── s01.unzip/
├── s02.tiff2bin/
├── s03.aggregate_200m/
└── s04.ter_tile/
```

The above folders perform the following tasks in sequence:

1. Select the ASTER-GDEM GeoTiff data that fit the reference domain
2. Rasterize the results from the previous step to the FB format using GDAL and FORTRAN
3. Aggregate the results from the previous step to a 0.002° resolution, and

4. Arrange the results from the previous step onto the reference tiles.

Land use is processed in *lulc/*. The LDD data is processed in the first step, followed by the MODIS data. The last step is to combine the LDD (for areas within Thailand) and MODIS (for areas outside Thailand) data altogether. The file structure of *lulc/* is as follows:

```
lulc/
├── s01. ldd/
├── s01. shp_tile/
├── s02. scheme_ldd2usgs /
├── s03. ldd_tile/
├── s04. usgs_tile/
├── s05. rasterize/
├── s06. comb_prune/
├── s07. aggregate_200m/
├── s02. modis/
├── s01. hdfmerge2bin/
├── s02. fix_undef_by_mode/
├── s03. rid_irrig/
├── s04. lulc_tile/
├── s05. modis2usgs/
├── s03. ldd_modis_comb/
```

For *s01. ldd/*, the following tasks are performed in sequence:

1. Partition the original LDD shapefiles (of the year 2009 and years 2012–2015) into shapefiles corresponding to the Thailand tiles.
2. Prepare matching tables for mapping the LDD classes to the USGS classes
3. Match LDD codes with USGS codes
4. Add USGS codes into the tile-wise LDD shapefiles
5. Rasterize the shapefiles from the previous step to the FB format using GDAL and FORTRAN

6. Combine the results (of 2009 and those of 2012–2015 altogether) from the previous step and then clean the results by keeping only non-sea grid cells. Note that, for each grid cell, first consider the data of 2012–2015 for use if available; otherwise, use those of 2009. and
7. Aggregate the results from the previous step to a 0.002° resolution and arrange them onto the reference tiles.

For *s02.modis/*, the following tasks are performed in sequence:

1. Select the data that fit the reference domain and convert them to the FB format
2. Fix any grid cell with a missing value using the mode value from its neighboring cells
3. Partition the original RID irrigated-area shapefile onto the Thailand tiles, convert the results to FB format, and clean the results by keeping only non-sea grid cells
4. Arrange the results from the previous step onto the reference tiles, and
5. Convert the 17 MODIS IGBP classes to the 24 USGS land use classes using the developed matching scheme (Table 2.1).

For *s03.1dd_modis_comb/*, it is to combine the results from *s01.1dd/* and *s02.modis/* by replacing the MODIS-based data with LDD-based data only for areas inside Thailand.

Green fraction is processed in *vf/*, whose file structure is as follows:

vf/

└── *s01.hdfmerge2bin/*

└── *s02.monthly_average/*

└── *s03.ndvi2vf/*

└── *s04.vf_tile/*

The above folders perform the following tasks in sequence:

1. Select only the original MODIS NDVI tiles that fit the reference grid, extract, merge the tiles, and convert them to the FB format
2. Compute monthly average NDVI and save output in the FB format
3. Compute monthly vegetation fraction using the results from the previous step and
4. Arrange the results from the previous step onto the reference tiles.

LAI is processed in *lai/*, whose file structure is as follows:

lai/

```
├── s01.hdfmerge2bin/
├── s02.monthly_average/
├── s03.lai_tile/
├── s04.fix_undef_by_lulc/
```

The above folders perform the following tasks in sequence:

1. Select only the MODIS LAI tiles that fit the reference grid, merge the tiles, and convert them to the FB format
2. Calculate monthly average LAI and save output in the FB format
3. Arrange the results from the previous step onto the reference tiles, and
4. Calculate average LAI for each LULC type found over the reference domain. Fix any grid cell with a missing value by assigning the average LAI value for the land use class corresponding to the grid cell.

ALBEDO is processed in *albedo/*, whole structure is as follows:

albedo/

```
├── s01.merra_nc2bin/
├── s02.diffuse_frac/
├── s03.hdfmerge2bin/
├── s04.bsa_wsa/
├── s05.albedo/
├── s06.albedo_tile/
```

└── *s07.fix_undef_by_lulc/*

The above folders perform the following tasks in sequence:

1. Extract NIRDF (Surface Downward NIR Diffuse Flux), NIRDR (Surface Downward NIR Direct Flux), PARDF (Surface Downward PAR Diffuse Flux), and PARDR (Surface Downward PAR Direct Flux) variables from the monthly MERRA-2 data and save in the FB format
2. Calculate diffuse fraction using the variables from the previous step and save output in the FB format
3. Extract black-sky albedo and white-sky albedo from the MODIS data. Here, select only the MODIS tiles that fit the reference grid, merge the tiles, and convert them to the FB format
4. Calculate monthly average black-sky albedo and white-sky albedo
5. Calculate monthly average albedo using the diffuse fraction, black-sky albedo and white-sky albedo data from steps 2 and 4
6. Rearrange the results from the previous step onto the reference tiles, and
7. For a particular month, calculate average albedo for each land use class and then fix any grid cell with a missing value by assigning the average albedo value for the land use class corresponding to the grid cell.

Soil texture is processed in *soil/*, whose file structure is as follows:

soil/

└── *s01.extract/*

└── *s02.soil_tile/*

└── *s03.hwsd2wrf/*

└── *s04.fix_undef_by_mode/*

The above folders perform the following tasks in sequence:

1. Extract the top-layer and bottom-layer soil data from the HWSD database

2. Rearrange the extracted soil data onto the reference tiles and save output in the FB format
3. Change default HWSO soil texture codes to soil texture codes used by WRF, and
4. Fix any grid cell with a missing value using the mode value from its neighboring cells.

Once all the tasks in *proc/s01.prep_data/* are completed, *proc/s02.fill_domain/* is next implemented, whose file structure is as follows:

```

s02.fill_domain/
├── s01.qc_undef/
├── s02.fill/
│   ├── lcc/
│   ├── s01.lulc/
│   ├── s02.vf/
│   ├── s03.lai/
│   ├── s04.ter/
│   ├── s05.albedo/
│   ├── s06.soil_bottom/
│   ├── s07.soil_top/
│   ├── s08.land_mask/
│   └── merc/
│       ├── s01.lulc/
│       ├── s02.vf/
│       ├── s03.lai/
│       ├── s04.ter/
│       ├── s05.albedo/
│       ├── s06.soil_bottom/
│       ├── s07.soil_top/
│       └── s08.land_mask/

```

s01.qc_undef/ is only to quality-check the final results obtained from *proc/s01.prep_data/* by counting the amount of missing values present in each variable. *s02.fill/* is to fill those results onto user-specified modeling domains one variable at a time, according to which map projection is desired (here, Mercator or Lambert conformal conic). For land use, top-layer soil texture, and bottom-layer soil texture, to find the dominant category for a target cell uses the simple bin filling with cell assignment as mode whereas to find the fraction of a category for a target cell also uses the same method but with cell assignment as the ratio of initial cells from that category found in the target cell to those from all categories. For green fraction, LAI, terrain height, and albedo, the simple bin filling is employed with cell assignment as average. In the end, LANDMASK is additionally created as a binary variable (1 as land and 0 as water) using the “water” class in the land use data.

The final step of *proc/* is to replace default values in the GEIGRID output files with new or updated ones, which is carried out by *s03.update2wrf_nc/*. Its file structure is simple as follows:

```

└── s03.update2wrf_nc/
    ├── lcc/
    └── merc/

```

Both *lcc/* and *merc/* share the same contents but support different map projections. They contain a number of FORTRAN codes perform replacement, each of which corresponds to a variable considered for updating. They were written based primarily on `read_wrf_nc.f90`, WRF utility program available at <http://www2.mmm.ucar.edu/wrf/users/utilities/util.htm>.

3.7 Illustrated results

A number of maps from GEOGRID output generated using the default and updated surface data are illustrated and compared. It is noted that the

GEOGRID default output files of the three modeling domains were directly obtained from the HAIL, and GUPROC was run on the files for updating. In this chapter, the results of the third (i.e., finest) modeling domain are shown and discussed for four regions covering entire Thailand, which are Northern, Northeastern, Central & Eastern, and Southern (east and west coasts combined), based on the five climatic regions of the TMD (TMD 2017) (Figure 3.1). The domain-wide results of all individual three modeling domains are also given as supplement and arranged in Appendix. For terrain height (Figures 3.2–3.5), GTOPO30 and GMTED2010 are the default data, and ASTER-GDEM is the new data (see Chapter 2). From the figures, the default and updated maps appear to be similar in terms of pattern and magnitude in every region. However, the difference maps (with respect to ASTER-GDEM) show the presence of non-negligible differences (>20 m in magnitude) over many areas, especially over mountainous areas. Nevertheless, the degree of difference appears to decrease substantially for GMTED2010 (recent), compared to GTOPO30 (old). As for land use (Figures 3.6–3.9), the results bear apparent discrepancy in every region. For example, “urban and built-up” in the Central & Eastern regions (Figure 3.8) is well present in the updated map but is not in the default map. In the Southern region (Figure 3.9), many areas designated as “irrigated cropland and pasture” or “deciduous broadleaf forest” in the default map become “cropland/woodland mosaic” in the updated map. For monthly variables (here, green fraction, LAI, and albedo), only results of April and October are representatively shown for the dry and wet seasons, respectively. For green fraction (Figures 3.10–3.13), it is obvious that the updated maps is superior to the default maps that appear to lack adequate spatial details due to a low (or coarse) resolution of the default surface data. The updated maps satisfactorily reflect intensified green fraction in October (i.e., the wet season). They are also consistent with the updated land use maps because both variables are from MODIS-based derivation. For example, “urban and built up” areas in Central & Eastern regions (Figure 3.8) have relatively low green fraction, particularly in the wet season (Figure 3.12d). Similar results and reasons apply to LAI (Figures 3.14–3.17) and albedo (Figures 3.18–3.21). For

top-layer soil texture (Figures 3.22–3.25) and bottom-layer soil texture (Figures 3.26–3.29), the updated maps are dominantly but unrealistically designated as “loamy sand” with much less spatial variation in every region. These are in quite contrast with those found in the default maps and possibly attributed to the soil data inventorized in the HWSO data for Thailand and its neighboring countries being relatively old and then not as representative.

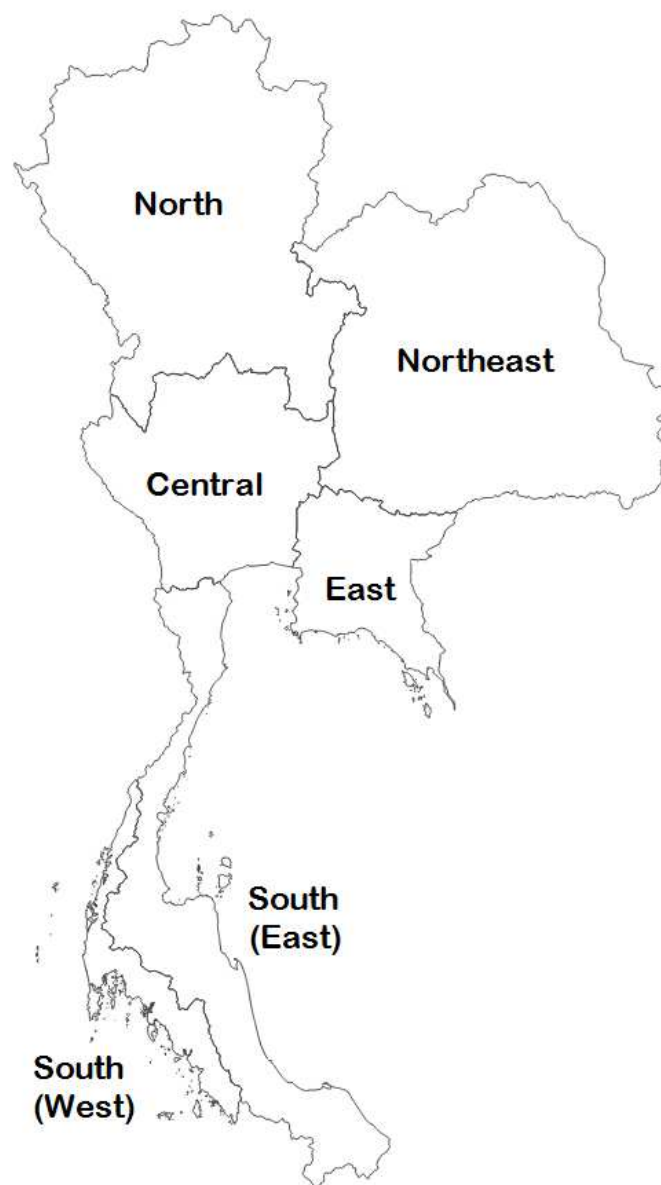
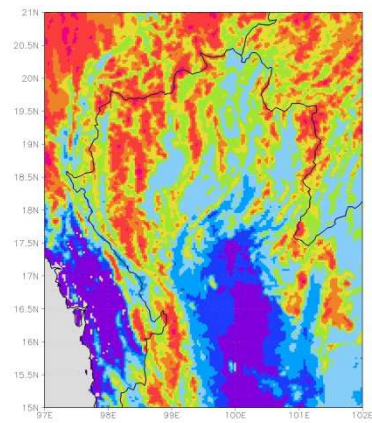
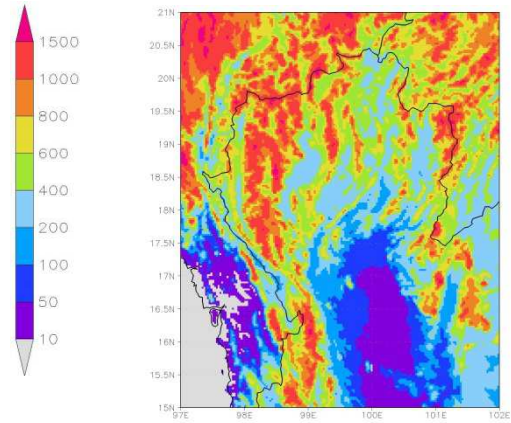


Figure 3.1 Regions of Thailand according to the TMD.

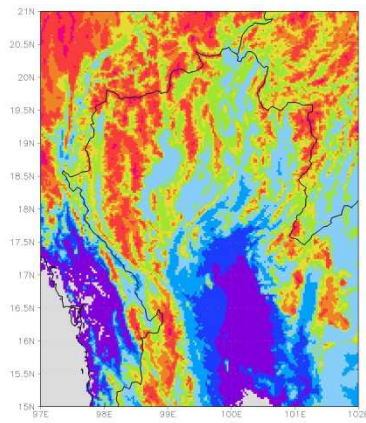
a) Default GTOP030



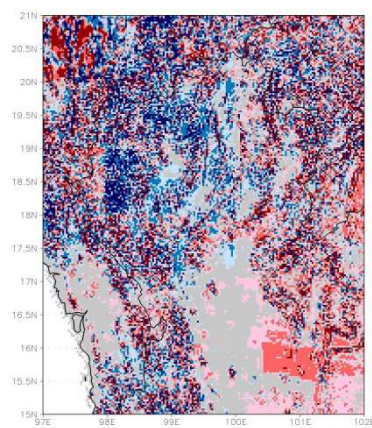
b) Default GMTED2010



c) Updated ASTER-GDEM



d) GTOP030 minus ASTER-GDEM



e) GMTED2010 minus ASTER-GDEM

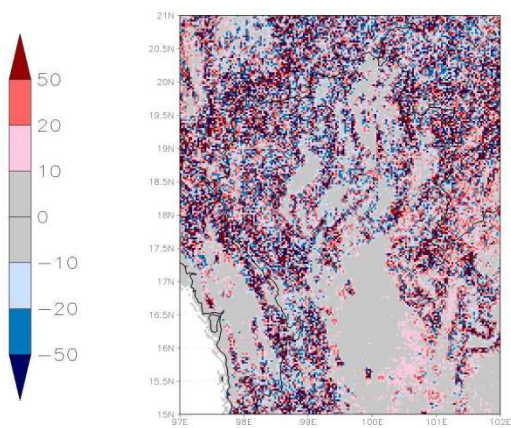
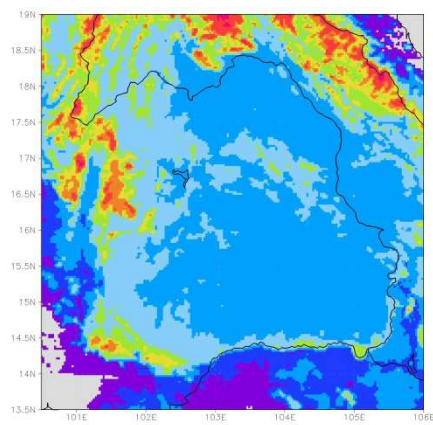
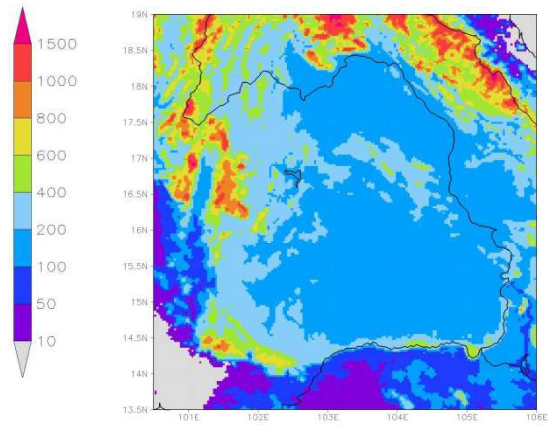


Figure 3.2 Terrain height (m) in Northern region.

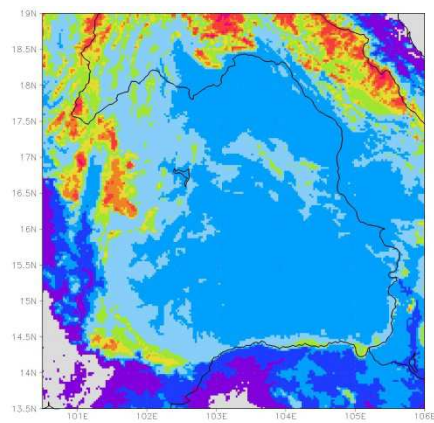
a) Default GTOP030



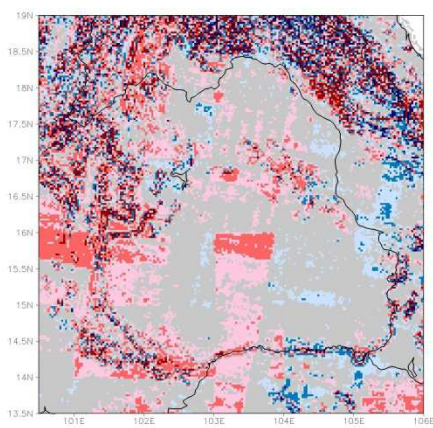
b) Default GMTED2010



c) Updated ASTER-GDEM



d) GTOP030 minus ASTER-GDEM



e) GMTED2010 minus ASTER-GDEM

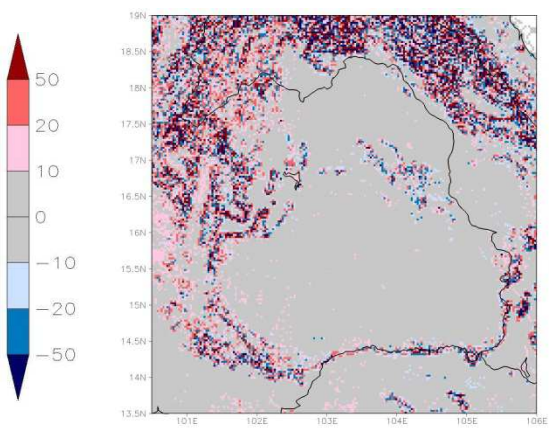
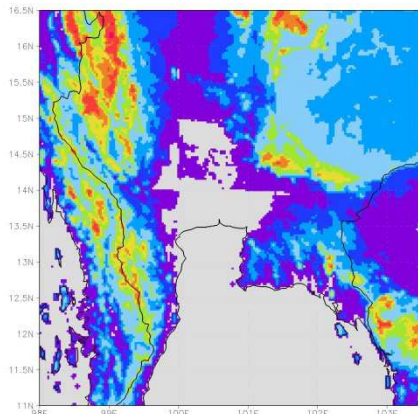
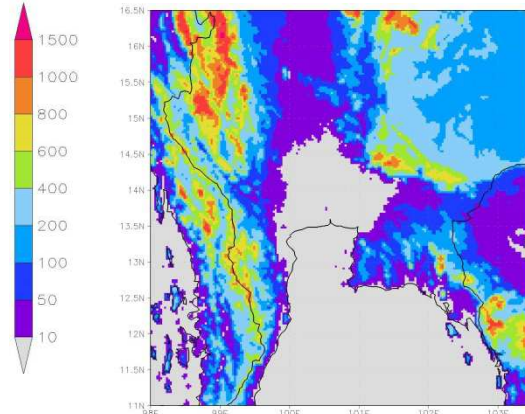


Figure 3.3 Terrain height (m) in Northeastern region.

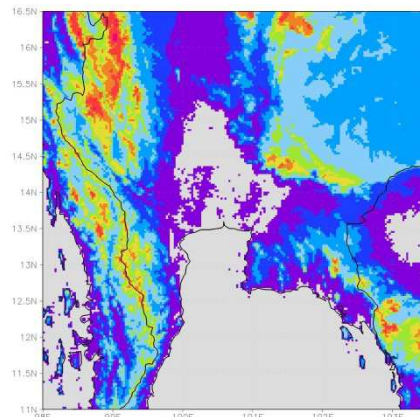
a) Default GTOP030



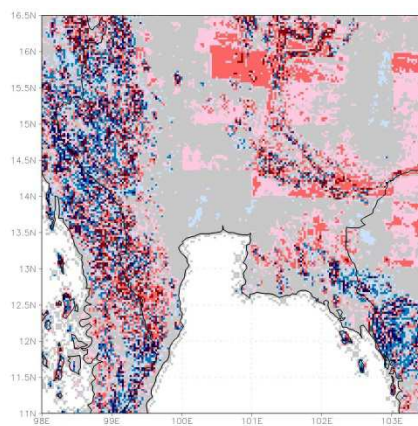
b) Default GMTED2010



c) Updated ASTER-GDEM



d) GTOP030 minus ASTER-GDEM



e) GMTED2010 minus ASTER-GDEM

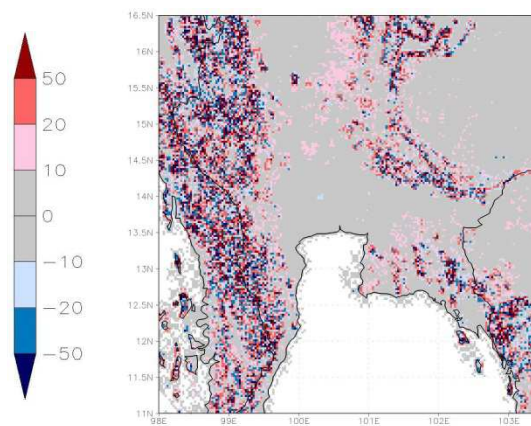
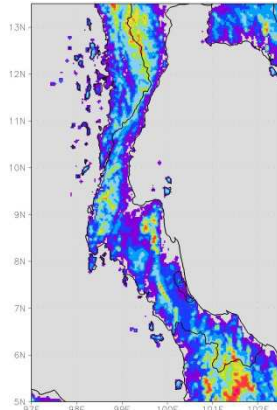
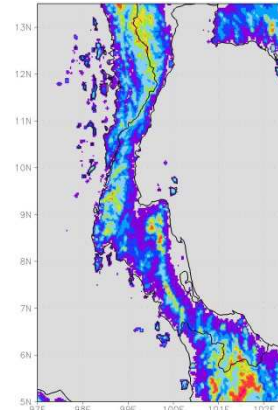


Figure 3.4 Terrain height (m) in Central & Eastern regions.

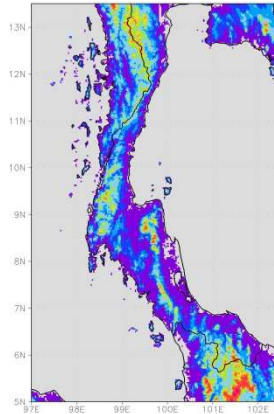
a) Default GTOP030



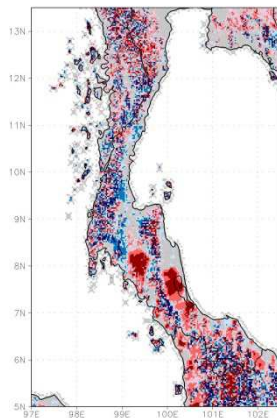
b) Default GMTED2010



c) Updated ASTER-GDEM



d) GTOP030 minus ASTER-GDEM



e) GMTED2010 minus ASTER-GDEM

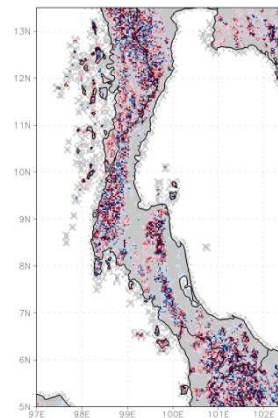
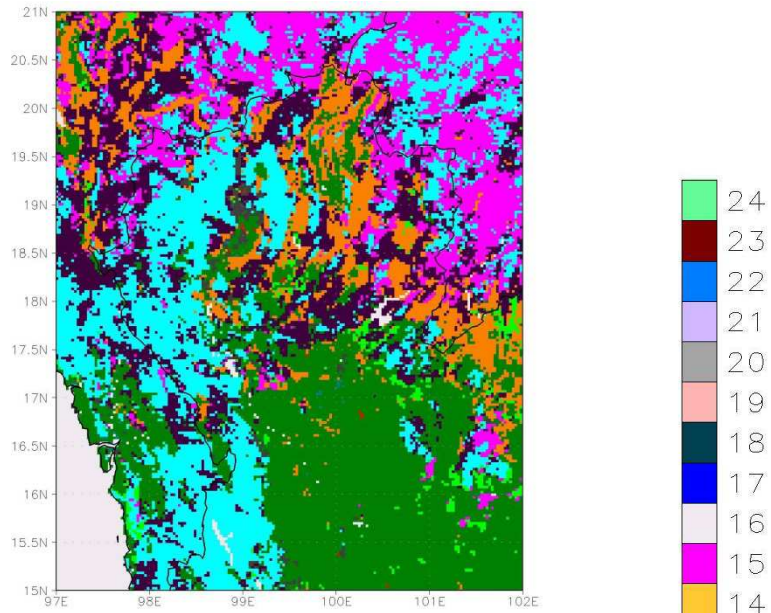


Figure 3.5 Terrain height (m) in Southern region.

a) Default



b) Updated

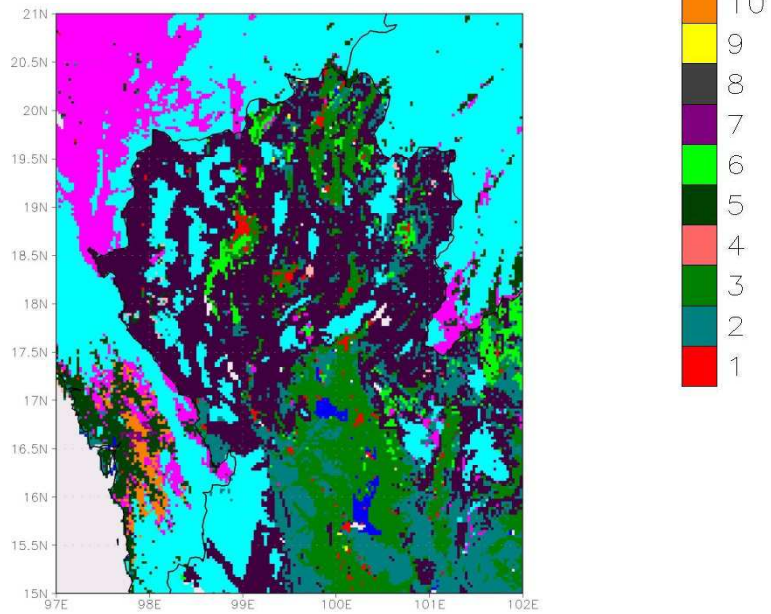
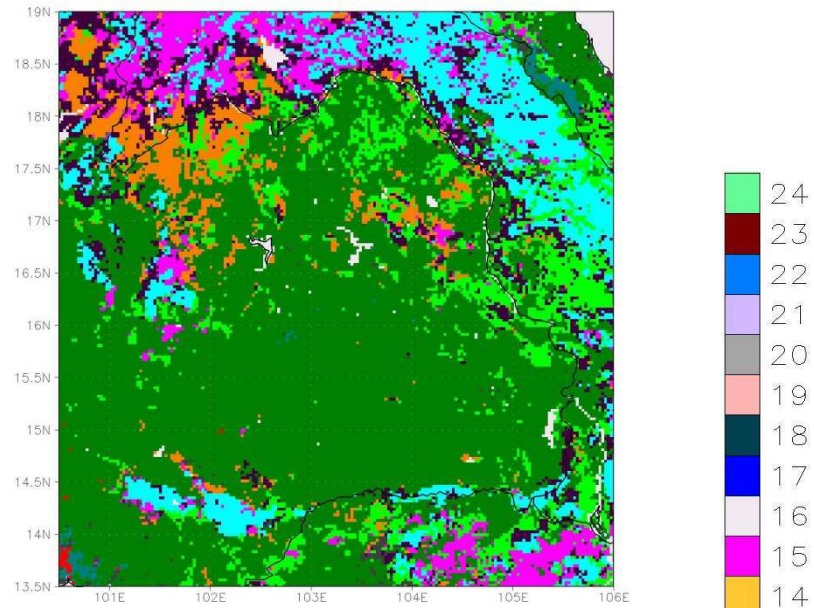


Figure 3.6 Dominant land use class in Northern region.

a) Default



b) Updated

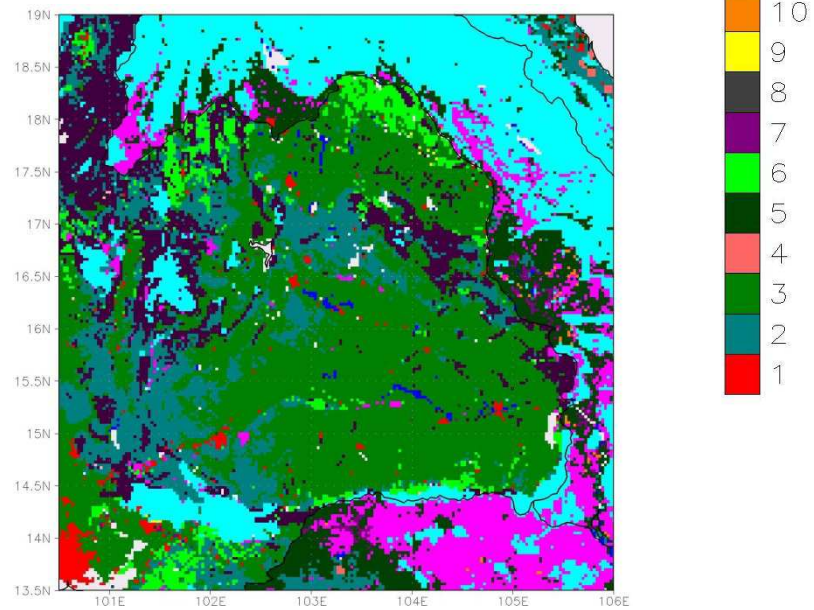
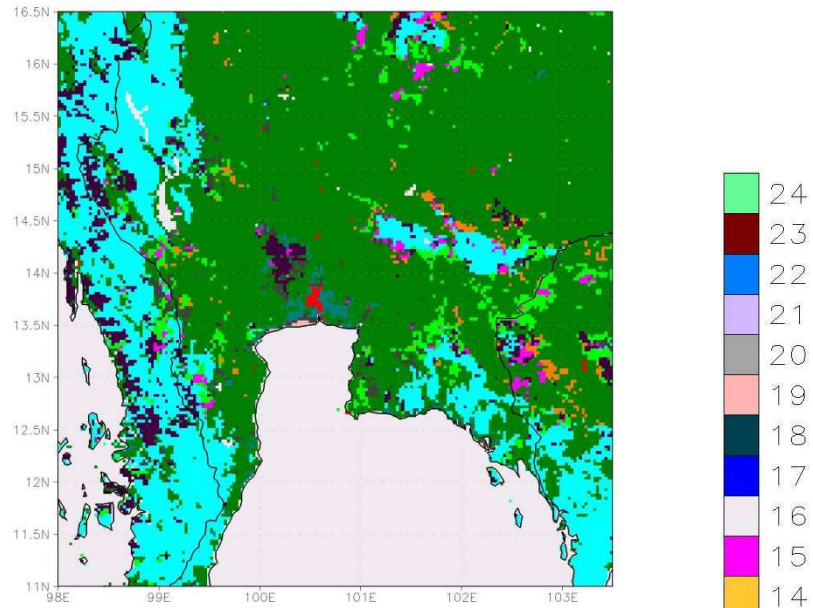


Figure 3.7 Dominant land use class in Northeastern region.

a) Default



b) Updated

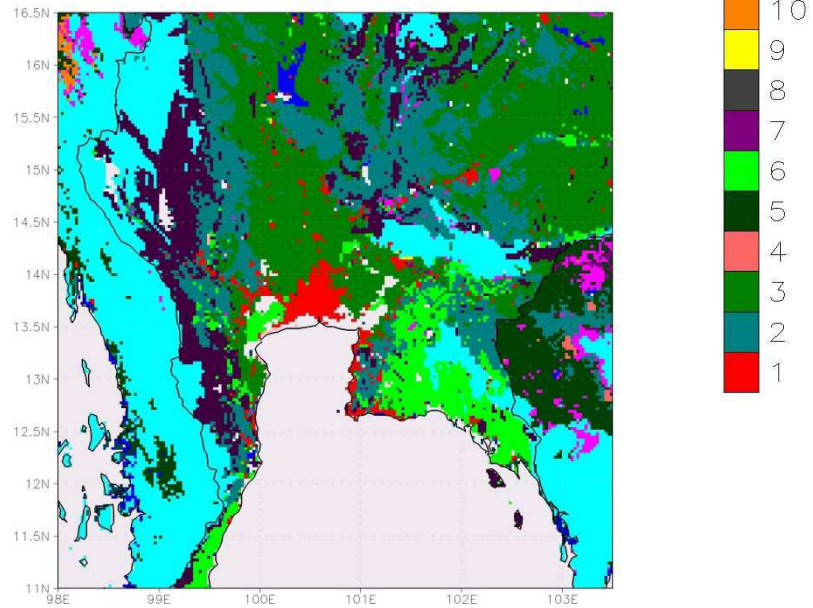
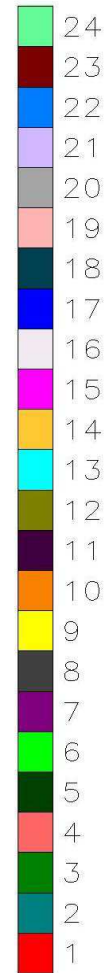
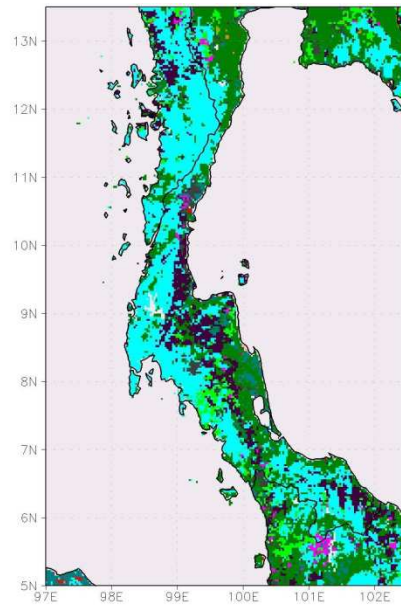


Figure 3.8 Dominant land use class in Central & Eastern regions.

a) Default



b) Updated

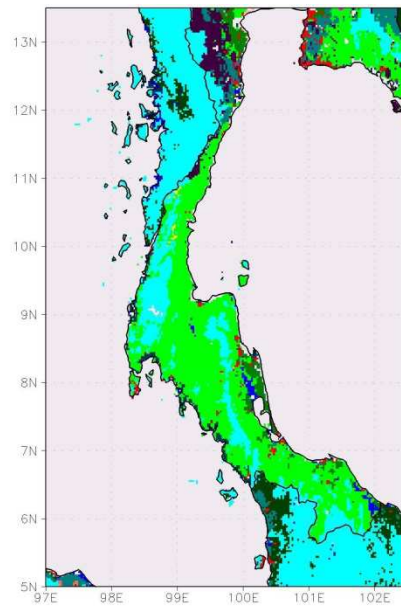


Figure 3.9 Dominant land use class in Southern region.

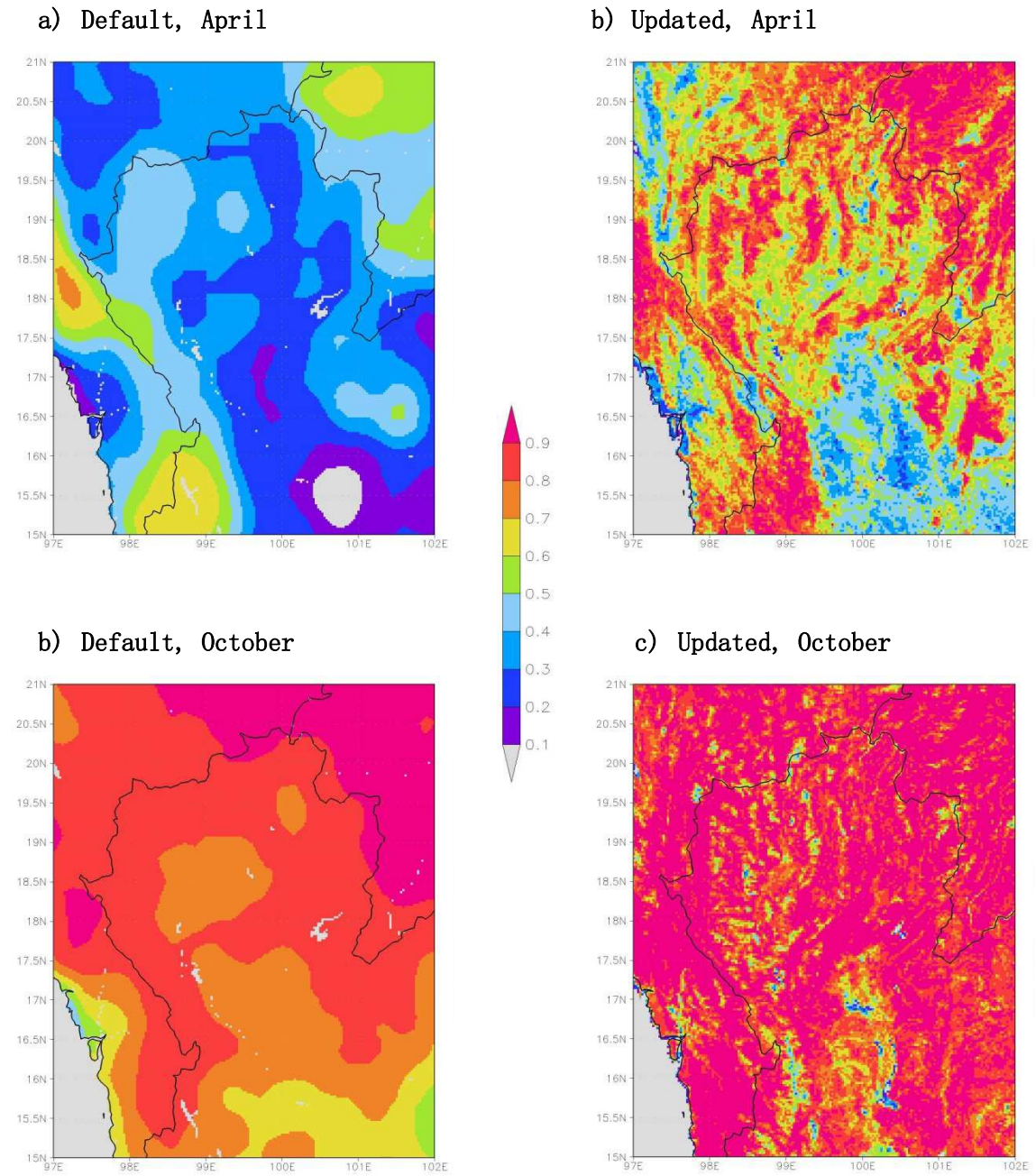


Figure 3.10 Green fraction in Northern region.

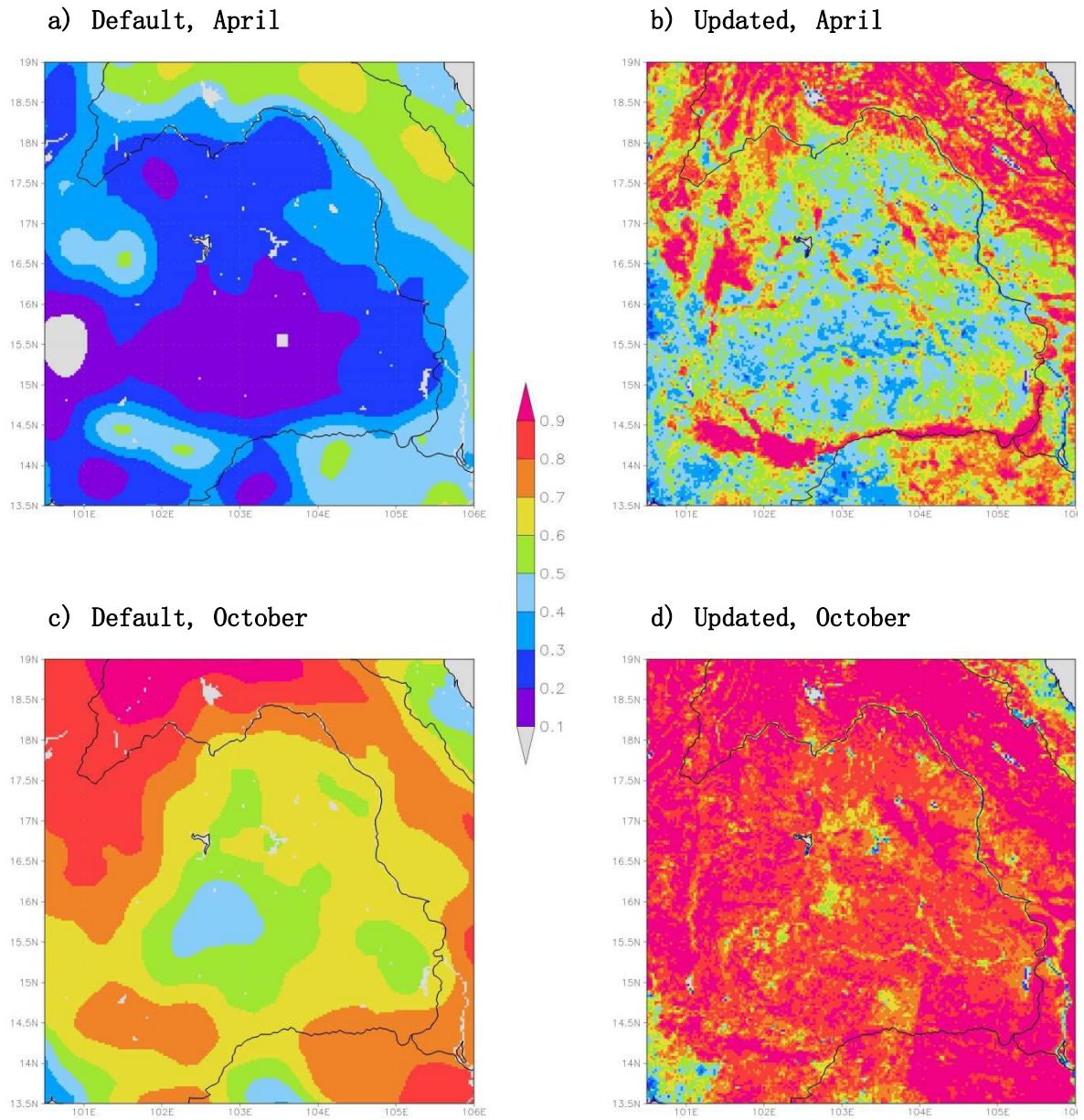


Figure 3.11 Green fraction in Northeastern region.

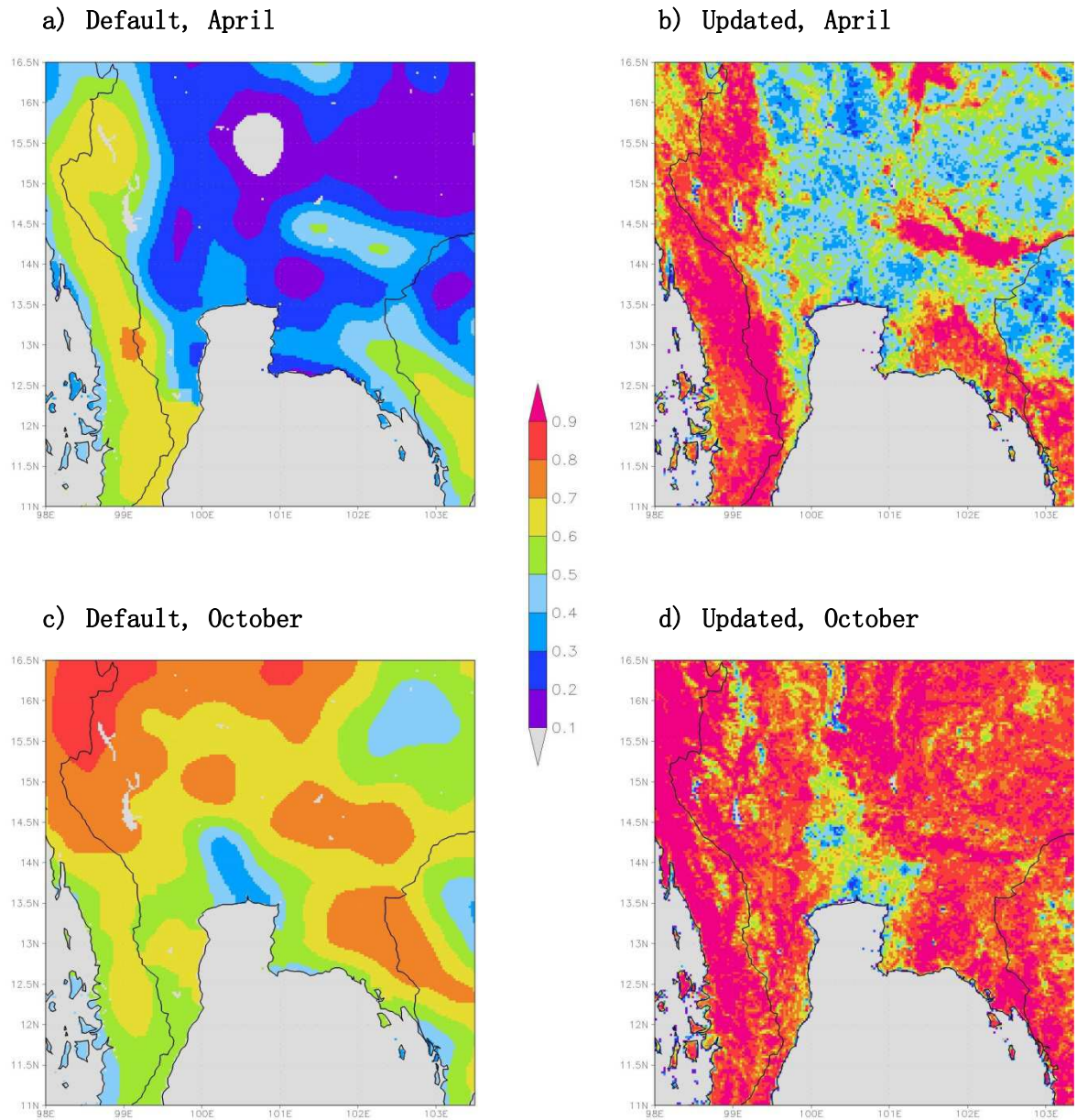
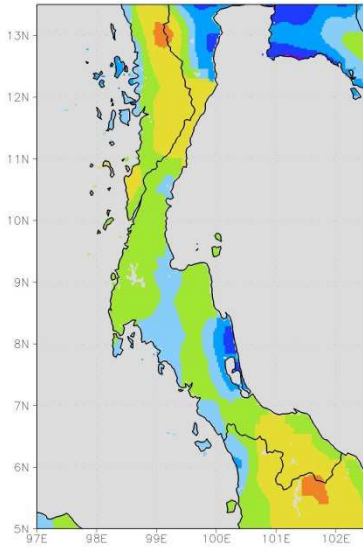
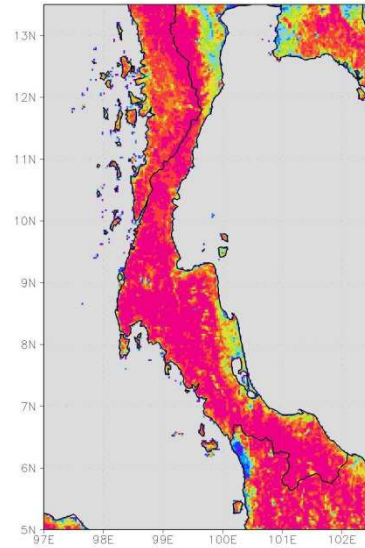


Figure 3.12 Green fraction in Central & Eastern regions.

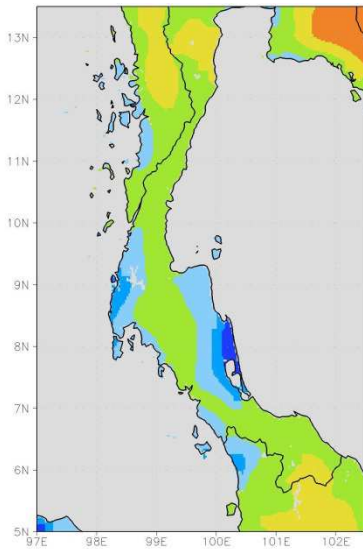
a) Default, April



b) Updated, April



c) Default, October



e) Updated, October

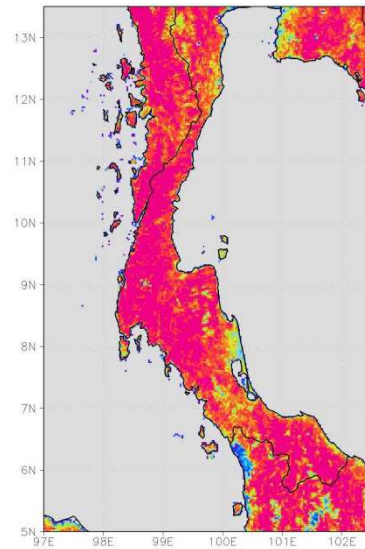


Figure 3.13 Green fraction in Southern region.

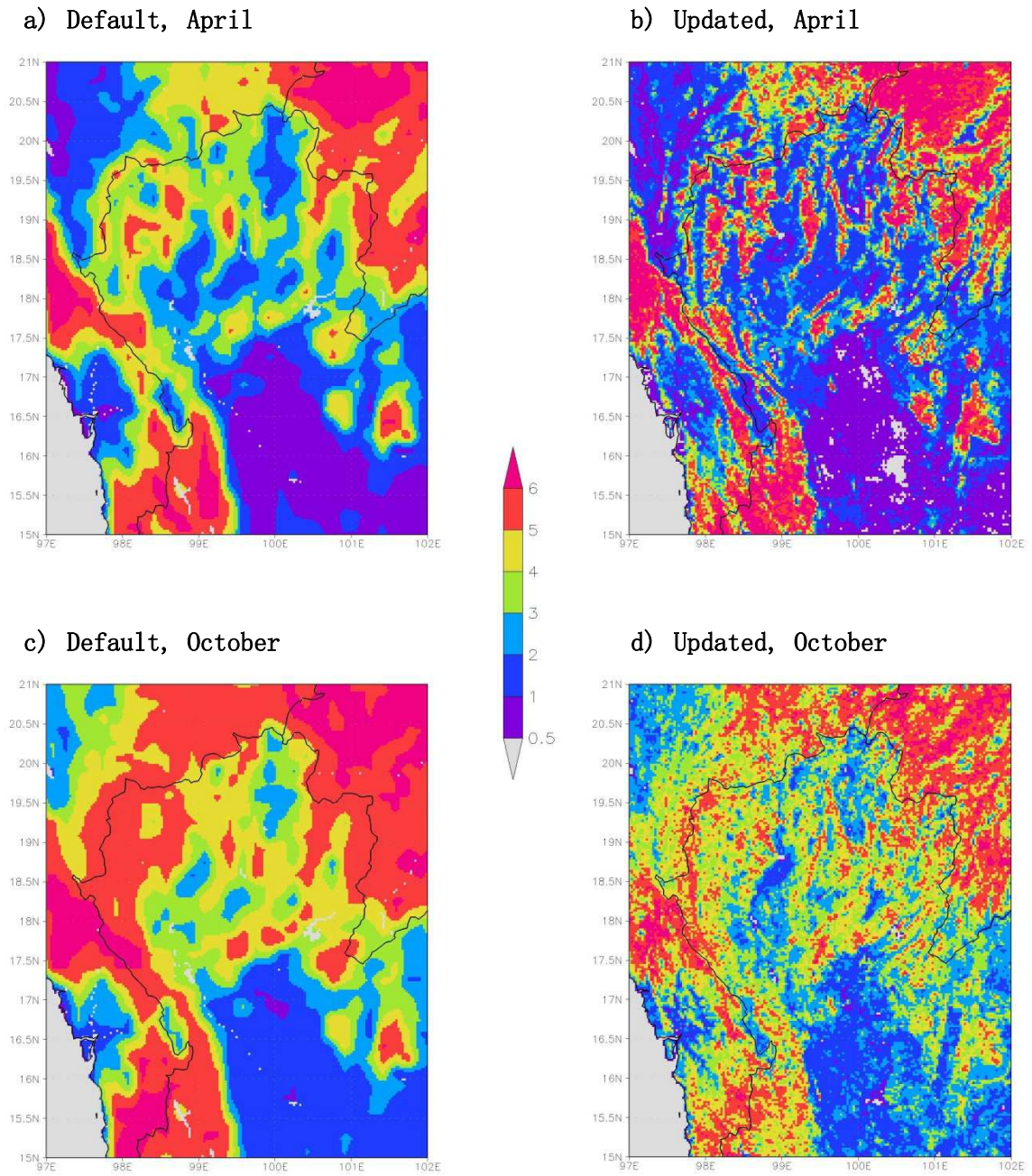


Figure 3.14 Leaf area index ($\text{m}^2 \text{m}^{-2}$) in Northern region.

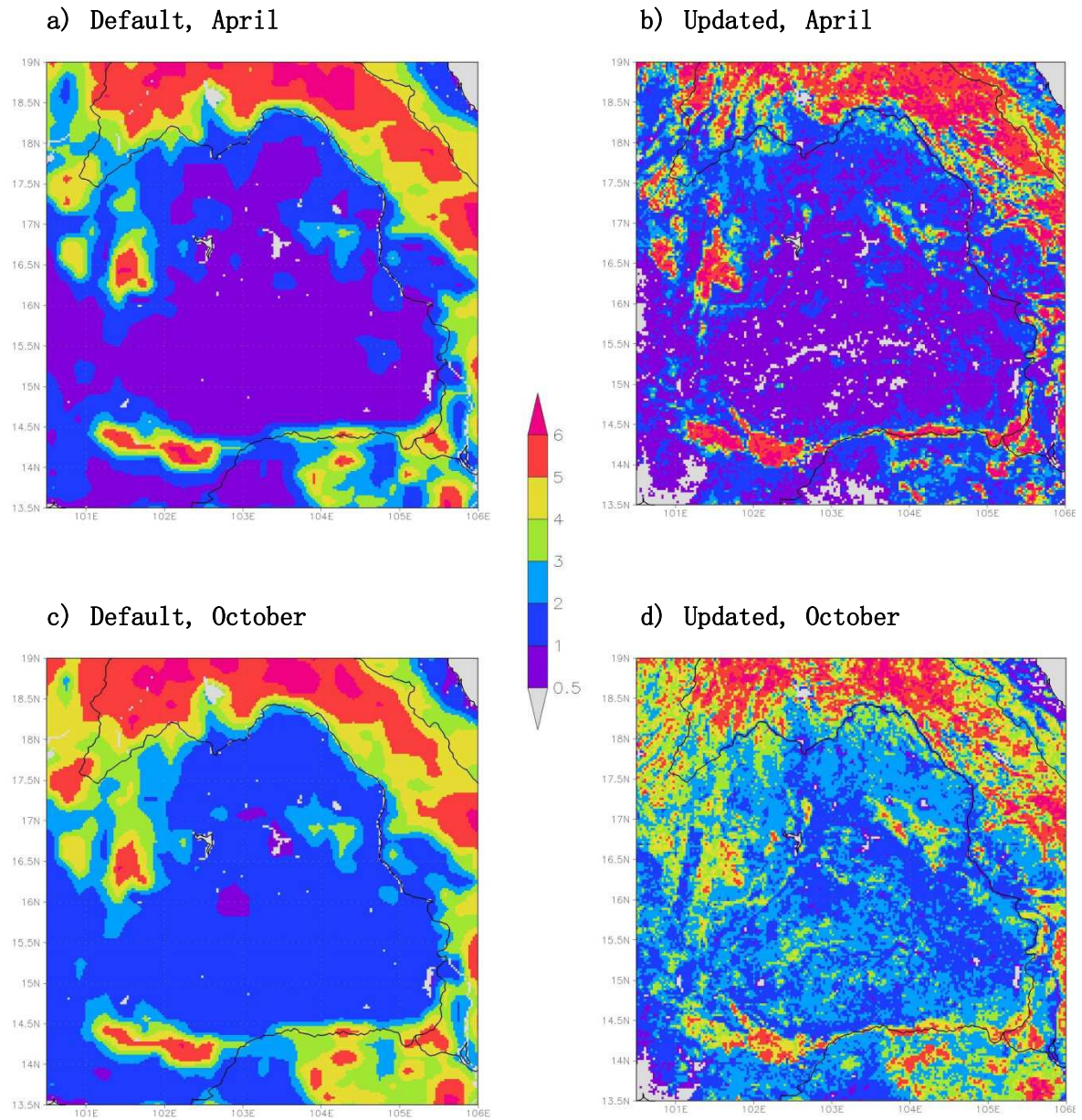


Figure 3.15 Leaf area index ($\text{m}^2 \text{m}^{-2}$) in Northeastern region.

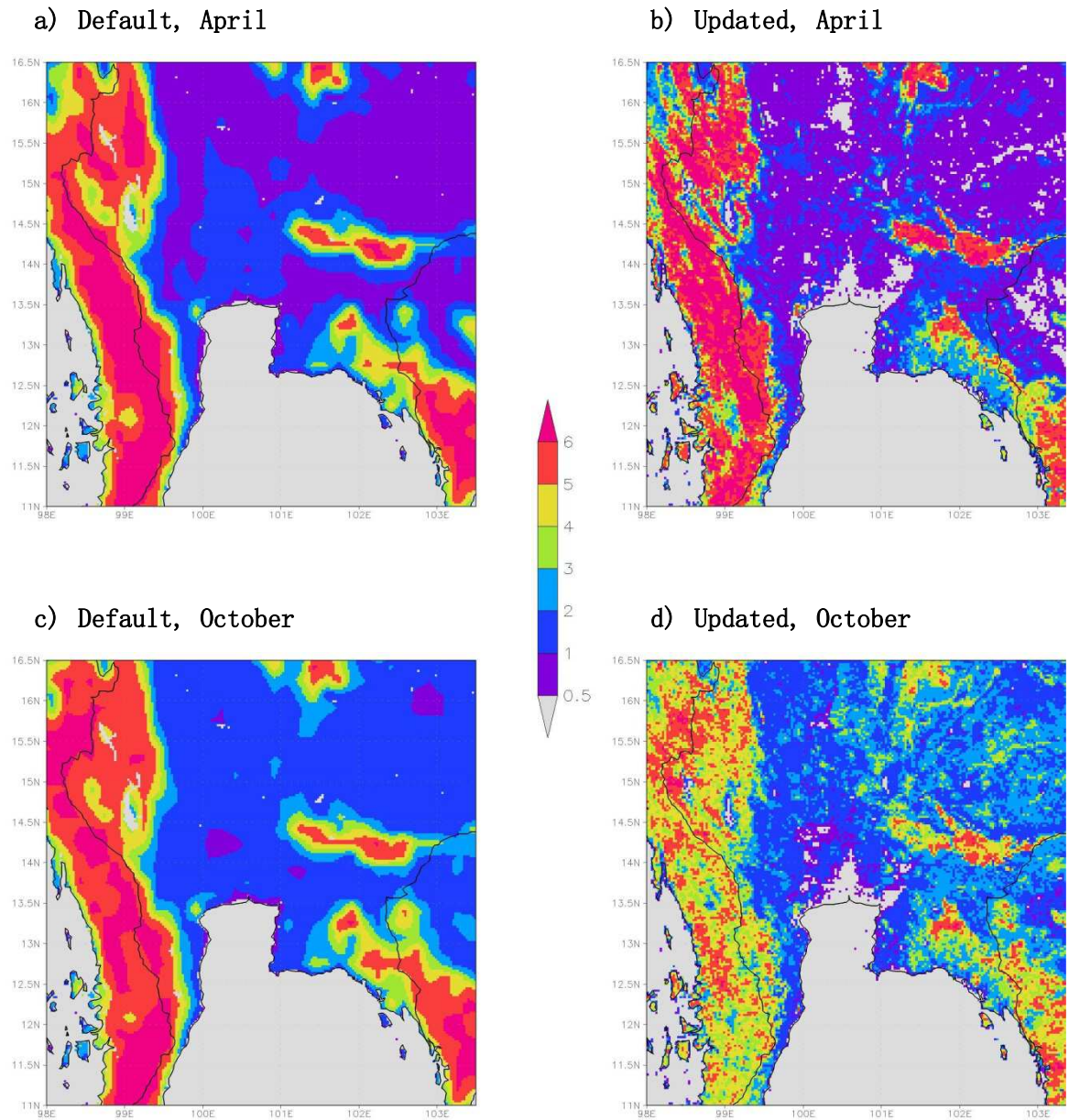
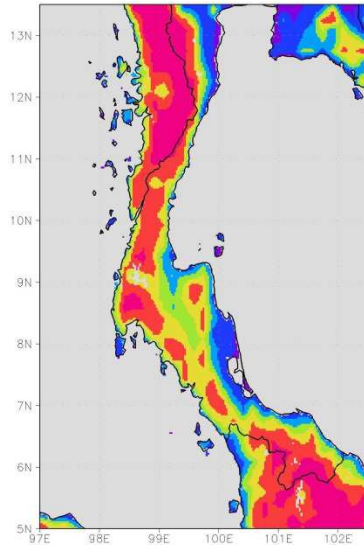
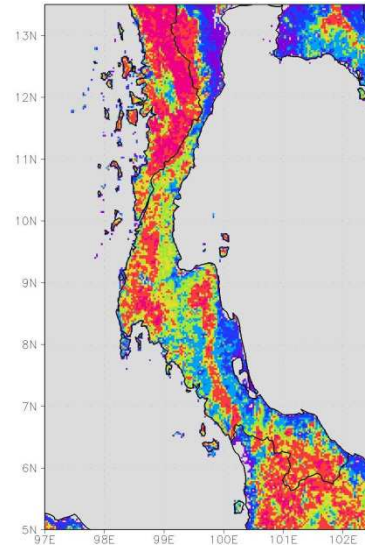


Figure 3.16 Leaf area index ($\text{m}^2 \text{m}^{-2}$) in Central & Eastern regions.

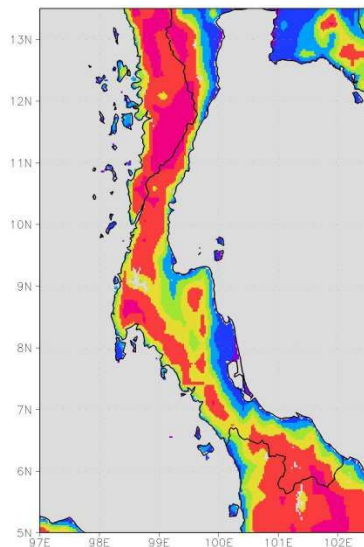
a) Default, April



b) Updated, April



c) Default, October



d) Updated, October

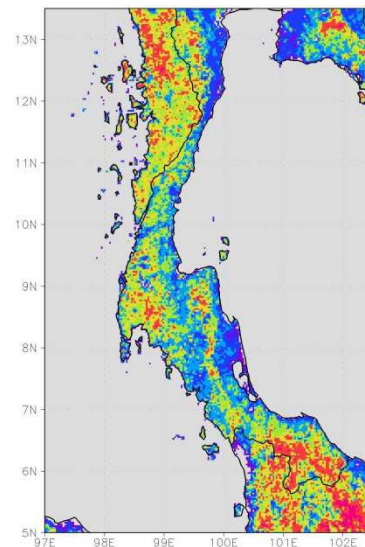


Figure 3.17 Leaf area index ($\text{m}^2 \text{m}^{-2}$) in Southern region.

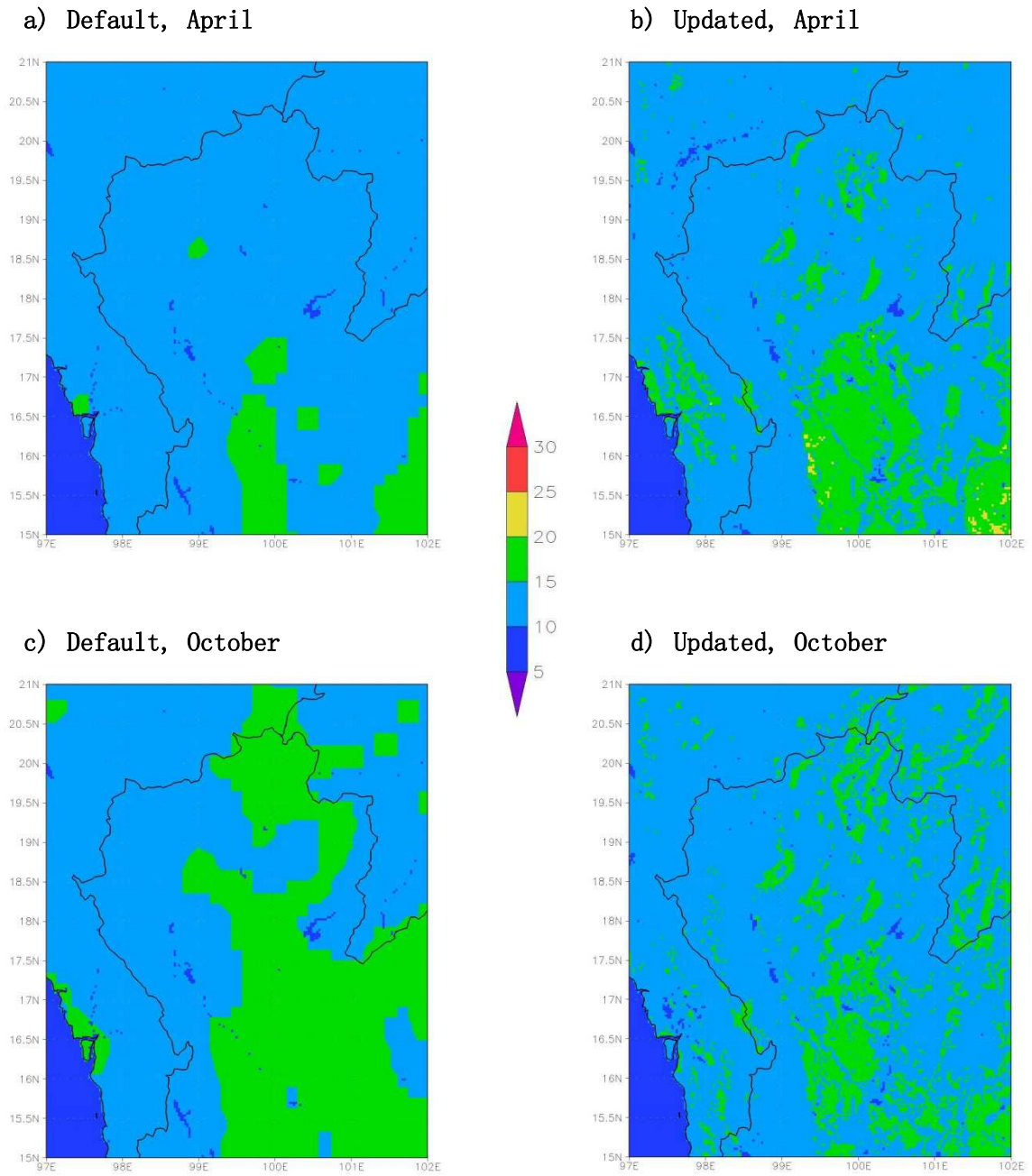


Figure 3.18 Surface albedo (%) in Northern region.

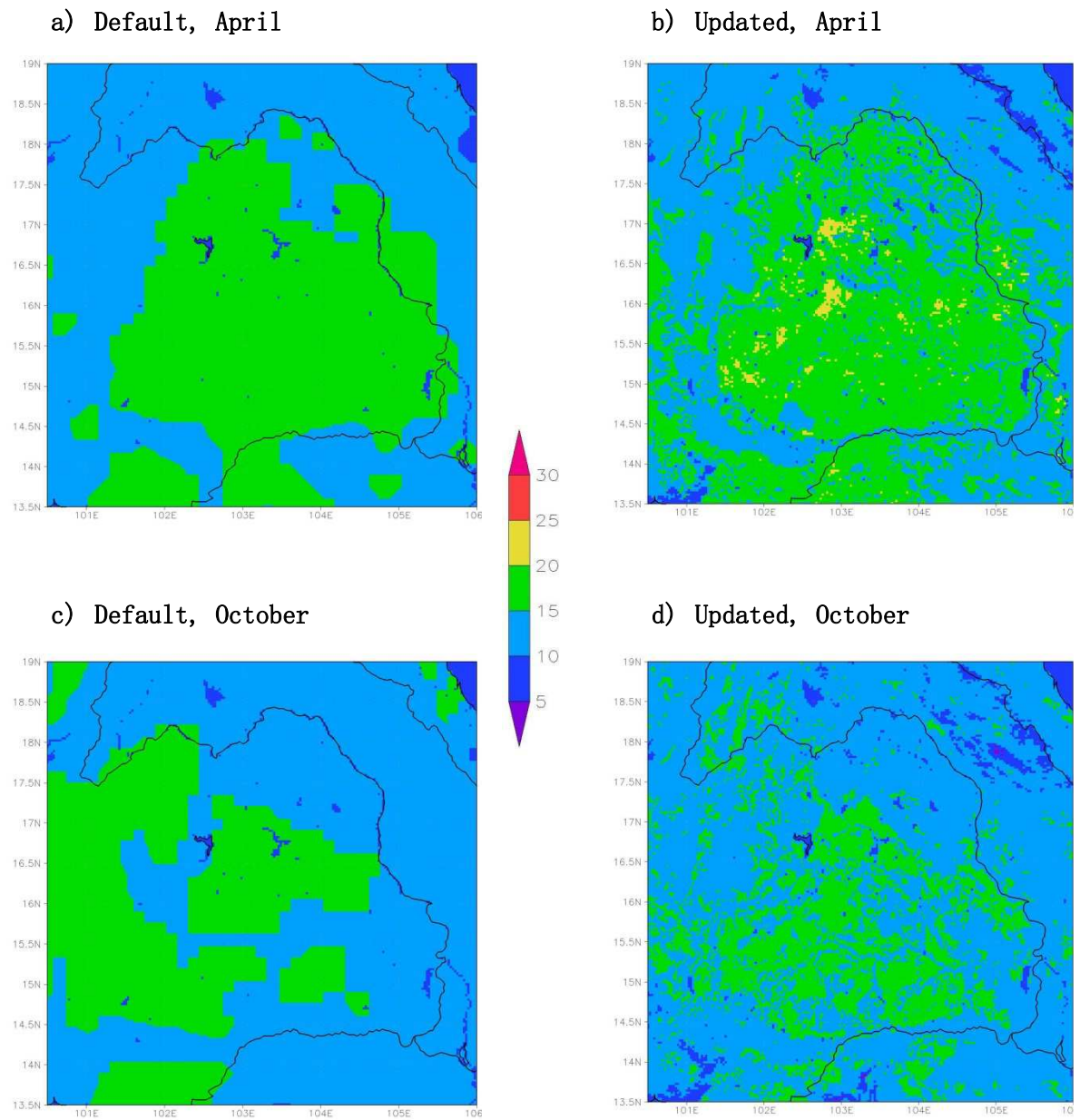


Figure 3.19 Surface albedo (%) in Northeastern region.

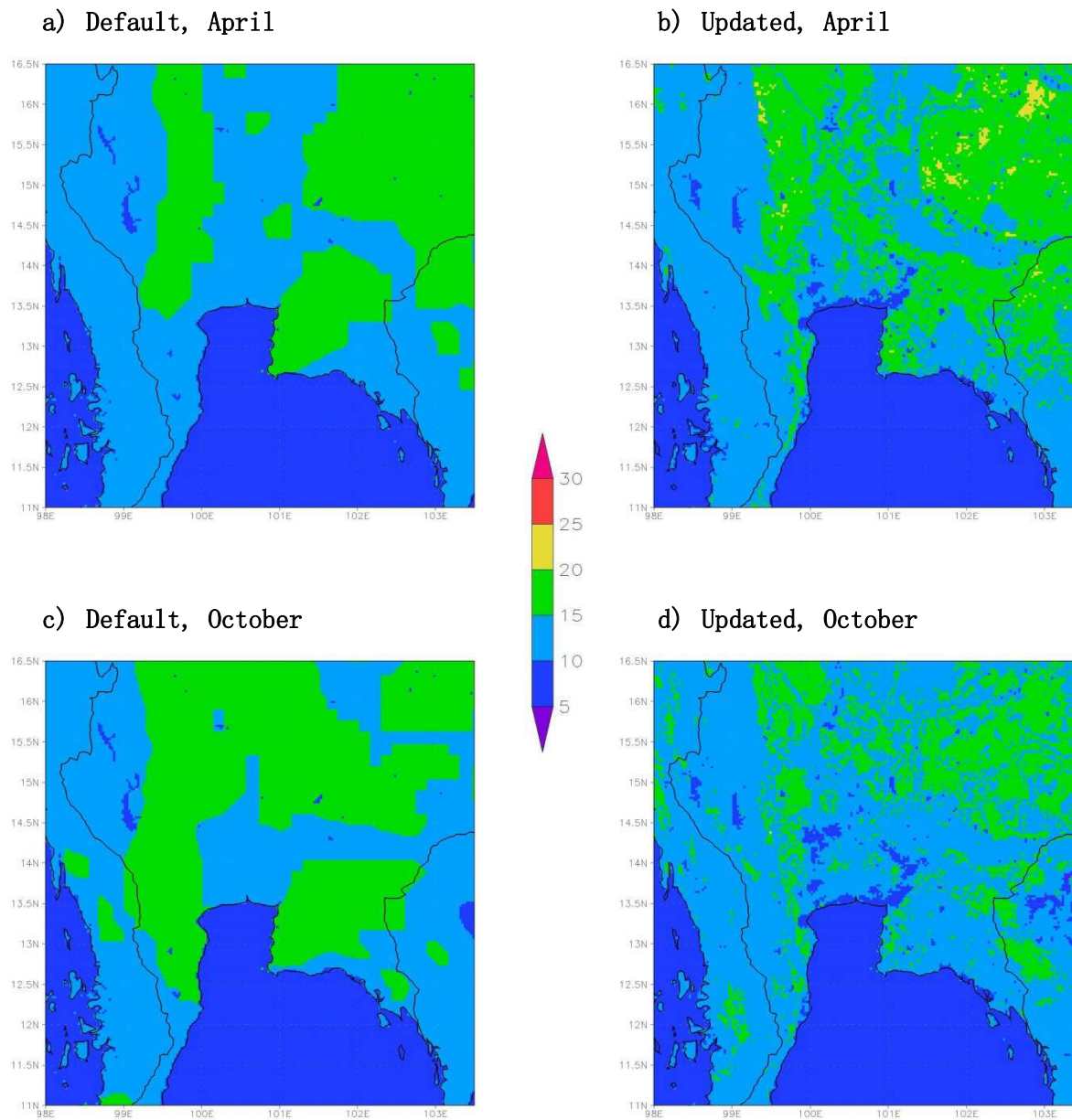
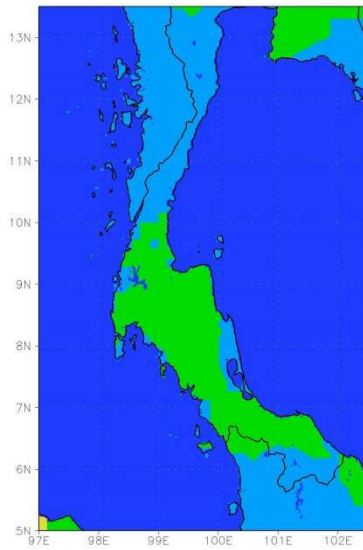
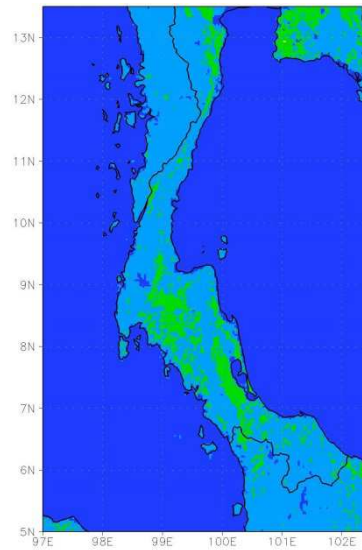


Figure 3.20 Surface albedo (%) in Central & Eastern regions.

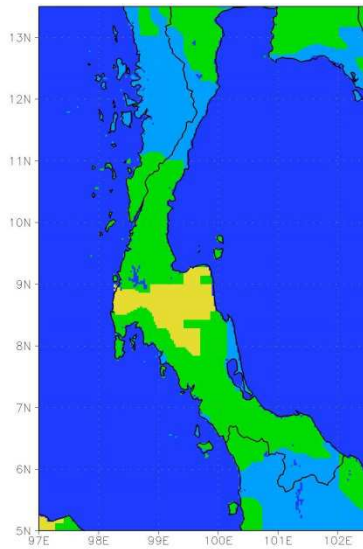
a) Default, April



b) Updated, April



c) Default, October



d) Updated, October

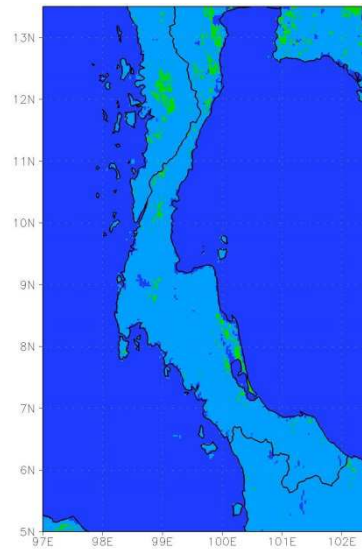
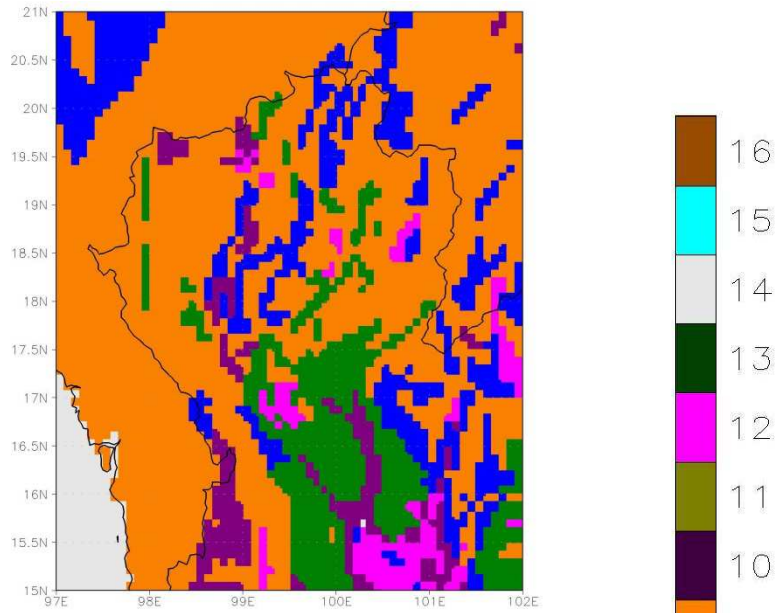


Figure 3.21 Surface albedo (%) in Southern region.

a) Default



b) Updated

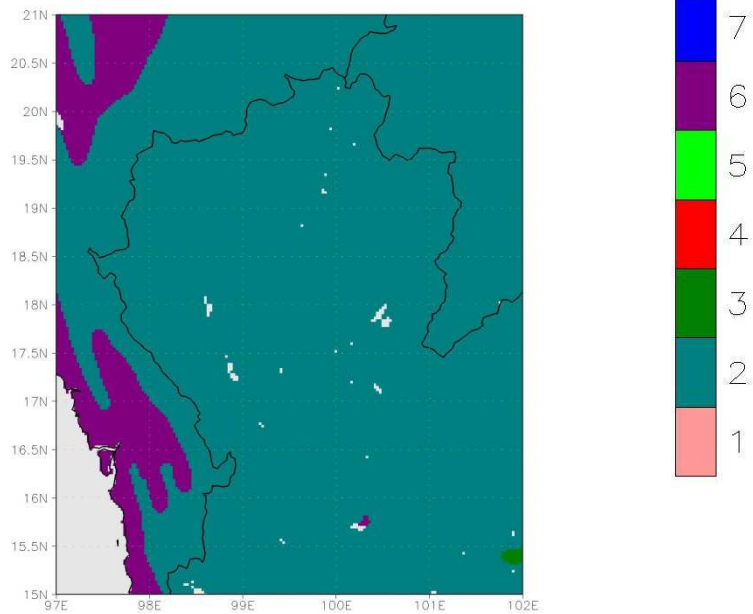
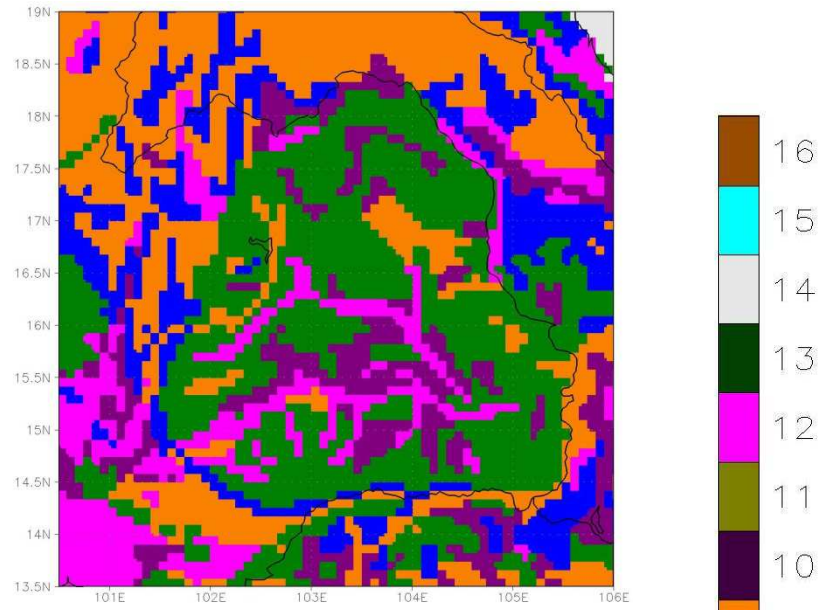


Figure 3.22 Dominant top-layer soil texture in Northern region.

a) Default



b) Updated

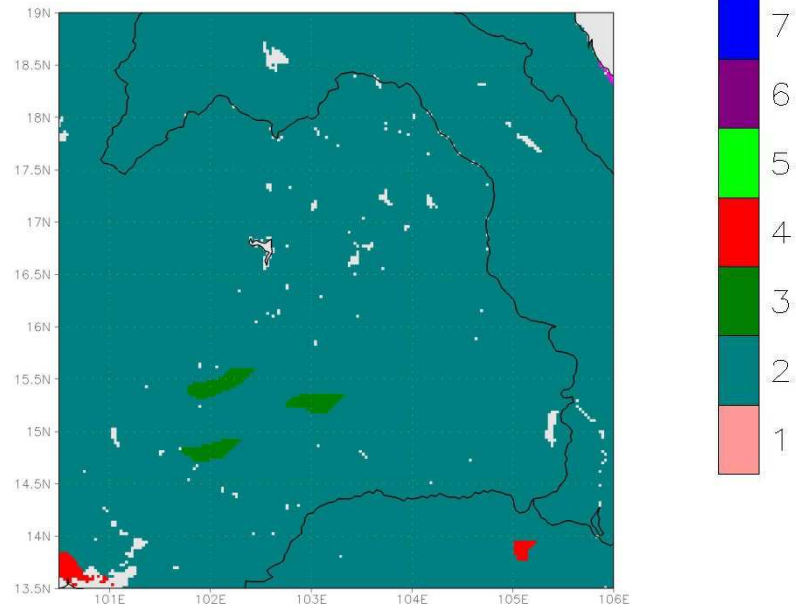
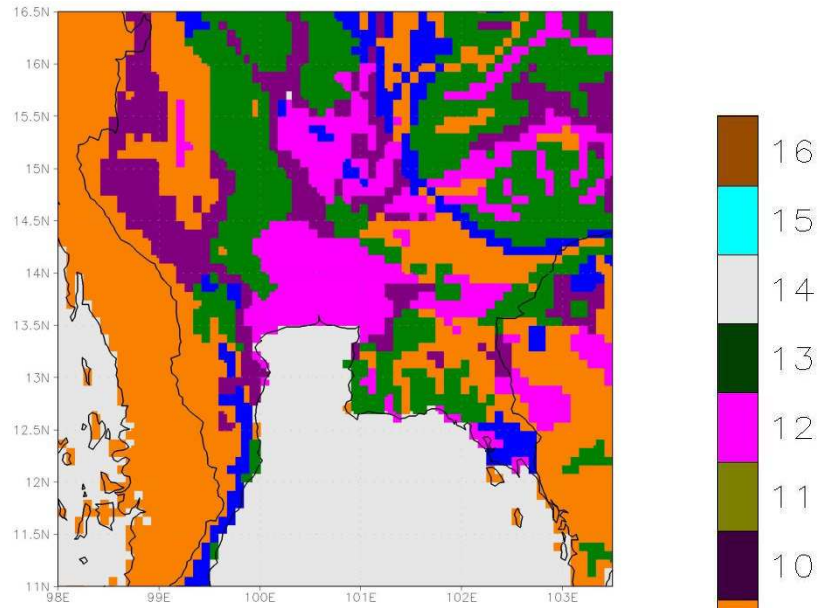


Figure 3.23 Dominant top-layer soil texture in Northeastern region.

a) Default



b) Updated

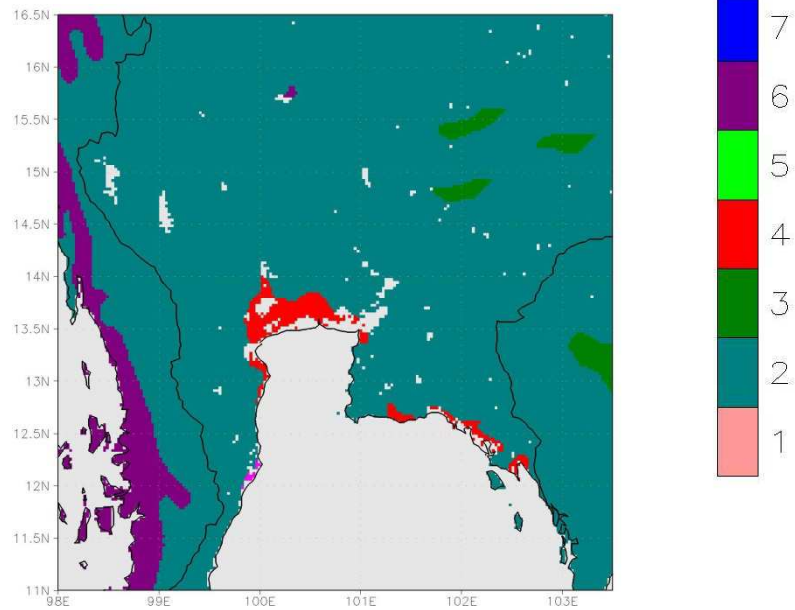
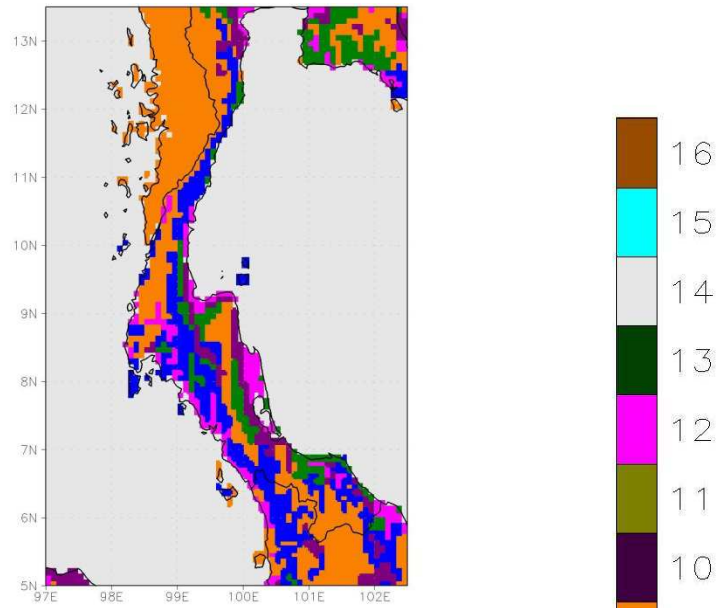


Figure 3.24 Dominant top-layer soil texture in Central & Eastern regions.

a) Default



b) Updated

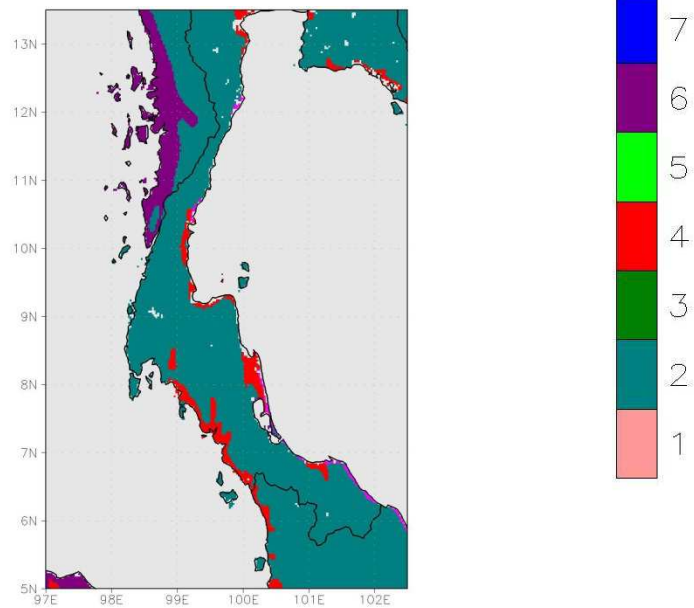
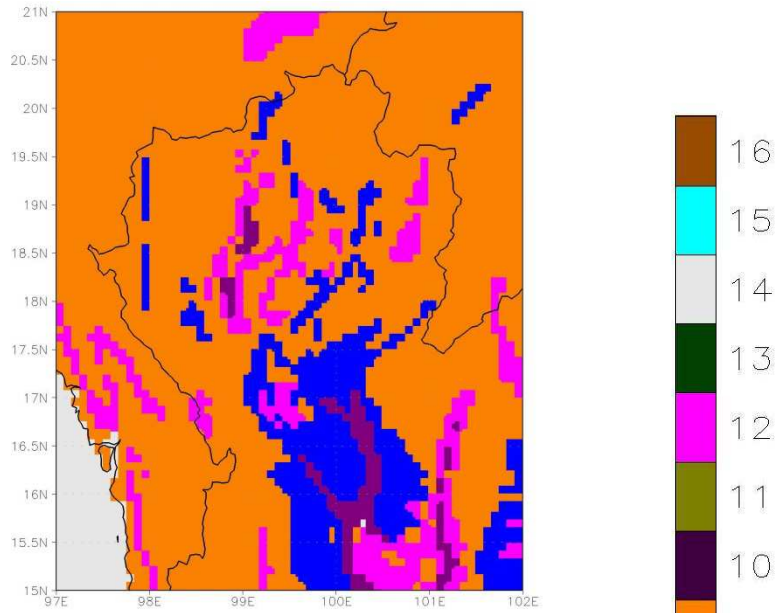


Figure 3.25 Dominant top-layer soil texture in Southern region.

a) Default



b) Updated

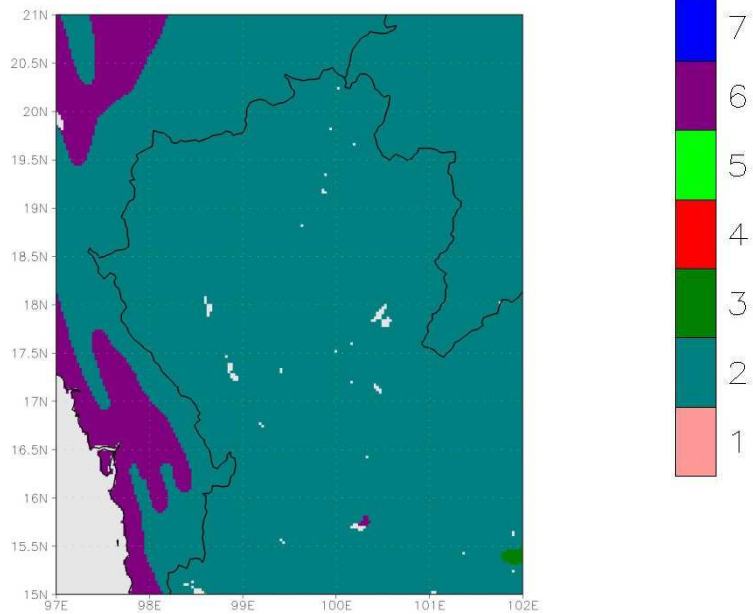
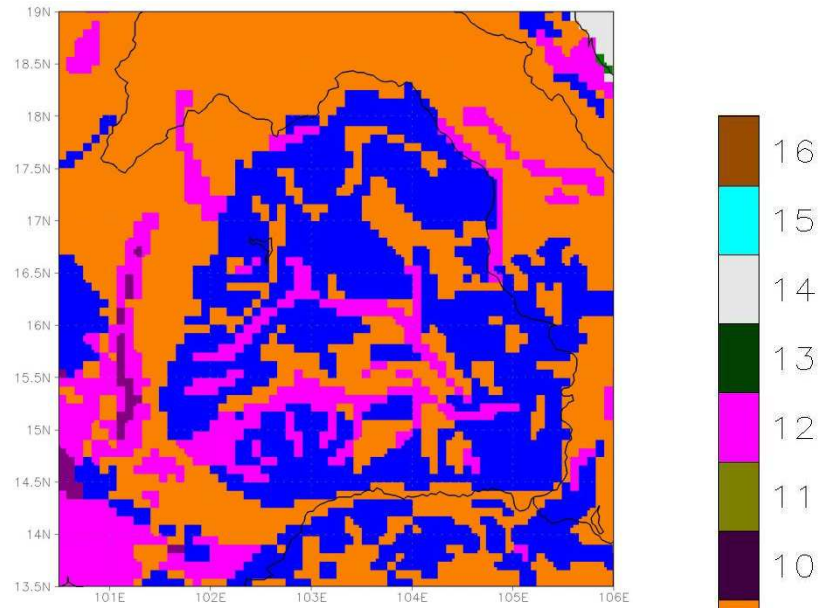


Figure 3.26 Dominant bottom-layer soil texture in Northern region.

a) Default



b) Updated

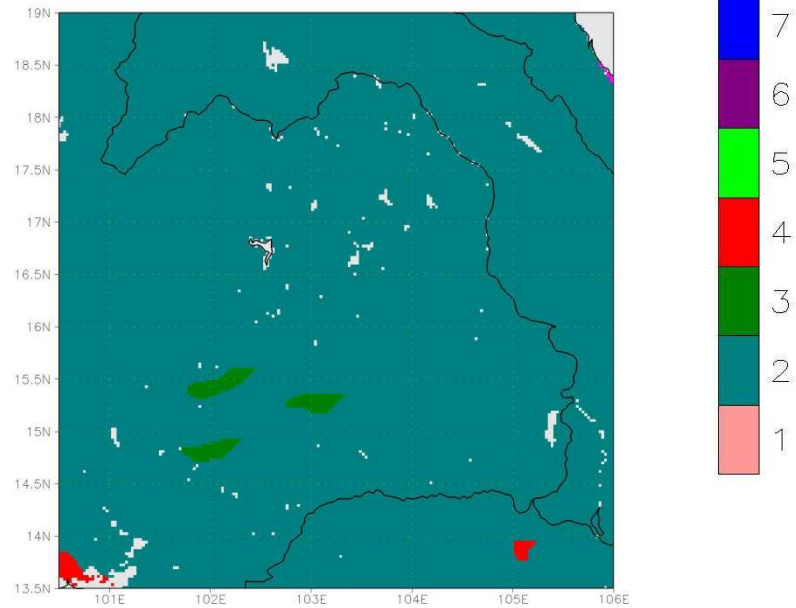
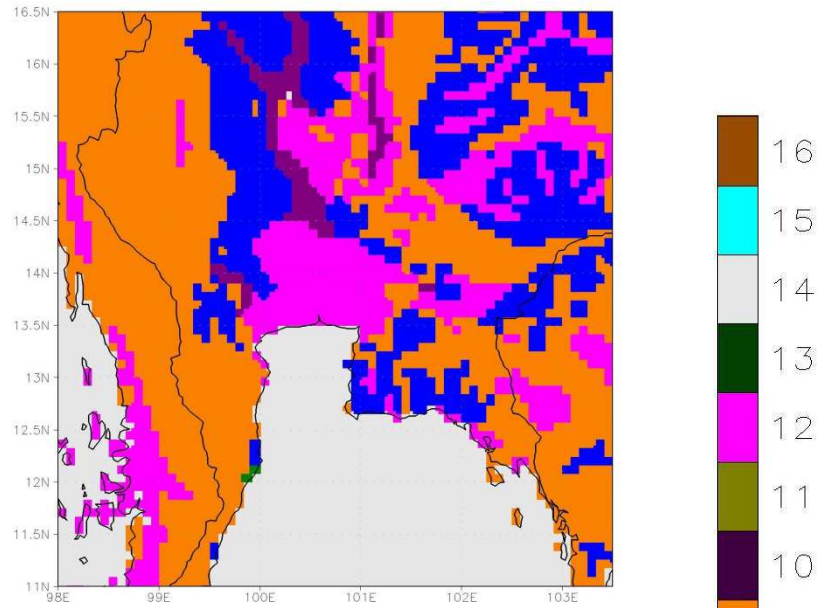


Figure 3.27 Dominant bottom-layer soil texture in Northeastern region.

a) Default



b) Updated

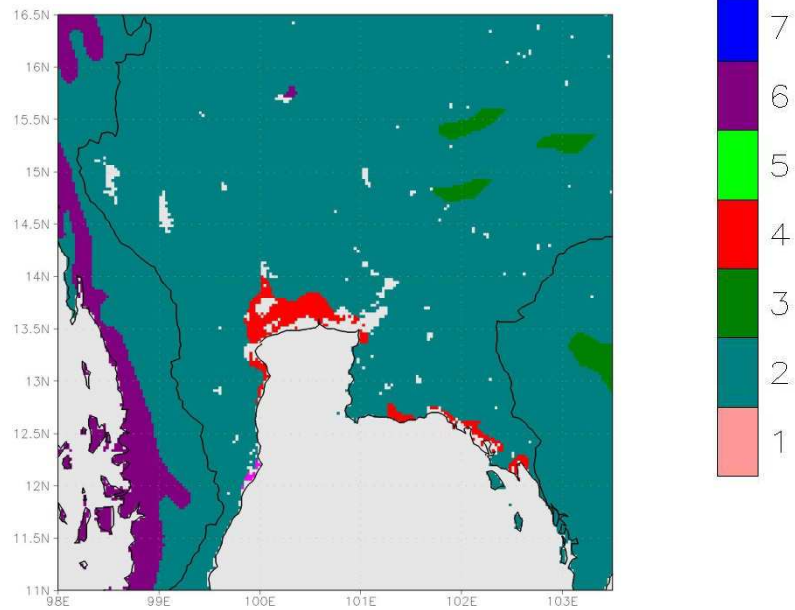
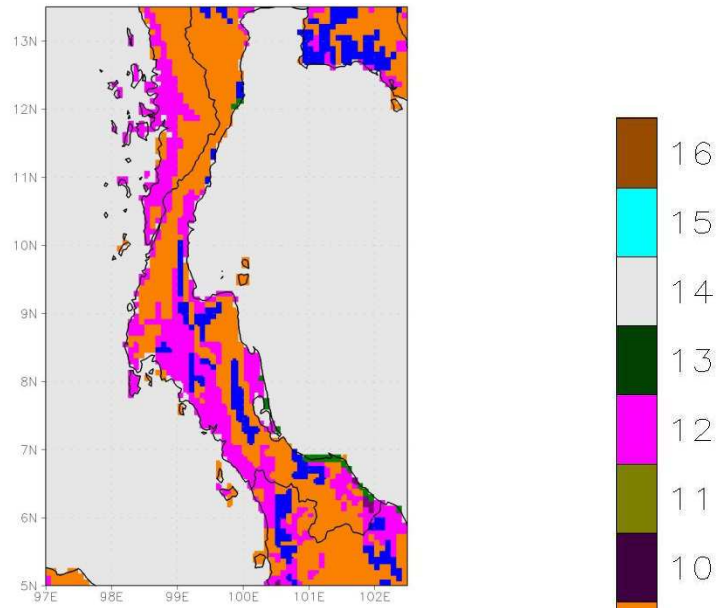


Figure 3.28 Dominant bottom-layer soil texture in Central & Eastern regions.

a) Default



b) Updated

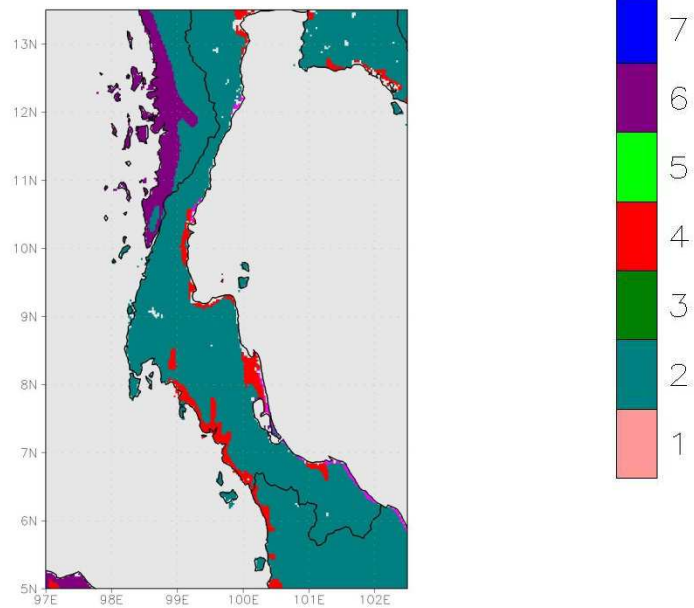


Figure 3.29 Dominant bottom-layer soil texture in Southern region.

Chapter 4

Conclusions and Recommendations

4.1 Conclusions

This study has completed the development of a suite of computer programs, called GUPROC (GEOGRID Updating Processor), which is capable of updating input surface data for WRF modeling to reduce uncertainty in the modeling due to unrealistic or unrepresentative surface data. GUPROC incorporates new surface data compiled from various sources into GEOGRID (a main WRF preprocessor) using an offline approach, by which GEOGRID default output variables are replaced with new surface data after running GEOGRID first. This approach does not interfere with any internal source codes, scripts, and input control files in GEOGRID. The GUPROC development relied mainly on open-source and/or freeware software and tools, with all codes and scripts written in standard computer languages. GUPROC can be installed and run on a simple standard PC or server (see Figure 4.1 as an example). A full electronic copy of GUPROC has already been transferred to the HAIL in support of its research and operational activities.

Currently, GUPROC is able to update the following variables:

1. Terrain height
2. USGS land use (dominant and individual fractions)
3. Monthly green fraction
4. Monthly leaf area index
5. Monthly surface albedo
6. Top-layer soil texture (dominant and individual fractions)
7. Bottom-layer soil texture (dominant and individual fractions) and
8. Land mask.

Moreover, it can handle both Mercator or Lambert conformal conic map projections and support modeling studies over areas in Thailand and the Lower Mekong River region at resolutions of 1 km or coarser. The updating process by GUPROC considerably involves with data re-gridding, as cell aggregation from finer to coarser resolutions or as resampling from coarser to finer resolutions. GUPROC utilizes a simple bin-filling method to support such re-gridding tasks in order to achieve acceptable quality. GUPROC was demonstrated using the WRF modeling domains used in the HAI forecasting operation. It was found that the updated results given by GUPROC show significant differences from the GEOGRID default results for all variables considered. For terrain height, both default and updated results are generally comparable, especially the recently developed datasets, but their discrepancy tends to be amplified over mountainous areas. The updating satisfactorily yields more realistic results with improved spatial details for the land use, green fraction, leaf area index, and albedo variables. For soil texture (both top-layer and bottom-layer), the default results are shown to be superior because the new soil texture data selected for the updating are unfortunately somewhat outdated for Thailand and its neighboring countries. Hence, the default soil texture data should still be maintained for use.

4.2 Recommendations

Below are some practical recommendations as well as perspectives for GUPROC application and future enhancement:

1. WRF users should make an effort to configure or set up their modeling domains with input surface data of good quality. The developed GUPROC tool can be used to serve this purpose.
2. Sensitivity tests on WRF modeling with/without updating with the new surface data by GUPROC should be conducted to inspect how modeled results change due to the updating. In principle, although improved prediction performance is not warranted to achieve, the updating can assist selecting physics options of interest (especially, land surface

model) in WRF more robustly since it naturally reduces the potential influences from unrealistic or unrepresentative surface data on modeled results.

3. Currently, only a set of GEOGRID output variables are handled by GUPROC. Future work can be extended to include more variables. For example, WRF modeling with an urban canopy model has become more important to account for urbanization over cities or metropolitans, requiring variables associated with urban configuration. and
4. New surface data of higher or better quality, when available, can also be considered for incorporation into GUPROC.



Figure 4.1 Hardware platform used in GUPROC development.

References

ASTER GDEM Validation Team (2011) ASTER Global Digital Elevation Model version 2 – summary of validation results. Ministry of Economy, Trade and Industry (METI), Japan, and National Aeronautics and Space Administration (NASA), US.

Danielson JJ, Gesch DB (2011) Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010). Report No. 2011-1073, US Department of the Interiors and US Geological Survey.

Didan K (2015) MOD13A1 MODIS/Terra vegetation indices 16-day L3 global 500m SIN Grid V006. NASA EOSDIS Land Processes DAAC.

DiMiceli C, Carroll M, Sohlberg R, Kim DH, Kelly M, Townshend JRG (2017) MOD44B MODIS/Terra vegetation continuous fields yearly L3 global 250m SIN Grid V006. NASA EOSDIS Land Processes DAAC.

FAO (1977) The FAO–UNESCO soil map of the world, volume VII, South Asia. Food and Agriculture Organization of the United Nations.

FAO (1979) The FAO–UNESCO soil map of the world, volume IX, Southeast Asia. Food and Agriculture Organization of the United Nations.

FAO/IIASA/ISRIC/ISSCAS/JRC (2012) Harmonized World Soil Database (version 1.2). Food and Agriculture Organization and International Institute for Applied Systems Analysis (IIASA).

Foy BD, Molina LT, Molina MJ (2006) Satellite-derived land surface parameters for mesoscale modelling of the Mexico City basin. *Atmospheric Chemistry and Physics*, 6, 1315–1330.

Friedl MA, Sulla-Menashe D, Tan B, Schneider A, Ramankutty N, Sibley A, Huang X (2010) MODIS collection 5 global land cover: algorithm refinements and characterization of new datasets. *Remote Sensing of Environment*, 114, 168–182.

Gelaro R, McCarty W, Suarez MJ, Todling R, Molod A, Takacs L, Randles C, Darmenov A, Bosilovich MG, Reichle R, Wargan K, Coy L, Cullather R, Draper C, Akella S, Buchard V, Conaty A, da Silva A, Gu W, Kim G-K, Koster R, Lucchesi R, Merkova D, Nielsen JE, Partyka G, Pawson S, Putman W, Rienecker M, Schubert SD, Sienkiewicz M, Zhao B (2017) The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2). *Journal of Climate*. (in press)

Gutman G, Ignatov A (1997) Satellite-derived green vegetation fraction for the use in numerical weather prediction models, *Advances in Space Research*, 19, 477–480.

- Gutman G, Ignatov A (1998) The derivation of the green vegetation fraction from NOAA/AVHRR data for use in numerical weather prediction models. *International Journal of Remote Sensing*, 19, 1533–1543.
- Li M, Song Y, Huang X, Li J, Mao Y, Zhu T, Cai X, Liu B (2014) Improving Mesoscale modeling using satellite-derived land surface parameters in the Pearl River Delta region, China. *Journal of Geophysical Research: Atmospheres*, 119, 6325–6346.
- Lucht W, Schaaf CB, Strahler AH (2000) An algorithm for the retrieval of albedo from space using semiempirical BRDF models. *IEEE Transactions on Geoscience and Remote Sensing*, 38, 977–998.
- Manomaiphiboon K, Assareh N, Prabamroong T, Aman N (2016) Development of a short-term wind speed forecasting system: a case study of Lamtakong reservoir area. Final Report, the Joint Graduate School of Energy and Environment, King Mongkut's University of Technology Thonburi, supported by the Electricity Generating Authority of Thailand. (in Thai with English Abstract)
- Myneni RYK, Park T (2015) MCD15A2H MODIS/Terra+Aqua leaf area index/FPAR 8-day L4 global 500m SIN Grid V006. NASA EOSDIS Land Processes DAAC.
- Octaviani M, Manomaiphiboon K (2011) Performance of regional climate model RegCM3 over Thailand. *Climate Research*, 47, 171–186.
- Paton CP, Manomaiphiboon K (2013) A metropolitan wind resource assessment for Bangkok, Thailand part 1: wind resource mapping. *Journal of Sustainable Energy & Environment*, 4, 69–76.
- Ran L, Pleim J, Gilliam R, Binkowski FS, Hogrefe C, Band L (2016) Improved meteorology from an updated WRF/CMAQ modeling system with MODIS vegetation and albedo. *Journal of Geophysical Research*, 121, 2393–2415.
- Schaaf CB (2015) MCD43A3 MODIS/Terra+Aqua BRDF/albedo daily L3 global 500m V006. NASA EOSDIS Land Processes DAAC.
- Shchepetkin AF, McWilliams JC (2005) The regional oceanic modeling system (ROMS): a split-explicit, free-surface, topography-following-coordinate oceanic model. *Ocean Modelling*, 9, 347–404.
- Siebert S, Henrich V, Frenken K, Burke J (2013) Global Map of Irrigation Areas (GMIA) version 5. Rheinische Friedrich-Wilhelms-University and Food and Agriculture Organization of the United Nations.

Skamarock WC, Klemp JB, Dudhia J, Gill DO, Barker DM, Duda MG, Huang X-Y, Wang W, Powers JG (2008) A Description of the Advanced Research WRF Version 3. NCAR Technical Note: NCAR/TN - 475+STR.

Sugimoto S, Sato T, Sasaki T (2015) Impact of land-use change on winter precipitation in Hokkaido, Japan. Scientific Online Letters on the Atmosphere, 11, 95-99.

TMD (2017) The climate of Thailand. Thai Meteorological Department. (in Thai)
https://www.tmd.go.th/info/climate_of_thailand-2524-2553.pdf

Torsri K, Wannawong W, Sarinnapakorn K, Boonya-aroonnet S, Chiradon, R (2014) An application of air-sea model components in the coupled ocean atmosphere wave sediment transport (COAWST) modelling system over an Indochina peninsular sub-region: impact of high spatiotemporal SST on WRF model in precipitation prediction. 11th Annual Meeting of Asia Oceania Geosciences Society, Sapporo, Hokkaido, Japan.

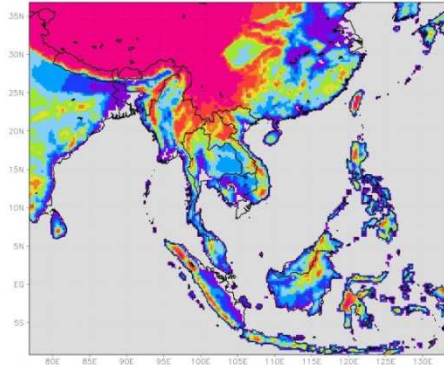
Vahmani P, Hogue TS (2014) High-resolution land surface modeling utilizing remote sensing parameters and the Noah UCM: a case study in the Los Angeles Basin. Hydrology and Earth System Sciences, 18, 4791-4806.

Wang YL, Feng JM, Gao H (2014) Numerical simulation of the impact of land cover change on regional climate in China. Theoretical and Applied Climatology, 115, 141-152.

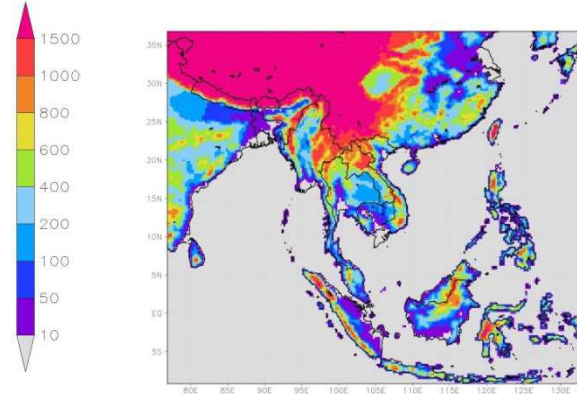
Warner JC, Armstrong B, He R, Zambon JB (2010) Development of a Coupled Ocean-Atmosphere-Wave-Sediment Transport (COAWST) modeling system. Ocean Modeling, 35, 230-244.

Appendix

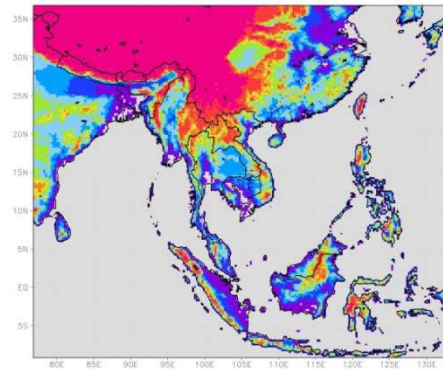
a) Default GTOP030



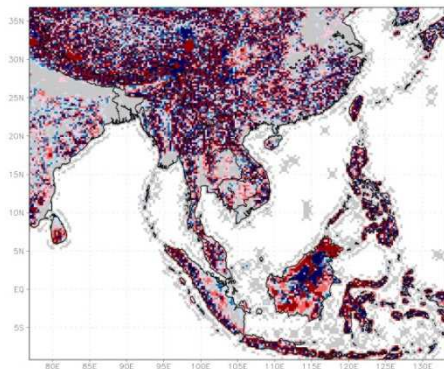
b) Default GMTED2010



c) Updated ASTER-GDEM



d) GTOP030 minus ASTER-GDEM



e) GMTED2010 minus ASTER-GDEM

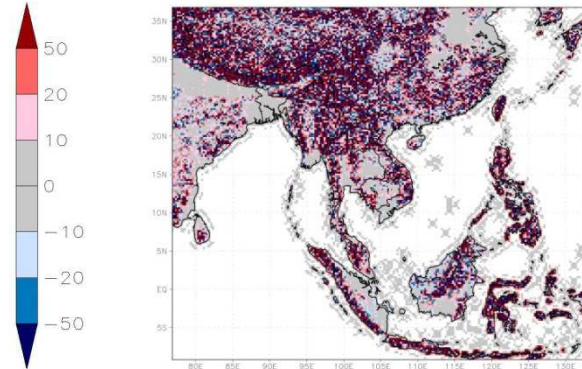
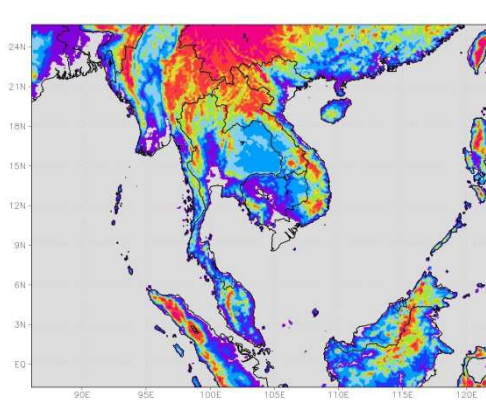
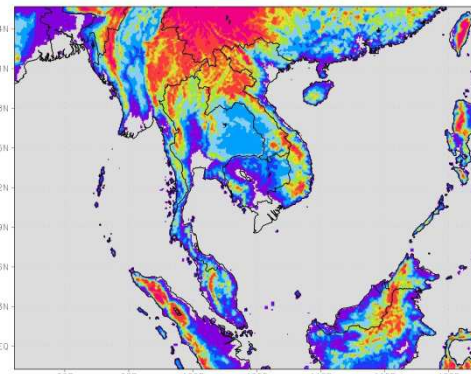


Figure A1 Terrain height in Domain 1.

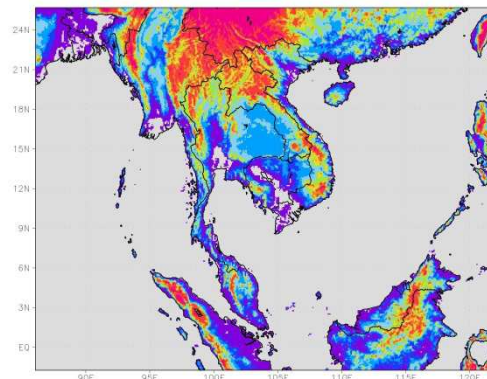
a) Default GTOP030



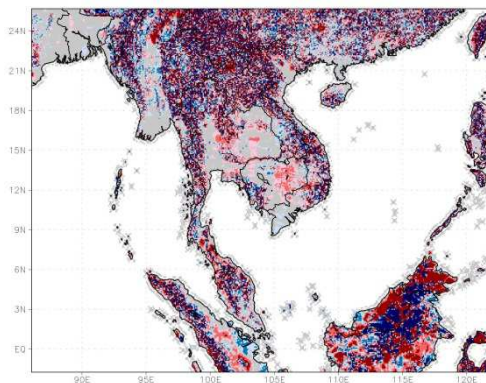
b) Default GMTED2010



c) Updated ASTER-GDEM



d) GTOP030 minus ASTER-GDEM



e) GMTED2010 minus ASTER-GDEM

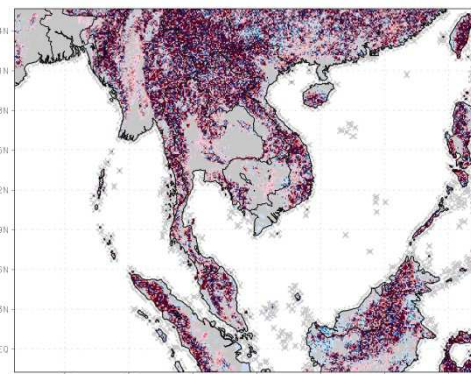
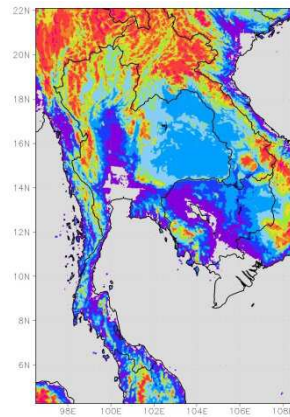
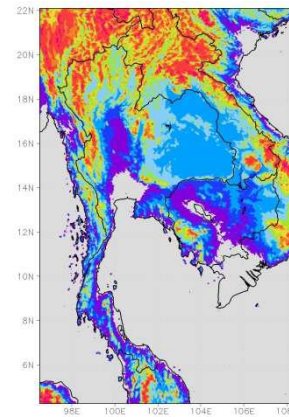


Figure A2 Terrain height in Domain 2.

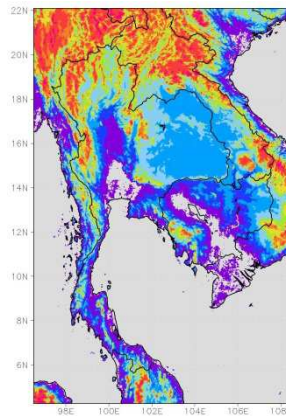
a) Default GTOP030



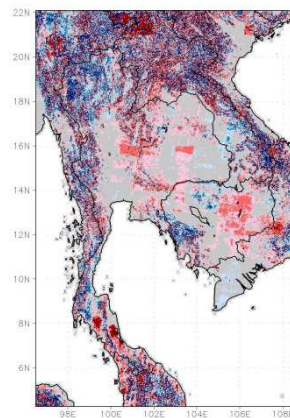
b) Default GMTED2010



c) Updated ASTER-GDEM



d) GTOP030 minus ASTER-GDEM



e) GMTED2010 minus ASTER-GDEM

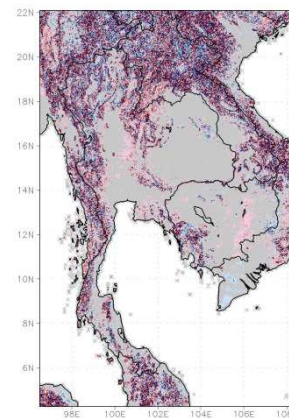
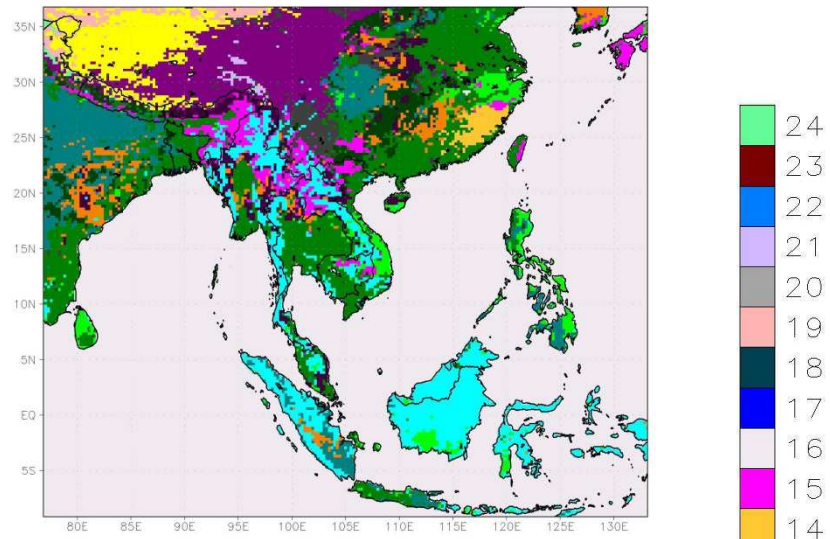


Figure A3 Terrain height in Domain 3.

a) Default



b) Updated

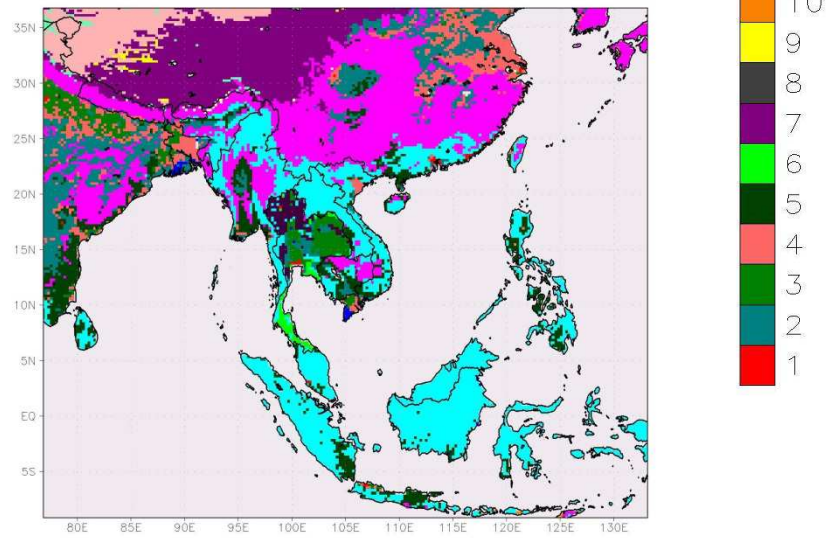
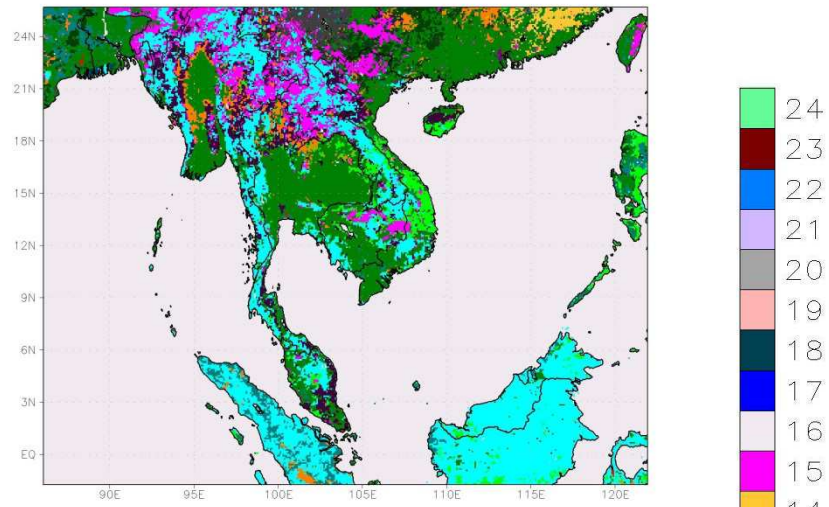


Figure A4 Dominant land use class in Domain 1.

a) Default



b) Updated

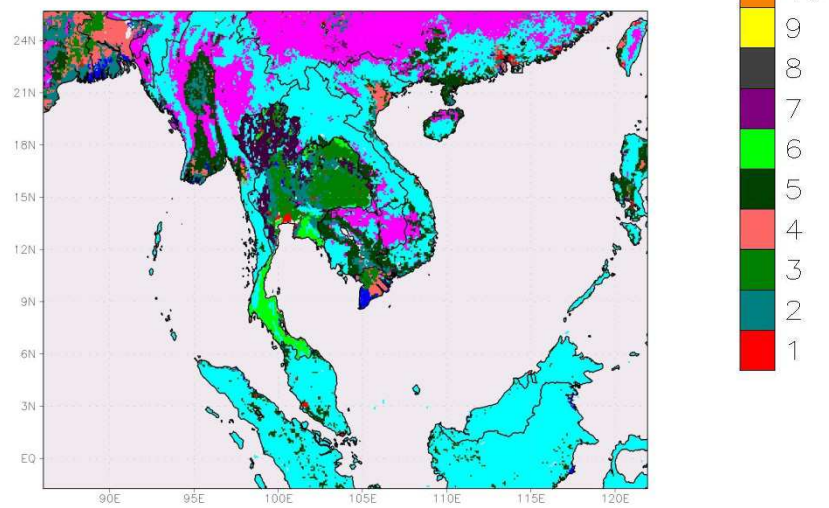
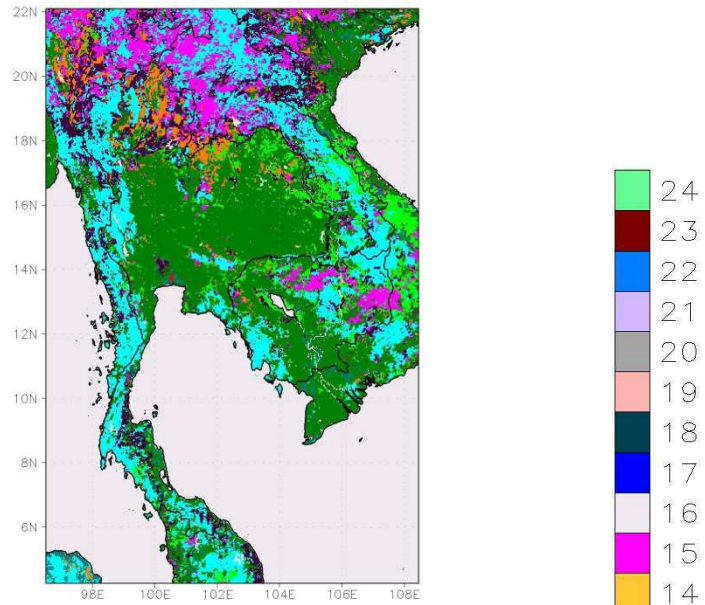


Figure A5 Dominant land use class in Domain 2.

a) Default



b) Updated

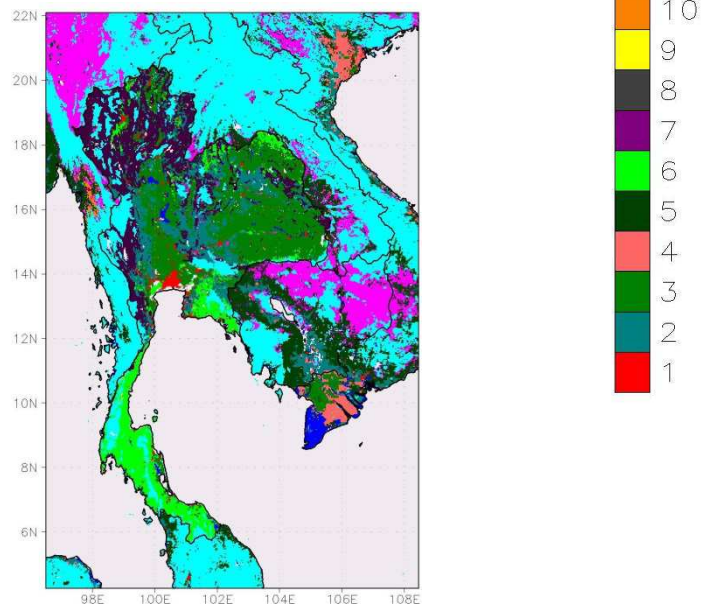
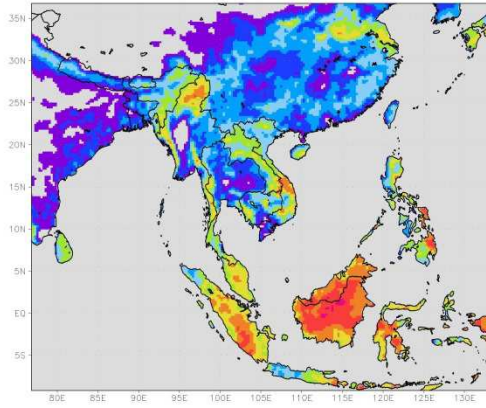
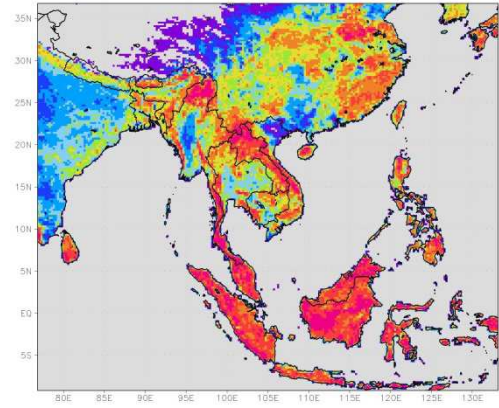


Figure A6 Dominant land use class in Domain 3.

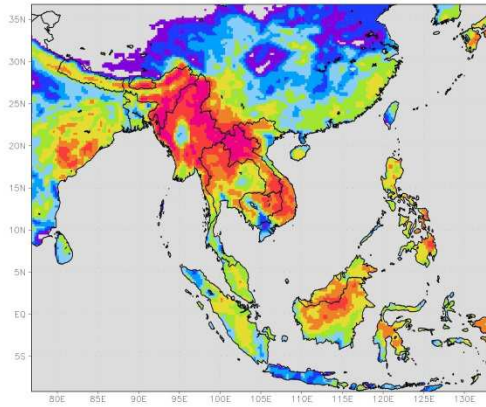
a) Default, April



b) Updated, April



c) Default, October



d) Updated, October

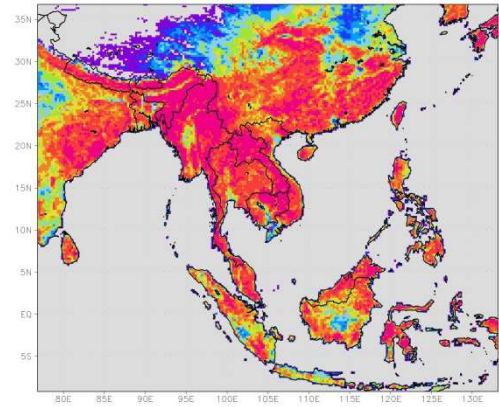
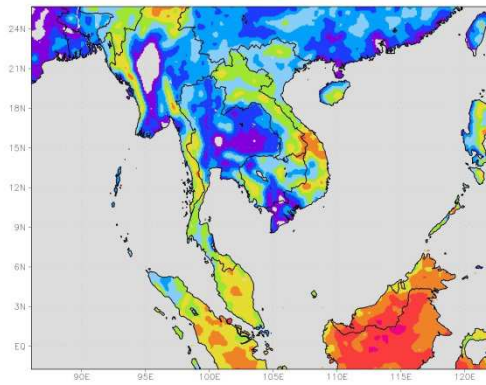
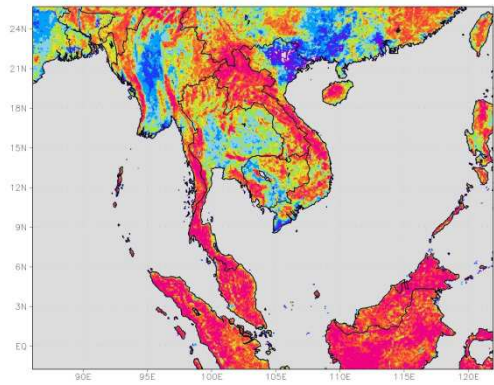


Figure A7 Green fraction in Domain 1.

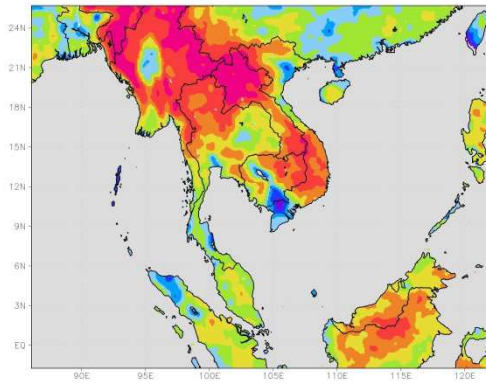
a) Default, April



b) Updated, April



c) Default, October



d) Updated, October

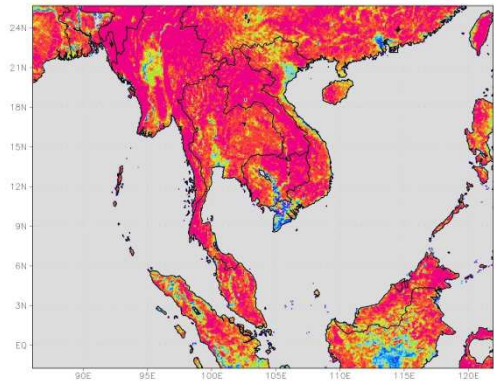
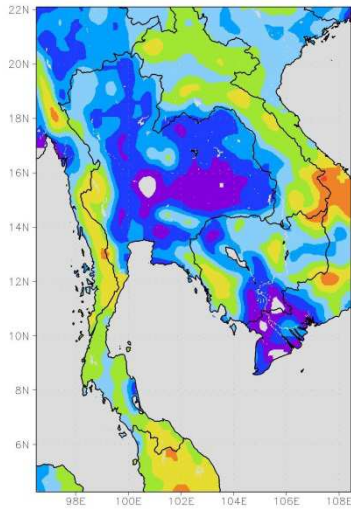
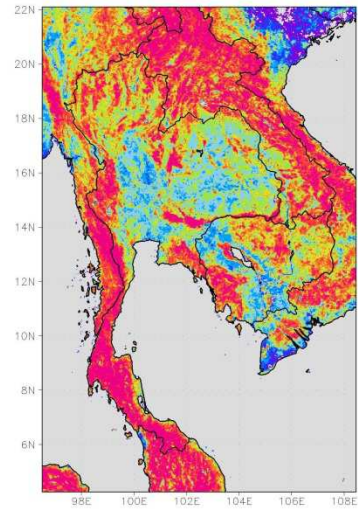


Figure A8 Green fraction in Domain 2.

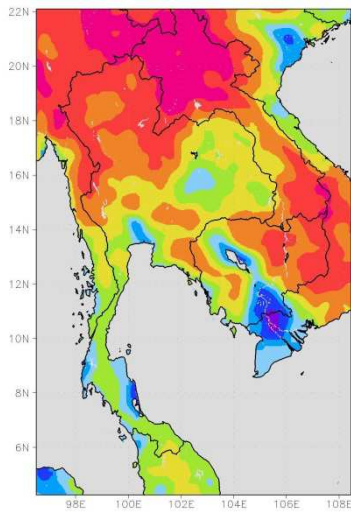
a) Default, April



b) Updated, April



c) Default, October



d) Updated, October

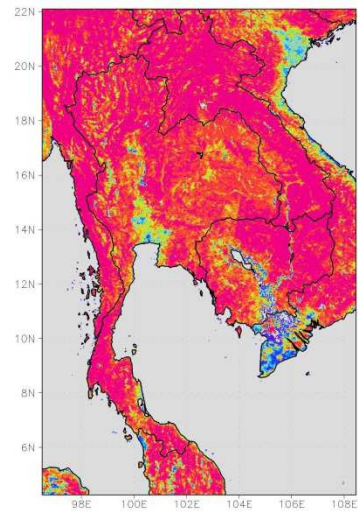
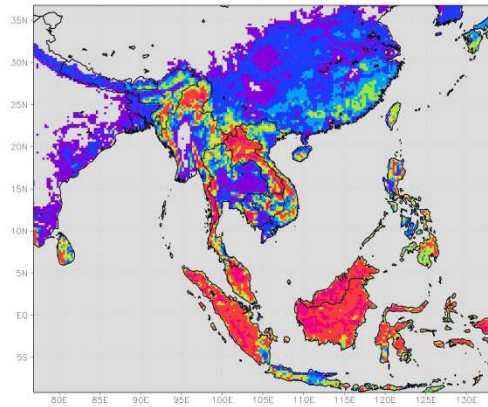
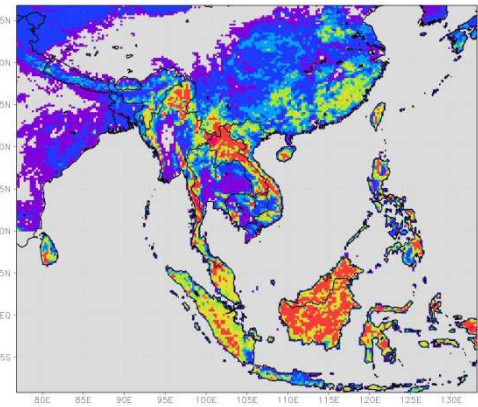


Figure A9 Green fraction in Domain 3.

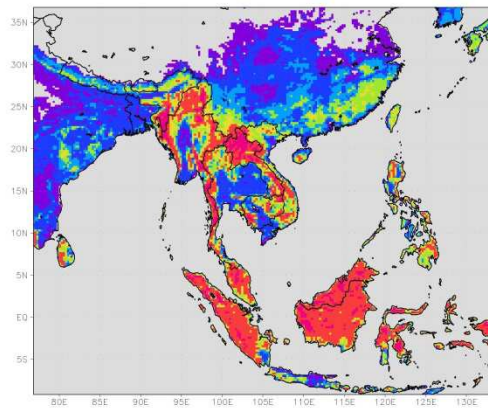
a) Default, April



b) Updated, April



c) Default, October



d) Updated, October

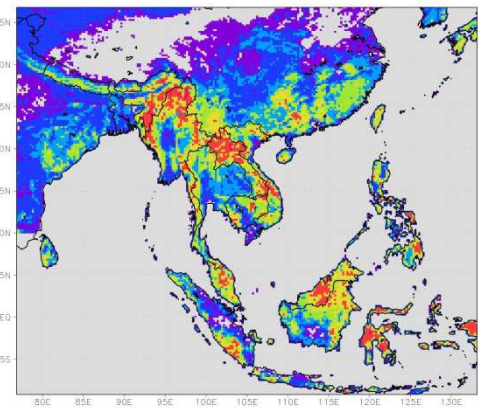
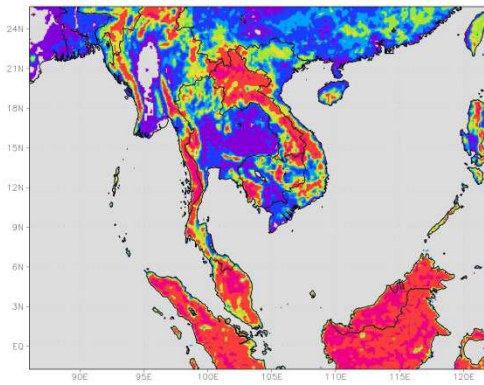
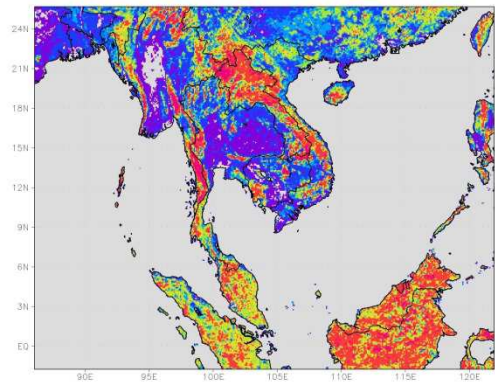


Figure A10 Leaf area index in Domain 1.

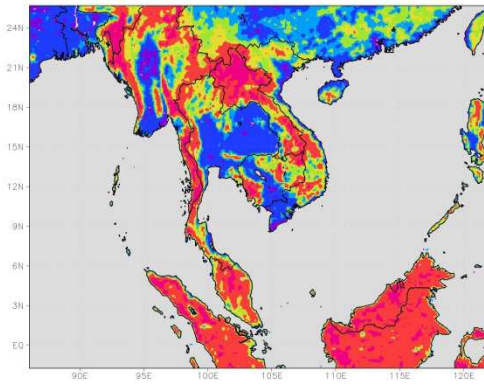
a) Default, April



b) Updated, April



b) Default, October



e) Updated, October

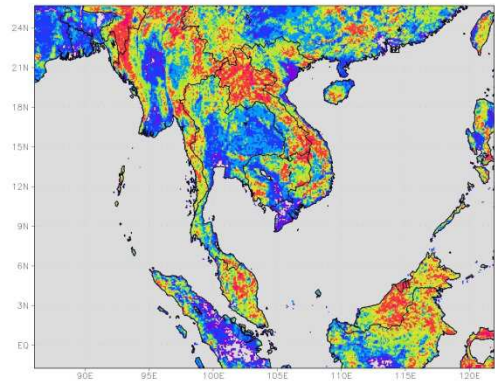
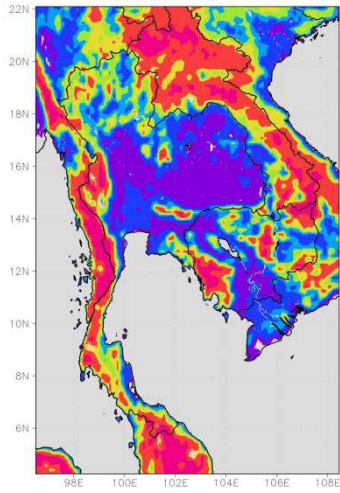
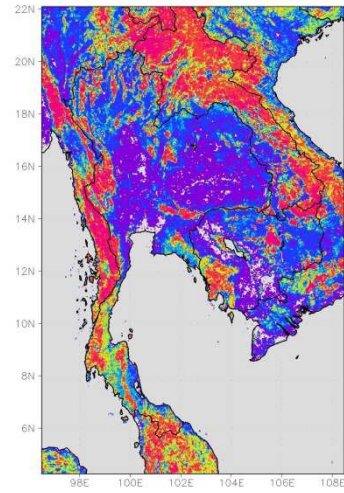


Figure A11 Leaf area index in Domain 2.

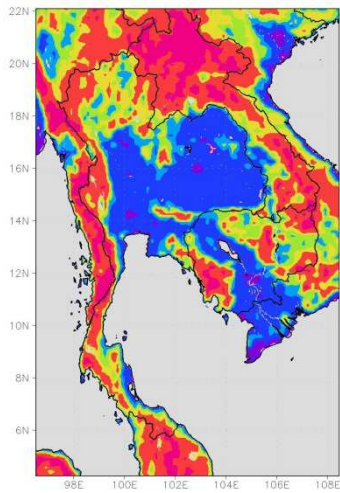
a) Default, April



b) Updated, April



c) Default, October



d) Updated, October

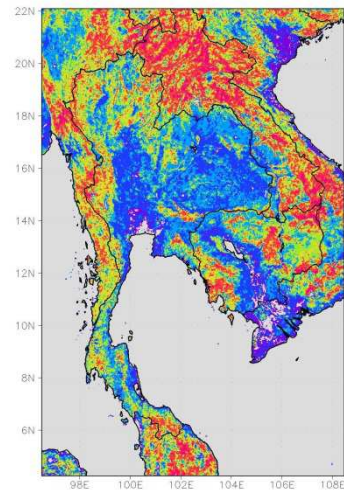
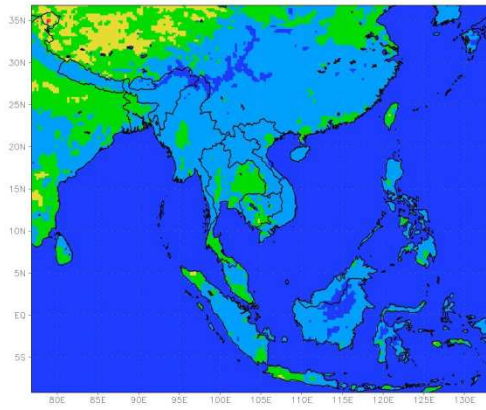
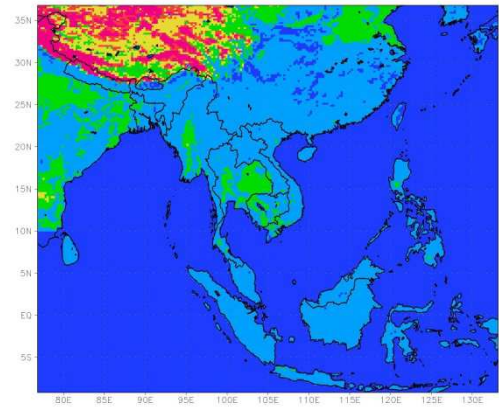


Figure A12 Leaf area index in Domain 3.

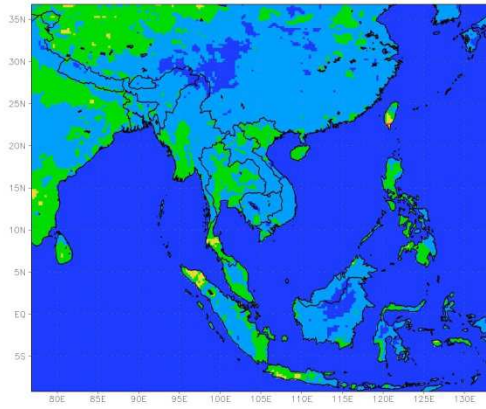
a) Default, April



b) Updated, April



c) Default, October



d) Updated, October

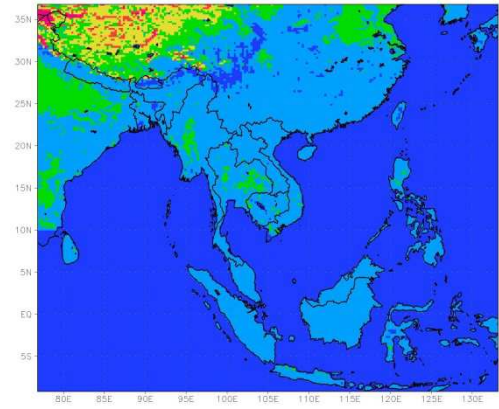
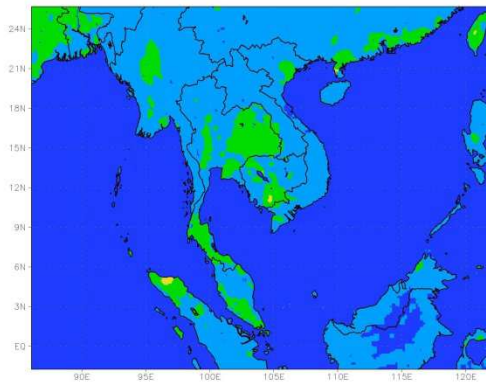
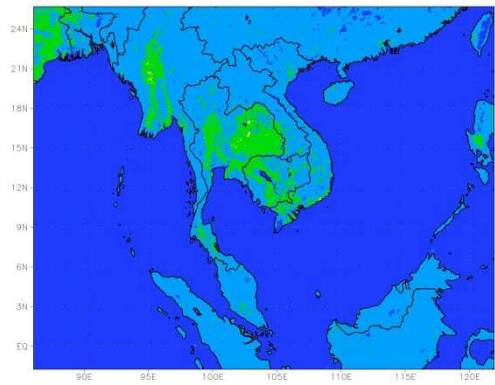


Figure A13 Surface albedo in Domain 1.

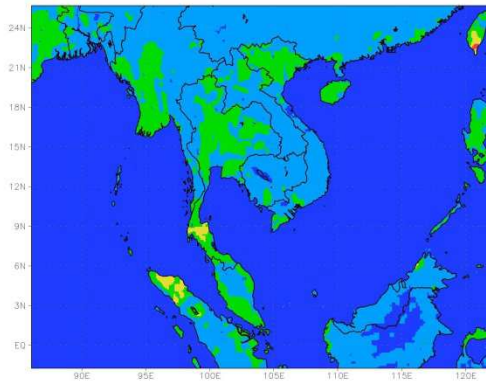
a) Default, April



c) Updated, April



d) Default, October



e) Updated, October

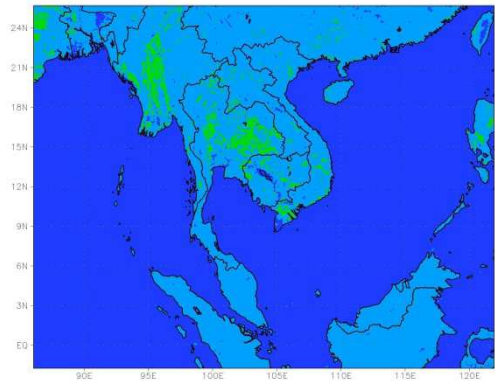
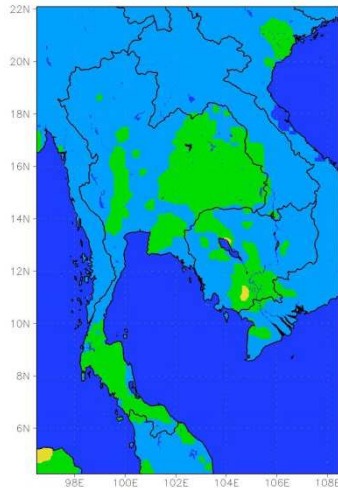
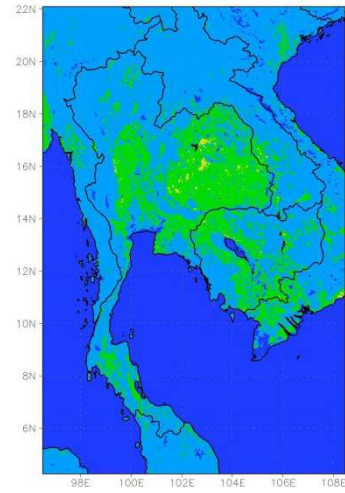


Figure A14 Surface albedo in Domain 2.

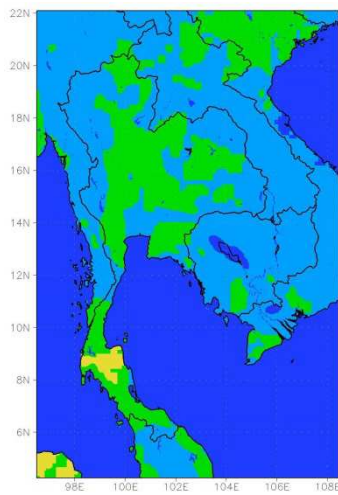
a) Default, April



b) Updated, April



c) Default, October



d) Updated, October

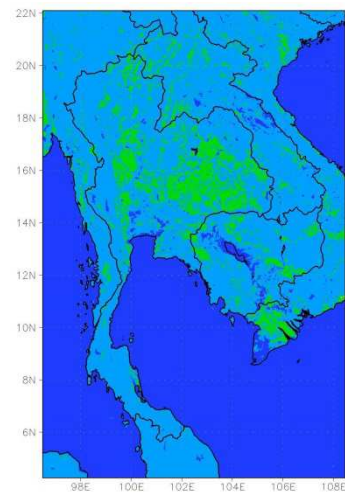
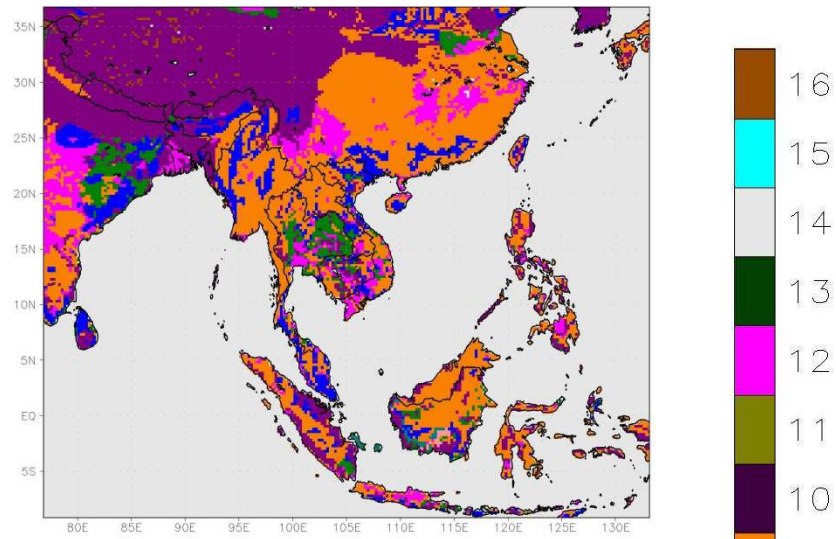


Figure A15 Surface albedo in Domain 3.

a) Default



b) Updated

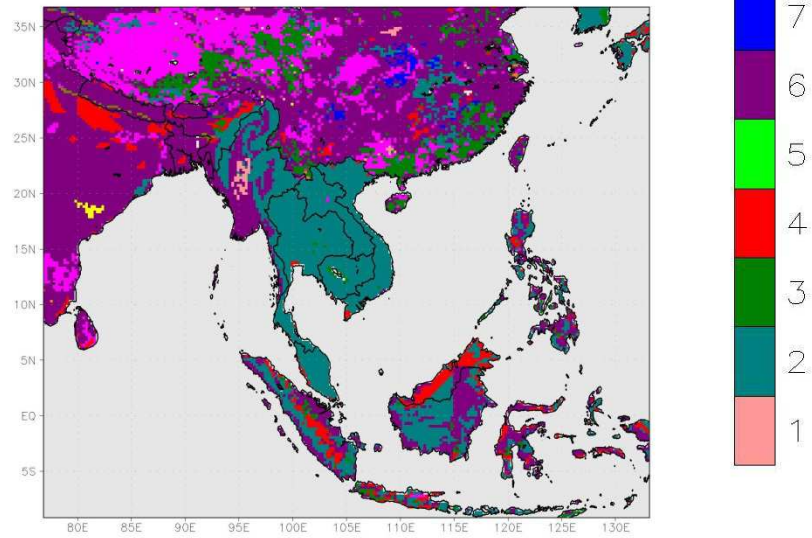
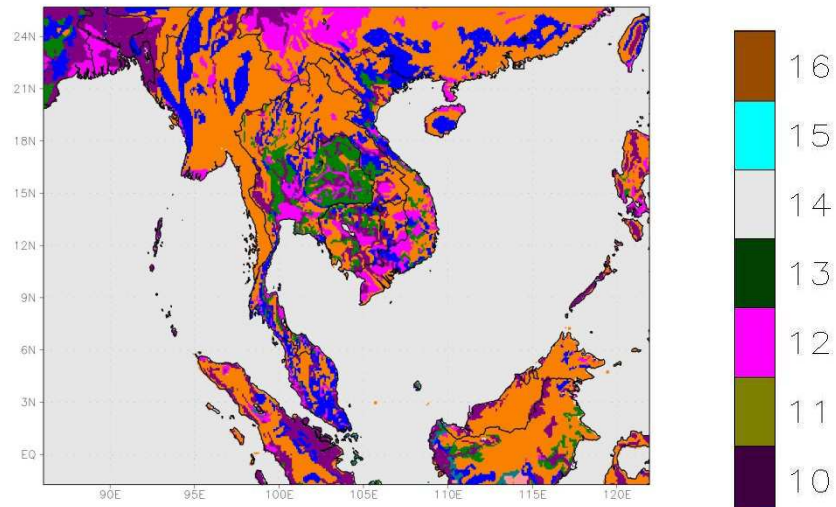


Figure A16 Dominant top-layer soil texture in Domain 1.

a) Default



b) Updated

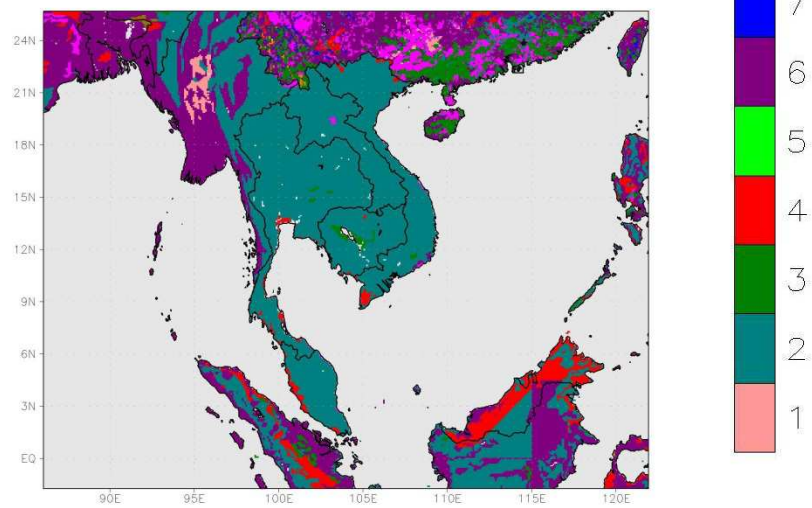
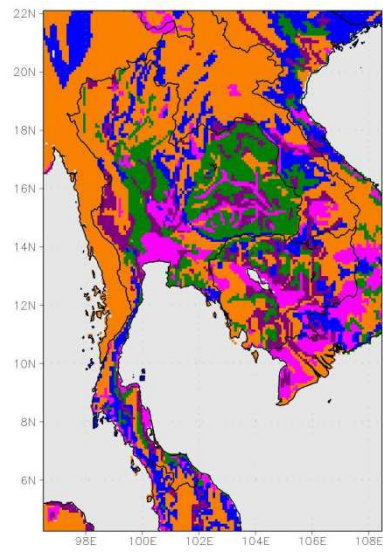


Figure A17 Dominant top-layer soil texture in Domain 2.

a) Default



b) Updated

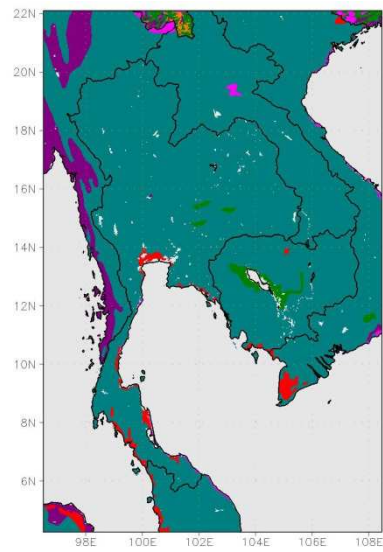
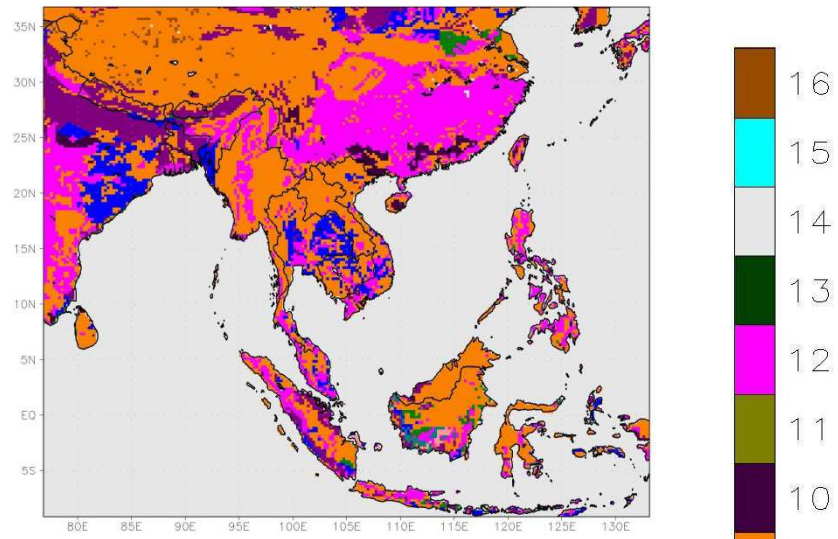


Figure A18 Dominant top-layer soil texture in Domain 3.

a) Default



b) Updated

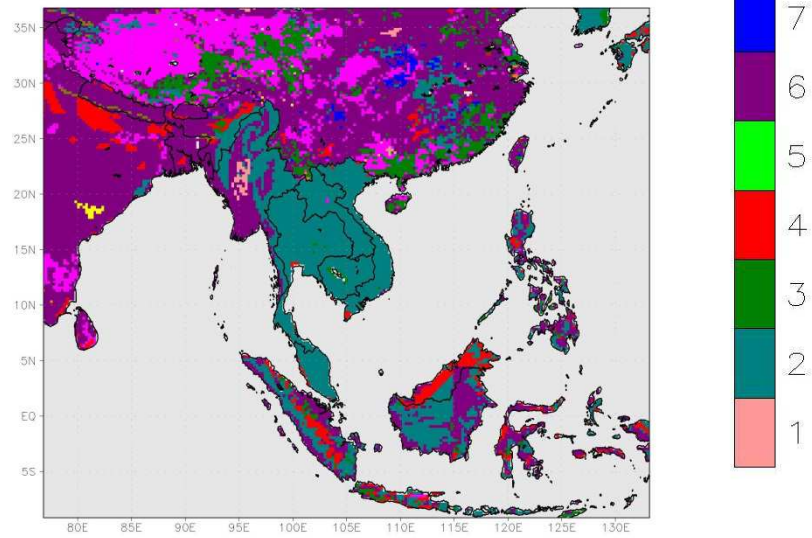
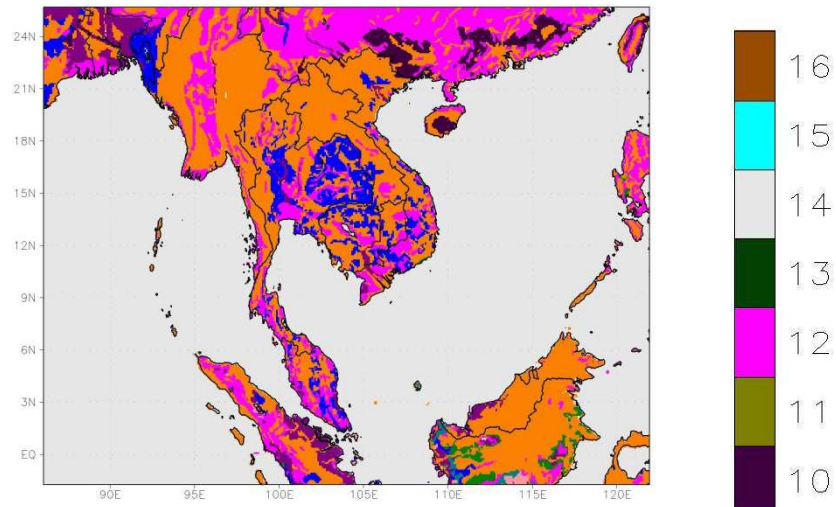


Figure A19 Dominant bottom-layer soil texture in Domain 1.

a) Default



b) Updated

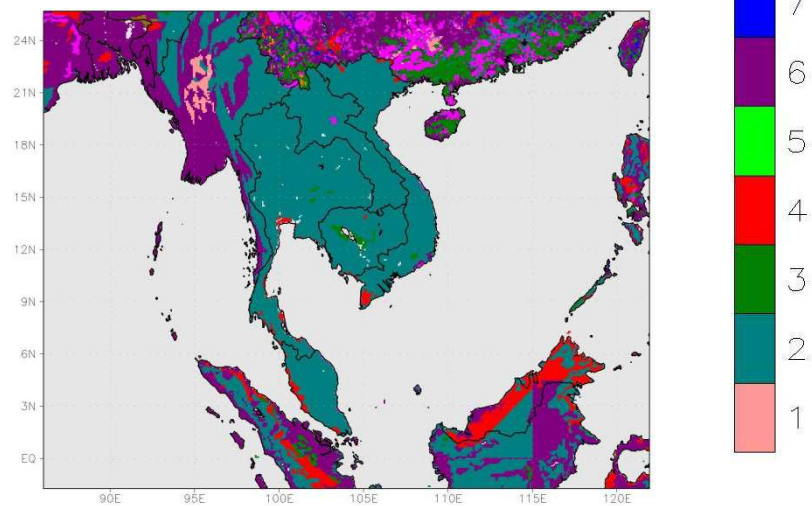
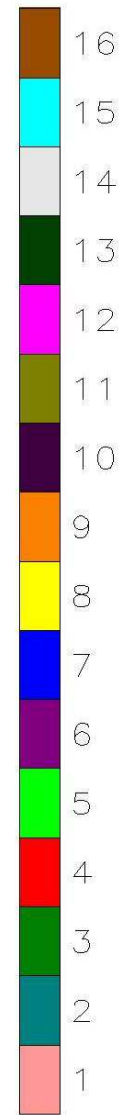
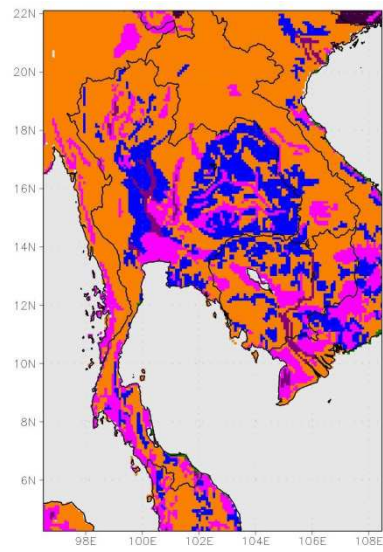


Figure A20 Dominant bottom-layer soil texture in Domain 2.

a) Default



b) Updated

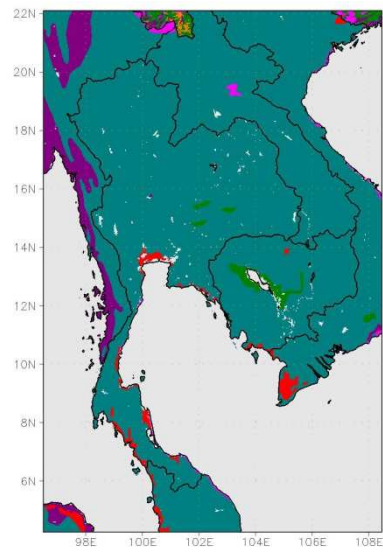


Figure A21 Dominant bottom-layer soil texture in Domain 3.

Researcher Information

1. Kasemsan Manomaiphiboon (PhD) (Thai: เกษมสันต์ มโนมัยพิบูลย์)

Associate Professor

The Joint Graduate School of Energy and Environment, King Mongkut's University of Technology Thonburi

Tel.: 02-470-7331, E-mail: kasemsan_m@jgsee.kmutt.ac.th and kasemsanm@hotmail.com

Education:

- PhD (2004) Environmental Engineering, Georgia Institute of Technology, US
- MS (2003) Industrial Engineering, Georgia Institute of Technology, US
- MS (1999) Environmental Engineering, University of Cincinnati, US
- BEng (1993) Environmental Engineering, Chulalongkorn University, Thailand

Expertise:

Analysis and numerical modeling in the areas of atmospheric science, air quality, weather, climate, and wind energy

2. Surajate Boonya-Aroonnet (DEng) (Thai: สุรเจตส์ บุญญารุณเนตร)

Director of Hydro Informatics Division

Hydro and Agro Informatics Institute, Ministry of Science and Technology

Tel: 02-642-7132 E-mail: surajate@haii.or.th

Education:

- DEng (2005) Environmental and Resource Engineering, Hokkaido University, Japan
- MEng (2002) Integrated Water Resources Management, Asian Institute of Technology, Thailand

- BEng (1990) Civil Engineering, King Mongkut' s University of Technology North Bangkok, Thailand

Expertise:

Hydroinformatics and modeling of water resource systems, hydraulic engineering, operational flood forecasting system, GIS and remote sensing, hydro-meteorological observation, and flood and drought risk management

3. Kanoksri Sarinnapakorn (PhD) (Thai: กนกศรี ศรีนนภากร)

Senior Researcher

Hydro and Agro Informatics Institute, Ministry of Science and Technology

Tel: 02-642-7132 E-mail: kanoksri@haii.or.th

Education:

- PhD (2007) Electrical and Computer Engineering, University of Miami, US
- MS (2002) Computer Science, Fairleigh Dickinson University, US
- MS (1997) Statistics, Kasetsart University, Thailand
- BS (1990) Statistics, Kasetsart University, Thailand

Expertise:

Advanced statistical analysis and data science, mathematical modeling for short-term and seasonal forecasting of rainfall and extreme events, and data management and quality assurance for hydro-meteorological data

4. Nosha Assareh (Thai: โนชา อาซาเรห์)

PhD Candidate

The Joint Graduate School of Energy and Environment, King Mongkut' s University of Technology Thonburi

Tel.: 02-470-7332 E-mail: nosha_assare@yahoo.com

Education:

- MS (2012) Chemical Engineering, Shahid Bahonar University (Kerman), Iran

- BEng (2009) Chemical Engineering, University of Guilan (Rasht), Iran

Expertise:

Analysis and numerical modeling in the areas of fluid mechanics, atmospheric science, air quality, and wind energy

5. Nishit Aman (Thai: นิชิต อามาน)

PhD Candidate

The Joint Graduate School of Energy and Environment, King Mongkut's University of Technology Thonburi

Tel.: 02-470-7332 E-mail: aman.nisit@gmail.com

Education

- MS (2012) Environmental Engineering, Indian Institute of Technology Bombay, India
- BEng (2010) Agricultural Engineering, Tamilnadu Agricultural University, India

Expertise

Analysis and numerical modeling in the areas of atmospheric science and air quality

6. Peraya Tantianuparp (Thai: พิรญา ตันตื้อนุภาพ)

Head of GIS research section

Hydro and Agro Informatics Institute, Ministry of Science and Technology

Tel: 02-642-7132 E-mail: peraya@haii.or.th

Education:

- MA (2002) Geography, Chulalongkorn University, Thailand
- BA (1997) Geography, Thammasat University, Thailand

Expertise:

GIS, photogrammetry, and remote Sensing

7. Thippawan Thodsan (Thai: ทิพย์วรรณ ทอดแสน)

Researcher

Hydro and Agro Informatics Institute, Ministry of Science and Technology

Tel: 02-642-7132 E-mail: thippawan@haii.or.th

Education:

- MS (2011) Information Systems, King Mongkut' s Institute of Technology Ladkrabang, Thailand
- BS (2005) Information System Technologies, Rajamangala University of Technology Rattanakosin, Thailand

Expertise:

Data quality management for hydro-meteorological research, meteorological early warning systems, and weather and climate prediction modeling

8. Aungkana Pratumthong (Thai: อังคณา ประทุมทอง)

Head of Project Management and Coordination Section, Hydro Informatics Division, Hydro and Agro Informatics Institute, Ministry of Science and Technology

Tel: 02-642-7132 E-mail: aungkana@haii.or.th

Education:

- MS (2008) Urban planning, Chulalongkorn University, Thailand
- BS (2004) Geography, Silpakorn University, Thailand

Expertise:

Project management and coordination