FINAL REPORT

Improvement of MM5 Surface Characteristics

August 31, 2008

Submitted by:

Daewon W. Byun, PI
Fong Ngan, Fang-Yi Cheng, Hyun-Cheol Kim, and Soontae Kim

Institute for Multidimensional Air Quality Studies
University of Houston
4800 Calhoun Rd.
Houston, TX 77204

PREPARED FOR

Texas Commission on Environmental Quality

Project Representative: Clint Harper (e-mail: charper@tceq.state.tx.us)
Technical Monitor: Doug Boyer (e-mail: dboyer@tceq.state.tx.us)

TCEQ - Air Quality Division
P.O. Box 13087/MC-164
Austin, TX 78711-3087
Office: 512.239.1674, fax: 512.239.1500

Project Summary: Improvement of MM5 Surface Characteristics

Accurate representation of surface characteristics in a mesoscale model is critically important to the production of meteorological inputs used for air quality modeling. The MM5 meteorological model, originally designed for regional meteorological modeling, requires various model inputs characterizing the land-surface conditions. Technological improvements have been made to generate improved inputs such as topography, land use and land cover, vegetation, and subsoil classification. Incorporation of these data in mesoscale modeling is expected to improve performance of meteorological and air quality modeling. These improvements are necessary to support the next State Implementation Plan (SIP) submissions required by the federal Clean Air Act.

In 2007, University of Houston developed a method to incorporate the high resolution satellite observation-based land use and land cover (LULC) data from the University of Texas Center for Space Research (UT-CSR) for the MM5 modeling of the East Texas area. In the present project, UH has improved key model inputs for MM5 characterizing the land surface conditions such as land-sea masks, vegetation fractions, and clay subsoil types to be consistent with the highly accurate LANDSAT-based LULC data for the Eastern Texas region. We have also improved the method of handling missing data in satellite observations to generate more consistent sea surface temperature inputs for the MM5 simulations.

Several MM5 simulations were performed for the TexAQS2000 episode for the 12-km and 4-km grid resolution domains with different combinations of input changes representing the improved surface characterization data. The sensitivity simulations show only small differences in the regional average time series plots of temperature and wind at the CAMS locations, but larger differences exist for the ground temperature at individual sites where vegetation fractions and soil classifications are changed. Spatial analysis of the sensitivity simulations reveals distinguishable effects of the input changes such as the convergent flow patterns and timing of sea breeze and thunderstorm events as well as distribution and magnitudes of daily maximum temperatures. Although additional evaluation with high resolution in situ and remote sensing measurements are needed, it seems that use of improved surface data affects simulation of mesoscale phenomena that are important in characterizing air pollutant source-receptor relations.

In 2007 we also reported the procedure to interpolate the sea surface temperature (SST) measurement from geostationary satellites. Since then, we have made two improvements in the processing of SST for the MM5 simulations. One is expansion of the time span to fill in missing satellite observations and the other is improvement of the spatial interpolation method to eliminate idiosyncratic patterns caused by cells with empty satellite data points in the interpolated SST fields.

One of the future projects would be incorporating the additional high resolution LULC data currently being assembled for the north-eastern Texas region by TAMU. Also, as

NCAR has stopped further development of the MM5 system, it would be necessary to prepare for the transition to the Weather Research and Forecasting (WRF) model as the meteorological driver for air quality modeling. We need to develop a localized WRF modeling system that inherits and further improves the current MM5 versions used by TCEQ and UH for modeling the Eastern Texas region. A future task extending the present work could be implementation of the LULC dataset into the WRF meteorological modeling system. For the last several years, University of Houston (UH) has been comparing the performance of WRF and MM5 simulation results. The main objectives of the new project would be to improve the land-surface and vegetation databases in WRF, and to perform comparative evaluation between the WRF and MM5 results.

List of Content

PROJECT SUMMARY……...…ii
1. Introduction 1
2. Tasks and Accomplishments 2
3. Improvement of Surface Characteristic Datasets for the MM5 Simulations with the High-Resolution LULC Data 4
   3.1 Topography data 4
   3.2 Improving land-water mask data 7
   3.3 Improving vegetation fraction data 9
   3.4 Improving clay soil type in MM5 16
   3.5 Processing of LULC, vegetation fraction, and soil type datasets for use in TERRAIN program 21
4. MM5 Simulations with the Updated Surface Characteristics 23
   4.1 Model configurations 23
   4.2 Time series analysis 24
   4.3 Spatial analysis 28
5. Processing of Sea Surface Temperature and MM5 Simulations 34
   5.1 Handling missing data 34
   5.2 Pre-interpolation of SST 35
6. Conclusive Remarks 36
7. References 38

List of Figures

Figure 1. (a) Terrain height (m) at 12-km domain with smoothing (b) Difference plot between the terrain height w/ and w/o smoothing. 5
Figure 2. (a) Terrain height (m) at 4-km domain with smoothing (b) Difference plot between the terrain height w/ and w/o smoothing. 6
Figure 3. (a) UT-CSR LULC data at 4-km domain and (b) is a smaller domain and also layered with the land/water mask data (land: black dot, water: pink circle). 8
Figure 4. Monthly vegetation fraction value from (a) WRF and (b) MM5 meteorological modeling systems. 10
Figure 5. (a) Dominant LULC type composite of the original MM5 and UT-CSR LULC datasets, (b) vegetation fraction derived from the composite LULC dataset, and (c) vegetation fraction in the original MM5 for the Texas 12-km resolution domain. 13
Figure 6. Difference of UTCSR-derived vegetation fraction and original MM5 vegetation fraction for the Texas 12-km resolution domain. 14
Figure 7. (a) Dominant LULC type composite of the original MM5, UT-CSR and TFS LULC datasets, (b) vegetation fraction derived from the composite LULC dataset, and (c) vegetation fraction in the original MM5 for the East Texas 4-km resolution domain. 15
Figure 8. Dominant soil type at 4-km resolution. 17
Figure 9. Blue color shows the areas with “water” LULC type from the UT-CSR LULC data at 30-m resolution. 18
Figure 10. (a) Clay soil type (red color) layered with the fractional water land use type from UT-CSR data for the 12-km resolution domain. (b) is same as (a) but for the 4-km resolution domain. 19
Figure 11. 4-km resolution soil type distribution from (a) the updated subsoil database with two new types defined by water fraction obtained from UT-CSR data and (b) the original MM5 subsoil database 20
Figure 12. Domain configuration of the 3-nested domains used for MM5 simulations. 23
Figure 13. Time series comparison of regional average 1.5 m temperature, wind speed and wind direction at 12-km domain for the Houston area. 24
Figure 14. Same as Figure 13 but for the Beaumont area. 25
Figure 15. Same as Figure 13 but for the Dallas area. 26
Figure 16. Time series comparison of 1.5 m temperature at 4-km domain at CAM site 112 (Beaumont) and 167 (Houston). 27
Figure 17. Same as Figure 16 but for wind speed at CAMS site 131 (Beaumont) and 603 (Houston). 27
Figure 18. Same as Figure 16 but for wind direction at CAMS site 131 (Beaumont) and 603 (Houston). 28
Figure 19. Difference plots of ground temperature for 12-km (left panel) and 4-km (right panel) resolution domain at 18 UTC 30 August 2000. 30
Figure 20. Difference plots of PBL height for 12-km (left panel) and 4-km (right panel) resolution domain at 23 UTC 30 August 2000. 31
Figure 21. Difference plots of horizontal wind at 1st layer for 12-km at 2 UTC 25 August 2000 (left panel) and 4-km (right panel) resolution domain at 23 UTC 30 August 2000. 32
Figure 22. Difference plots of horizontal wind at 1st layer (left panel) and precipitation for 4-km resolution domain at 17 UTC 23 August 200 33
Figure 23. Steps of GOES SST retrieval. (a) Raw GOES SST measurement at Sep. 1, 2006 17:00, (b) SST averaged between 15:00 ~ 19:00 is added. (c) Monthly averaged diurnal SST variation is added. 35
Figure 24. Example of SST pre-interpolation for fine resolution grid cells. Left panels show (a) 4km CMAQ domain grids and locations of GOES SST measurements, and (c) gridded SST. Shaded cells in (a) indicates grid cells with available SST measurements. Right panels show (b) doubled SST points by interpolation, and (d) corresponding SST retrieval in grid cells. 36

List of Table

Table 1. LULC categories from UT-CSR data. σ is the shading factor. 11

1. Introduction

The purpose of this project is to improve the Penn State/National Center for Atmospheric Research Mesoscale Model v. 5 (MM5) meteorological modeling system for East Texas, including Dallas-Fort Worth and the Houston-Galveston Area, and provide advanced training in MM5 to the modeling staff from the Texas Commission on Environmental Quality (TCEQ). The land-surface process in MM5 represents momentum, heat, and moisture exchange between the earth's surface and atmosphere. In combination with the planetary boundary layer schemes, it realizes the effects of the earth's surface on local, regional, and global weather and climate. This process requires a precise description of surface characteristics such as land-water demarcation and land use, vegetation, subsoil distribution, and topographic height. Key physical parameters affecting transfer of heat, momentum, and soil moisture are tabulated for each land use type. MM5 relies on a few key base datasets to describe these surface conditions.

The University of Houston (UH) Institute for Multidimensional Air Quality Studies (IMAQS) has demonstrated that meteorological inputs for air quality models can be improved by use of high resolution (~30 m) satellite-based land use (LU) and land cover (LC) data (Cheng and Byun, 2008; Cheng et al., 2008). For performing meteorological simulations for the Texas Air Quality Studies (TexAQS-2000 and 2006) domains, we have further improved the LULC dataset for the East Texas and Houston area through contributions from the University of Texas Center for Space Research (UT-CSR) as well as assistance from TCEQ, Texas Forest Service (TFS), and Houston Advanced Research Center (HARC).

For the previous project with TCEQ in 2007, University of Houston - Institute for Multidimensional Air Quality Studies (UH/IMAQS) implemented a method to utilize the UT-CSR high-resolution satellite observation based LULC data (Wells, 2006) in the MM5 system. Under the current project, UH has improved other key inputs to the MM5 system characterizing the land surface conditions, such as land-water mask, subsoil type data, and vegetation fraction data, to be consistent with the UT-CSR LULC datasets. The data and reports produced under this Grant Activity are to support the State Implementation Plan, computer modeling studies, and the goals of the TCEQ and State of Texas.

Section 2 describes the tasks and deliverables, Section 3 provides details of the surface input datasets that are updated. In section 4, MM5 meteorological modeling results with the improved input datasets are compared, and in Section 5 a few modifications made on the processing method of the GOES satellite-observed sea surface temperature are described. Conclusive remarks and future directions of the related research are presented in Section 6.

2. Tasks and Accomplishments

Task 1 - Grant Activity Description:

Deliverable 1: TCEQ approved Grant Activity Description

UH submitted the Grant Activity Description on February 8, 2008, and the final version was approved by TCEQ on February 19, 2008.

Task 2 - Monthly Status Report:

Deliverables 2: Monthly Status Reports

UH submitted monthly status reports to the TCEQ. They include a summary of events and activities completed during the previous month (past 30 calendar days) and planned events and activities for the next month (next 30 calendar days).

Task 3 – Collection of available high resolution terrain and subsoil datasets

Deliverables 3**:** High resolution terrain data and soil data

Coat et al. (2007) mentioned that meteorological models typically smooth the terrain height in order to avoid the formation of spurious waves due to 2 dx distances interacting with advection schemes. This smoothing causes errors at the fine grid scale. Following Coat et al. (2007), we re-derived the terrain height data from the same USGS generated digital elevation model (DEM) terrain data without smoothing the real features and compared.

We have searched for several soil data sources. The best data source is the FAO 5-minutes global soil data and STATSGO databases from the Natural Resources Conservation Service (NRCS), United States Department of Agriculture (USDA): http://www.ncgc.nrcs.usda.gov/products/datasets/statsgo/. After investigating the suitability of the data for MM5 modeling, we realized that the MM5 relies on similar data sources.

Task 4 - Development of methods and tools to implement updated topographic and soil data into the MM5 meteorological model

Deliverables 4**:** Tools (procedures and code), methods, and training to implement topographic and soil data into MM5

UH developed tools and methods for the implementation of the high resolution terrain data and soil data into MM5. Section 3 of this report provides a detailed description of

the processing method used to improve the use of the surface databases for the simulation of the MM5 system in the Eastern Texas area.

Task 5 – Quality Assurance of datasets, fine-tuning of methods and tools

Deliverables 5**:** All quality assured data, improved tools (procedures and code) and methods, summary of MM5 simulation results, training to implement topographic and soil data into MM5, and quality assurance report

UH made substantial efforts to ensure quality of data, and performed all necessary fine-tuning of tools and methods to implement the high resolution terrain and soil data through repeated MM5 simulations. UH provided two training sessions to the TCEQ modeling staff on the code changes and use of the tools. This report summarizes information on the program code modifications and analysis graphics used as part of the quality assurance activities.

Task 6 – Advanced Meteorological Modeling Training

Deliverables 6: Training of TCEQ staff on the advanced meteorological modeling

For the current project UH/IMAQS provided two training sessions to TCEQ staff on the implementation of these and other advanced improvements to the MM5 model, such as the use of satellite-based modeling inputs, data assimilation, and model evaluation tools.

Task 7 – Final Report:

Deliverables 7**:** Final report

Together with this final report, UH provides the following as the final project products:

- PowerPoint presentation file used for the TCEQ training sessions
- Animation files of surface temperature, wind, planetary boundary layer height, and cloud fraction
- Source code and run scripts for updating the LULC, vegetation fraction and soil type data and the soil parameter table used for the MM5 simulations

3. Improvement of Surface Characteristic Datasets for the MM5 Simulations with the High-Resolution LULC Data

The TERRAIN program of the MM5 system analyzes the necessary data from the latitude-longitude interval for the chosen mesoscale domains, including terrain data, mean deep-soil temperature, vegetation fractions, and subsoil types. In addition to constructing the model domain settings, the program performs nest interface adjustment and feedback for multiple domain nest simulations and outputs the terrain elevation and land use for the subsequent MM5 land-surface model (LSM) processing. In order to represent the surface exchange processes properly, all the base datasets for the TERRAIN program must be improved. Occasionally, a few various terrain datasets in the MM5 system are not self-consistent with the land use/land cover data and vegetation data, such as the vegetation fraction data. The initial expectation was that it would correspond to the land use data, but the TERRAIN process does not necessarily generate consistent maps between the two. There is also no clear description on the resolution and quality of the soil type and deep soil temperature datasets in the TERRAIN documentation.

This section describes how we improved three datasets in the TERRAIN program of MM5 (land-water mask, vegetation fraction, and subsoil types) utilizing the UT-CSR satellite observation-based high resolution LULC data.

3.1 Topography data

Meteorological models often smooth the underlying terrain for spatial scales of at least three times its original resolution. Therefore, it is essential to utilize a set of topography data that are spectrally consistent across the target grid resolutions of nesting domains. Because in MM5 the minimum terrain smoothing permitted is $3 \cdot \Delta x$ averaging, nominal 4-km grid spacing actually corresponds to the use of 12-km averaged terrain height (Coats, 2007).

Five resolutions of the source elevation data (i.e., 1-degree, 30-, 10-, 5-minute, and 30-second), together with the land-water mask files, are currently available. Figure 1a shows the terrain height at the 12-km domain with the smoothing treatment inside TERRAIN preprocessing. Figure 1b is the difference plot of the terrain height (smoothing minus no smoothing). Overall, the western area of the domain shows the most difference where the terrain height is above 1200 m. But near Houston and the surrounding areas, the difference is less than 20-m. Figure 2 is the same as Figure 1 but for the 4-km domain. Since the difference is not apparent in our simulation domain, we have decided not to change our current terrain preprocessing.

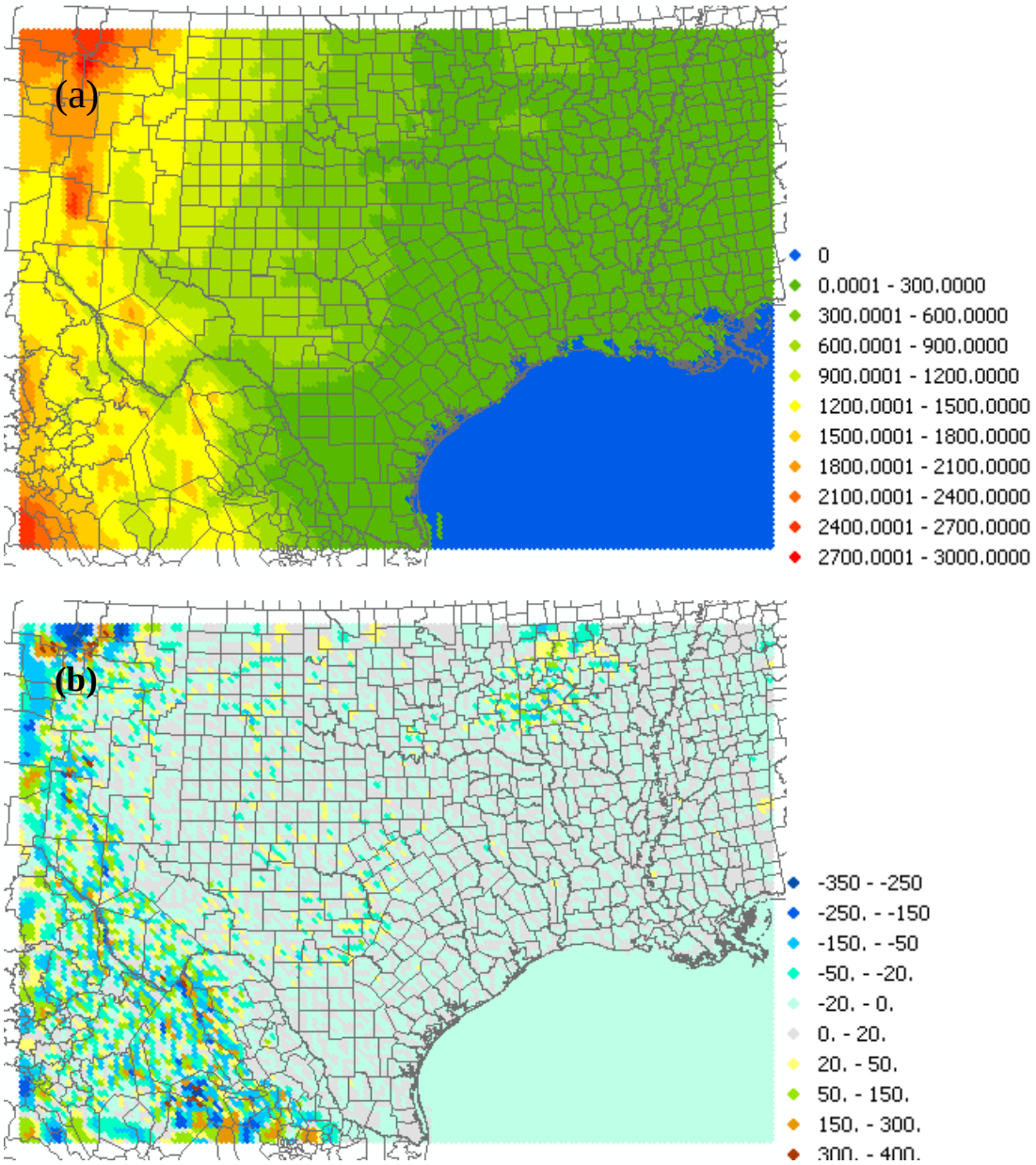


Figure 1. (a) Terrain height (m) at 12-km domain with smoothing (b) Difference plot between the terrain height w/ and w/o smoothing.

During the data collection process, we realized that the meteorological simulations for the Eastern Texas domain will not be much affected by the use of improved topography data other than yielding more accurate water-land demarcations. Because of the critical importance of the vegetation fraction input in representing the evapo-transpiration processes in the MM5 land surface module (LSM) in MM5, we instead concentrated our efforts on producing a reasonable vegetation fraction map for MM5 modeling.

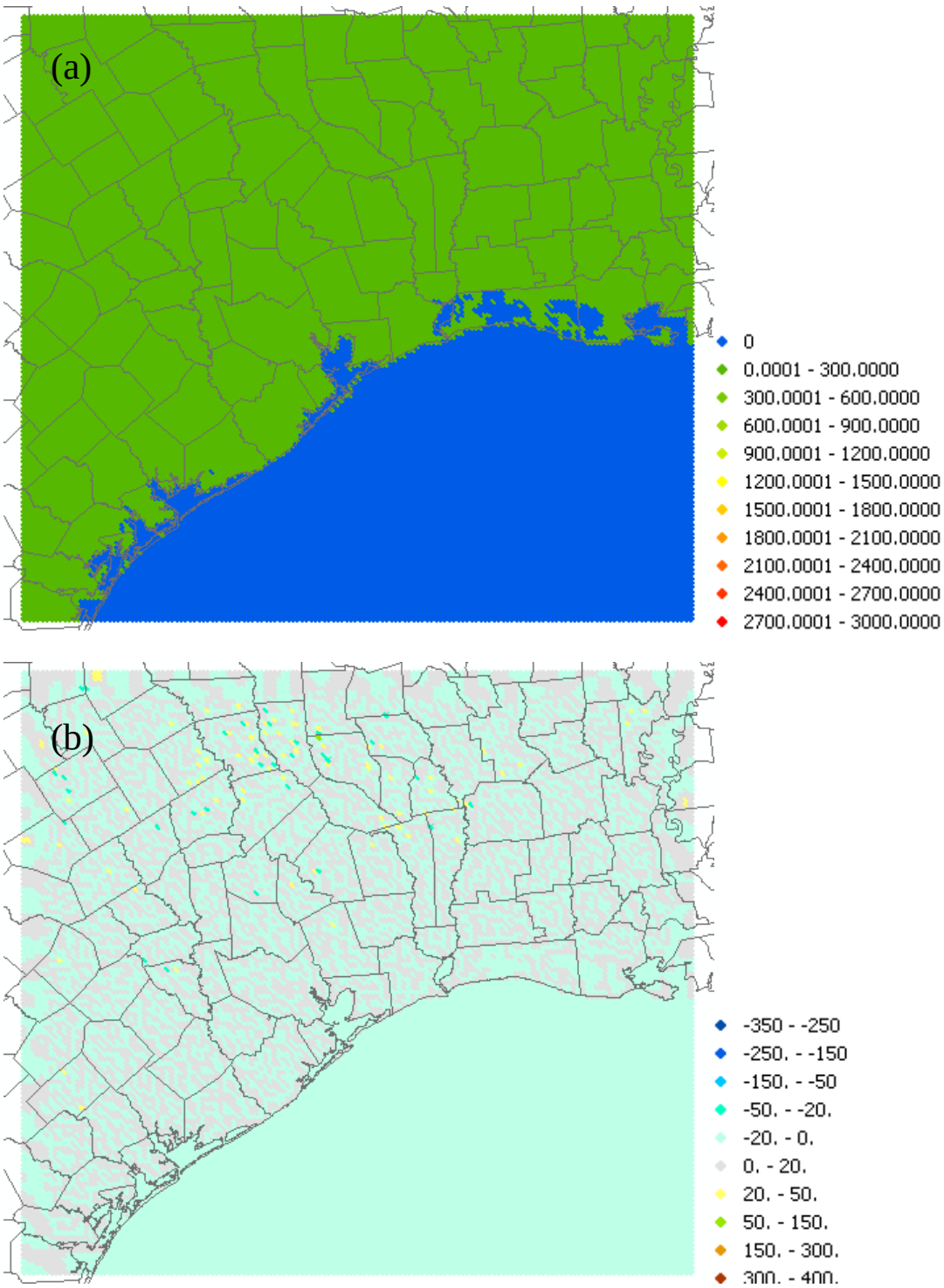
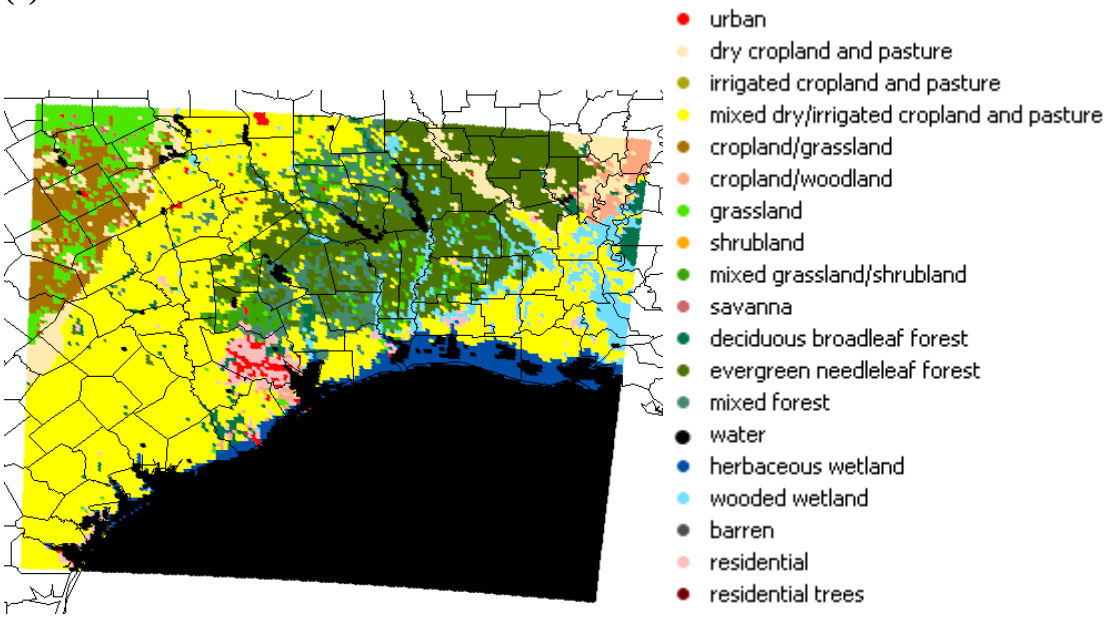


Figure 2. (a) Terrain height (m) at 4-km domain with smoothing (b) Difference plot between the terrain height w/ and w/o smoothing.

3.2 Improving land-water mask data

Instead of focusing on the terrain height, we noticed that there are some differences in the way land and water surfaces are represented between the UT-CSR LULC data and the land-water mask data, which is basically a flag maintained inside TERRAIN and used in MM5 LSM. MM5 uses different surface exchange algorithms for the sea and land. In order to better delineate the land-water boundary, we have relied on the UT-CSR LULC data to re-derive the land-water mask. Figure 3a shows the UT-CSR LULC covering at 4-km domain. Water is identified as a black color. In Figure 3b, we layered out land/water mask data from the default MM5 database where land is identified as small black dots and water is identified as pink circles. Grid points where total black points appeared locate where the land-water mask data from the default MM5 database is not matching with the land/water types in the UT-CSR LULC data. We identified mismatching grid cells with dark circles. Although only around 30 cells show inconsistency, they are along critical areas where significant emissions sources are located. Therefore, we have decided to replace the land-water mask in the MM5 database with those based on the UT-CSR LULC data.

(a)



(b)

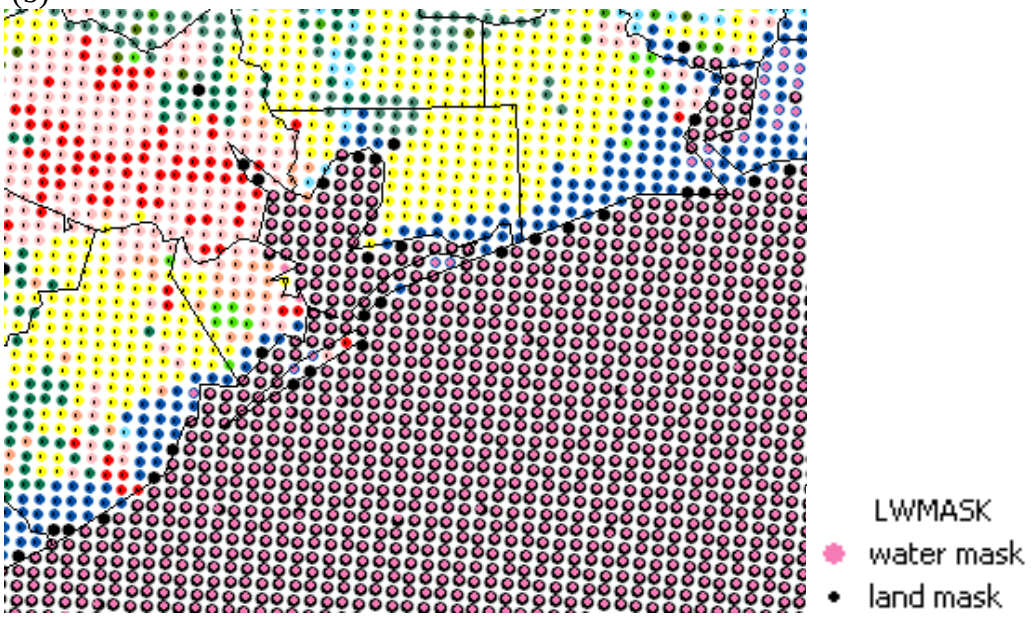


Figure 3. (a) UT-CSR LULC data at 4-km domain and (b) is a smaller domain and also layered with the land/water mask data (land: black dot, water: pink circle).

3.3 Improving vegetation fraction data

The Noah land surface model in MM5 makes use of vegetation fraction and soil type in handling evapotranspiration. During the modeling of the Eastern Texas area with MM5, we have noticed an artificial-looking surface temperature distribution near the Beaumont and Louisiana coastal areas. We traced the cause of this bias back to the strangely low values of vegetation fraction stretching near the coastal areas in the MM5 system. An alternative data source of the vegetation fraction is utilizing the data from the WRF system. Although the data sources are expected to be the same between the two systems, the procedures to derive the vegetation fraction data are different. Figure 4a and Figure 4b are from WRF and MM5 systems, respectively, and show the difference in the vegetation fraction data. The overall distribution is quite similar everywhere except for the coastal area. The strange straight line of very low vegetation value near the Beaumont and Louisiana coasts in the MM5 system does not appear in the WRF system.

In this project, the vegetation fraction data for the MM5 system were reconstructed to be consistent with the high-resolution TCEQ (Wells, 2006) land use and land cover dataset. One of the limitations of utilizing the UT-CSR LULC data is that we only collected data during the summer months. In MM5, the original vegetation fraction data is provided from the monthly 5-year climatology data derived from the Advanced Very High Resolution Radiometer (AVHRR) (Gutman and Ignatov, 1998) satellite data. In this study, we have generated a fixed 12 monthly vegetation fraction; therefore, the application of this data to drive the LSM inside MM5 should be restricted to the summer months only.

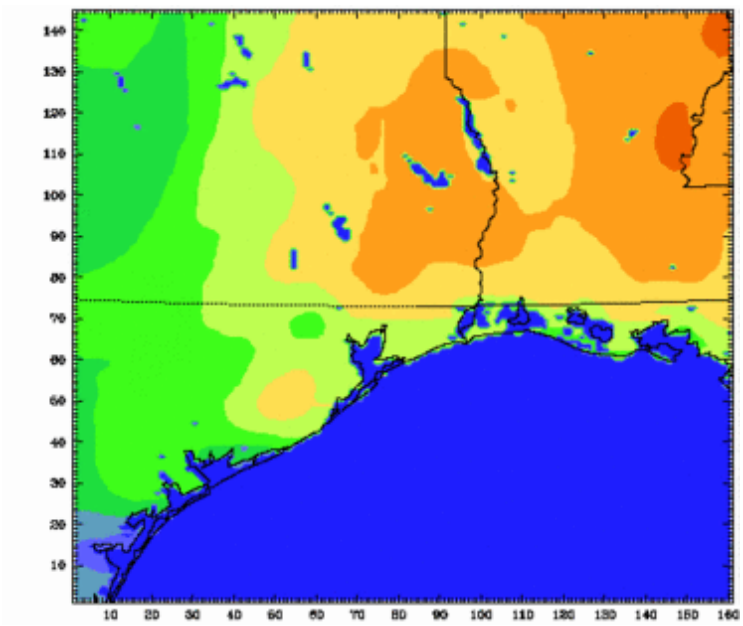
We calculated the vegetation fractional value by adding up the coverage areas of the LULC categories and multiplying a shading factor (σ) to each grid cell. For categories 2, 3, 4 and 5, we supplied the shading factor following the class descriptions in the National Land Cover Dataset (NLCD). For example, the definition of category (cat) 2 is given as:
"Class 2 – Developed Open Space: Includes areas with a mixture of some constructed materials, but mostly vegetation in the form of lawn grasses. Impervious surfaces account for less than 20 percent of total cover."

Based on this description, we have assigned σ = 0.8 for category 2. A similar approach is applied for categories 3, 4 and 5, respectively. For the other categories, we followed the values specified in the look-up table in the MM5. The σ value is listed in the last column of the Table 1. The following equation is the formula used for the estimation of the vegetation fractional value for each grid cell:

Vegetation fraction=(1/CELLAREA)* (cat1* 0 + cat2*0.8 + cat3*0.65 + cat4*0.35 + cat5*0.10 + cat6*0.07 + cat7*0.9 + cat8*0.6 + cat9*0.7 + cat10*0.7 + cat11*0.6+ cat12*0.9 + cat13*0.8 + cat14*0.95 + cat15*0.7 + cat16*0.8 + cat17*0.8+ cat18*0.7 + cat19*0.85 + cat20*0.6 + cat21*0.7 + cat22*0.7 + cat23*0.85+ cat24*0.6 + cat25*0.7 + cat26*0.2).

Here, CELLAREA represents the areas of the grid cell. For example, for 12-km resolution grid cell, CELLAREA equals 12x12 (144 km^2). Table 1 lists the LULC categories used in the UT-CSR data.

(a)



(b)

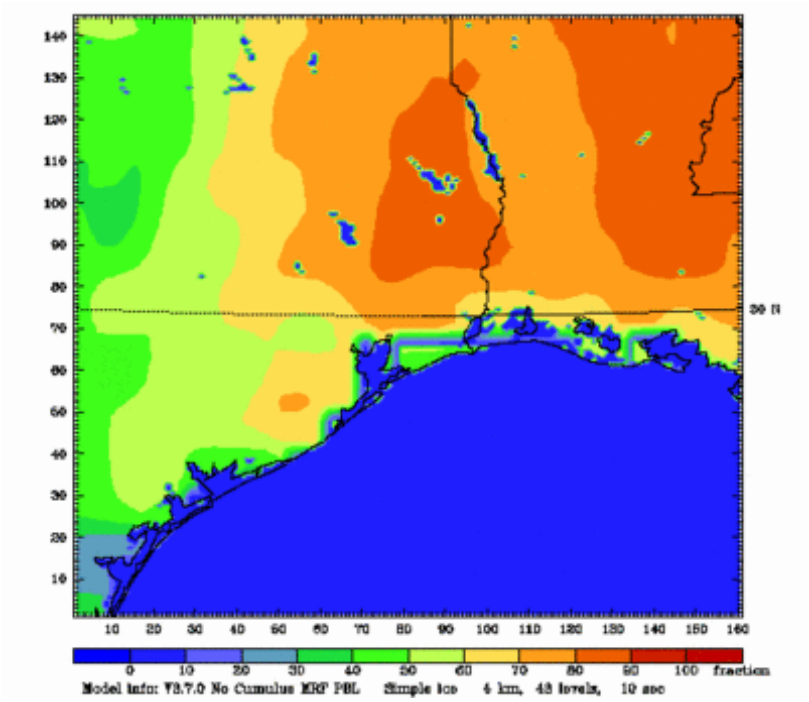


Figure 4**.** Monthly vegetation fraction value from (a) WRF and (b) MM5 meteorological modeling systems.

Table 1. LULC categories from UT-CSR data. σ is the shading factor.

Label	Acronym	Description	Mapping USGS	σ
1	OW	Open Water	16_water	0
2	DOS	Developed Open Space	7_grass	0.8
3	DLI	Developed Low Intensity	28_residential	0.65
4	DMI	Developed Medium Intensity	28_residential	0.35
5	DHI	Developed High Intensity	1_urban	0.1
6	BL	Barren Land (Rock/Sand/Clay/Unconsolidated Shore)	19_barren	0.07
7	HN	Herbaceous Natural	7_grass	0.9
8	HC	Herbaceous Cultivated	4_agriculture	0.6
9	WFR	Riparian Forested Wetland (not used)	18_wooded_wetland	0.7
10	WFS	Swamp Forested Wetland (not used)	18_wooded_wetland	0.7
11	WS	Shrub Wetland (do not have)	18_wooded wetland	0.6
12	WHE	Herbaceous Emergent Wetland	17_herbaceous wetland	0.9
13	FDC	Cold-Deciduous Forest	11_deciduous broadleaf forest	0.8
14	FEB	Broad-leafed Evergreen Forest	13_evergreen broadleaf forest	0.95
15	FEN	Needle-leafed Evergreen Forest	14_evergreen needleleaf forest	0.7
16	FM	Mixed Forest	15_mixed forest	0.8
17	CWV	Cultivated Woody Vegetation(don't have)	4_agriculture	0.8
18	WDC	Cold-Deciduous Woodland	6_cropland/woodland mosaic	0.7
19	WEB	Broad-leafed Evergreen Woodland	6_cropland/woodland mosaic	0.85
20	WEN	Needle-leafed Evergreen Woodland	6_cropland/woodland mosaic	0.6
21	WM	Mixed Woodland	6_cropland/woodland mosaic	0.7
22	SDC	Cold-Deciduous Shrub	9_mixed shrub/grass	0.6
23	SEB	Broad-leafed Evergreen Shrub	9_mixed shrub/grass	0.75
24	SEN	Needle-leafed Evergreen Shrub	9_mixed shrub/grass	0.6
25	SM	Mixed Shrub	9_mixed shrub/grass	0.7
26	SDS	Desert Scrub	8_shrubland	0.2

Figure 5(a) shows the dominant LULC categories from the original MM5 and UT-CSR LULC datasets for the12-km resolution Texas domain. Figure 5(b) is derived from the composite MM5 and UT-CSR LULC datasets as mentioned above. Figure 5(c) is from the original AVHRR satellite data utilized inside MM5. Figure 7(a) shows the dominant LULC categories from the UT-CSR and Texas Forest Service (TFS) combined LULC datasets for the4-km resolution Eastern Texas domain. Figure 7(b) is derived from the composite MM5 and UT-CSR LULC datasets as mentioned above.

In Eastern Texas and Louisiana state, the vegetation fraction values are reduced in the new dataset; near the coastal area and in western Texas, the values are increased. Inside the Houston downtown area, the vegetation fraction distribution varies in accordance to the UT-CSR LULC data. We expect that use of these improved vegetation data will improve representation of the exchange of heat and moisture simulated by the MM5 LSM.

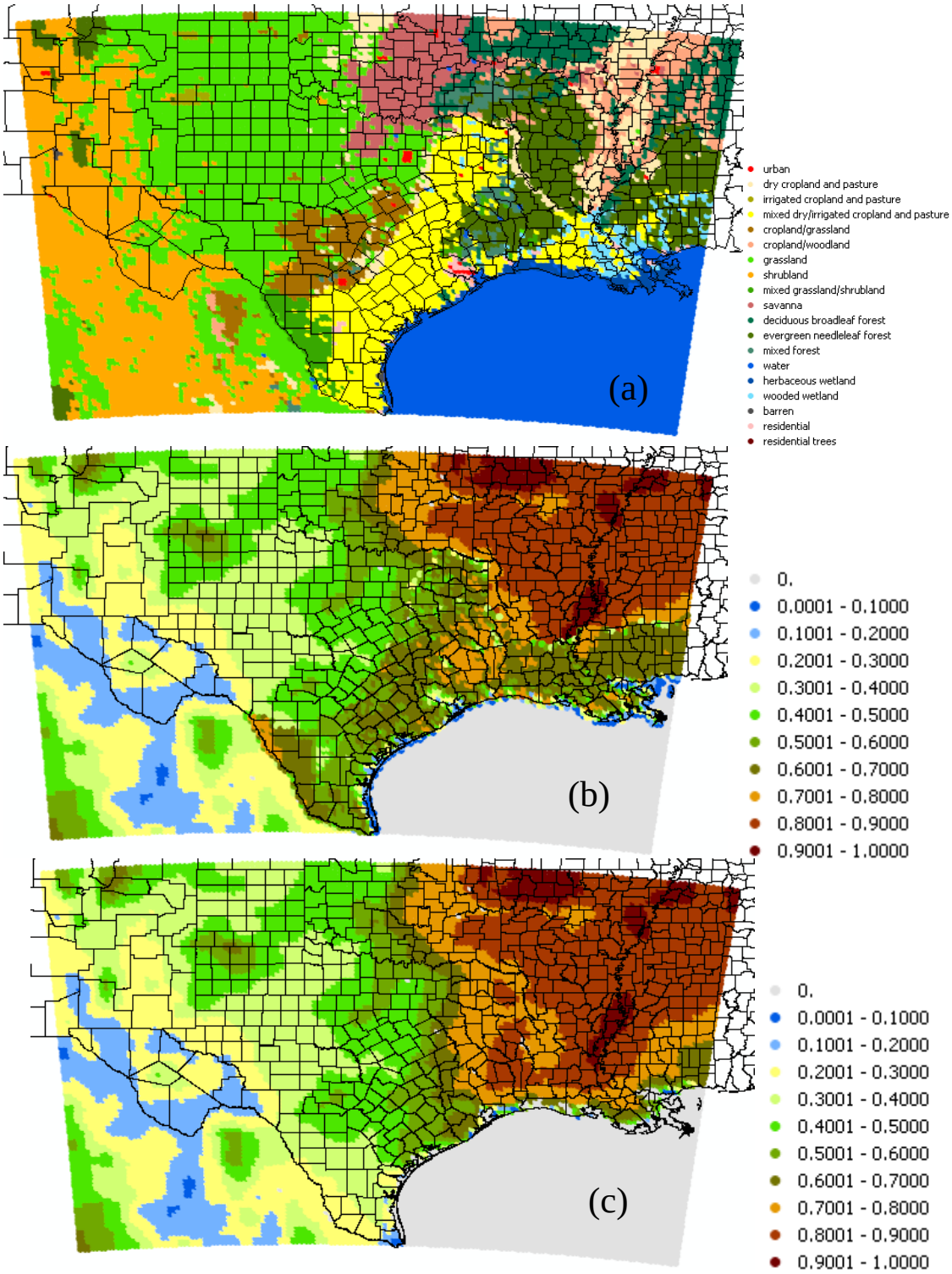


Figure 5**.** (a) Dominant LULC type composite of the original MM5 and UT-CSR LULC datasets, (b) vegetation fraction derived from the composite LULC dataset, and (c) vegetation fraction in the original MM5 for the Texas 12-km resolution domain.

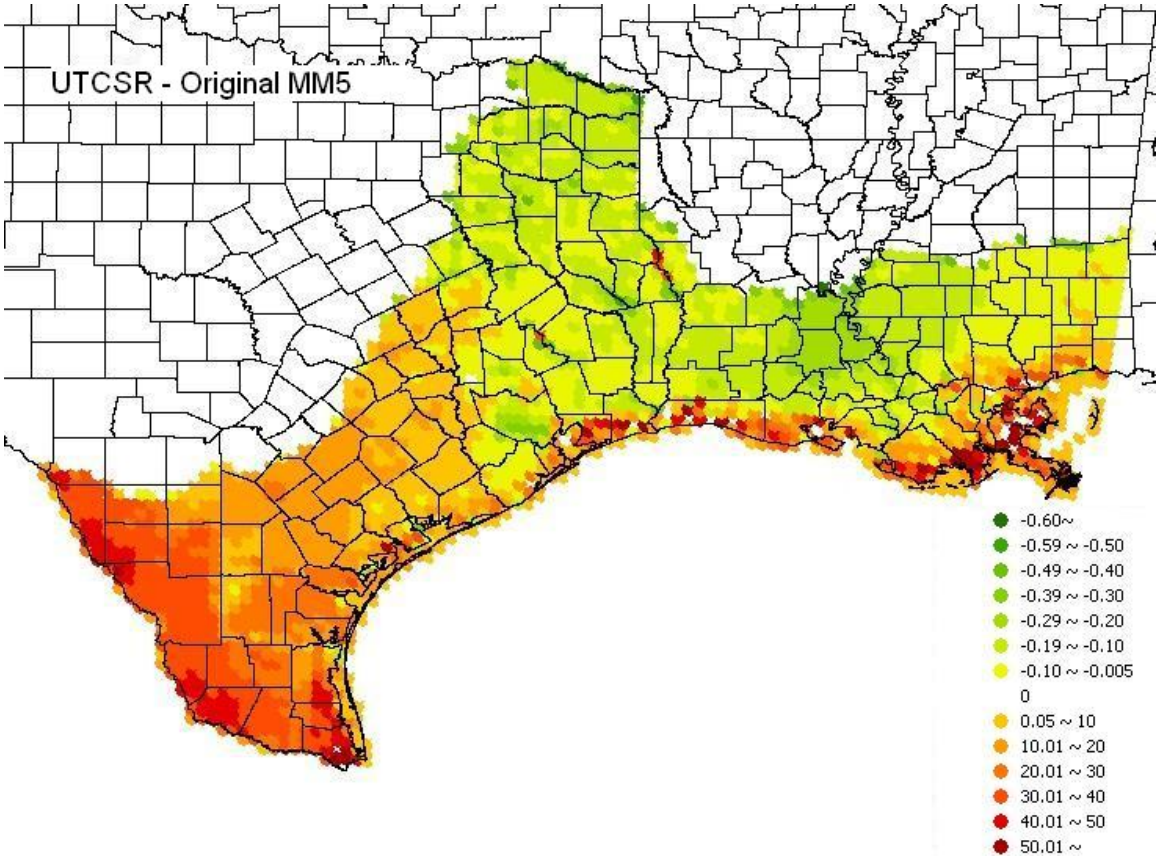


Figure 6. Difference of UTCSR-derived vegetation fraction and original MM5 vegetation fraction for the Texas 12-km resolution domain.

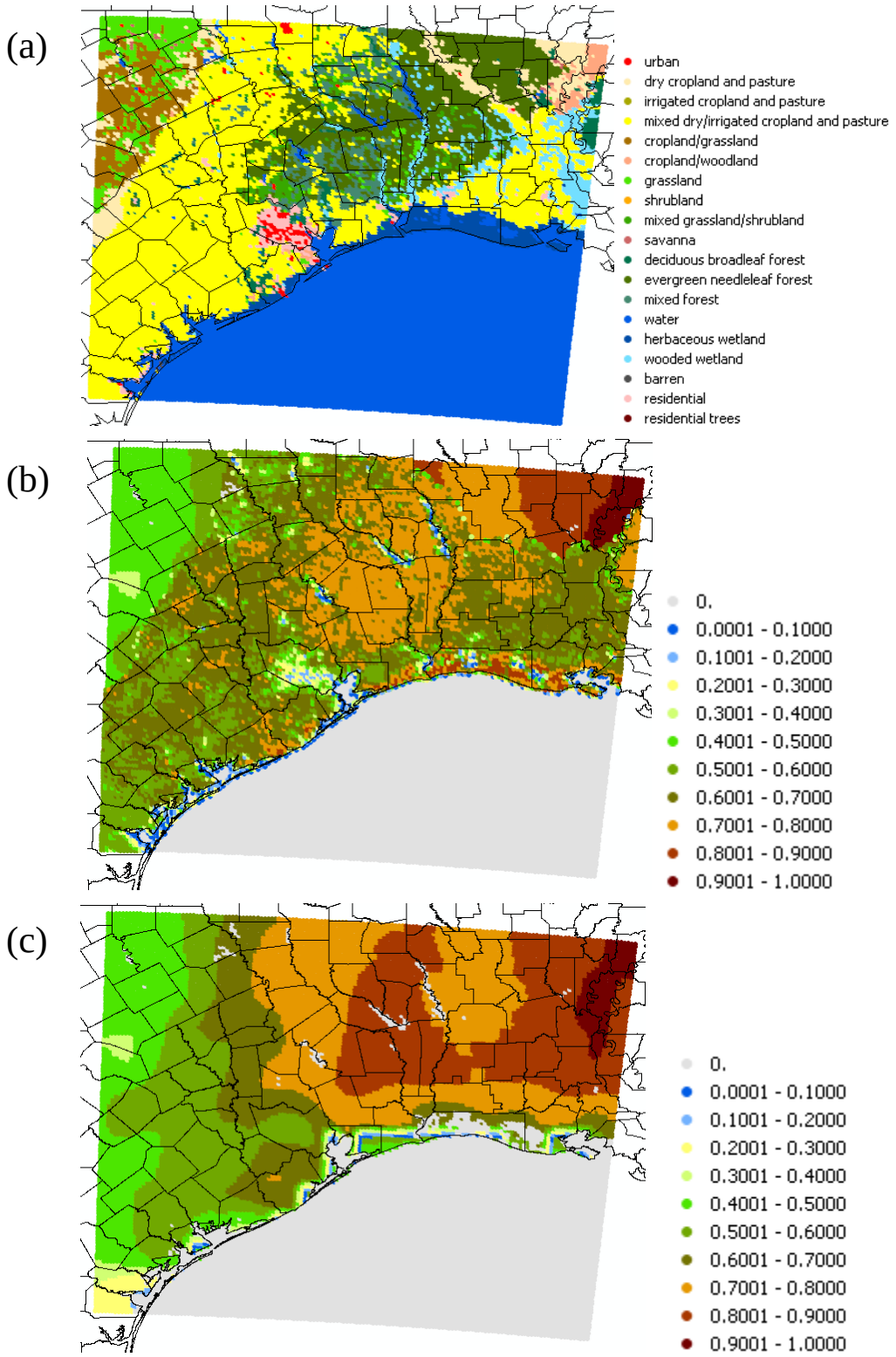


Figure 7**.** (a) Dominant LULC type composite of the original MM5, UT-CSR and TFS LULC datasets, (b) vegetation fraction derived from the composite LULC dataset, and (c) vegetation fraction in the original MM5 for the East Texas 4-km resolution domain.

3.4 Improving clay soil type in MM5

Mesoscale simulation for a limited area like for the Texas air quality modeling domain utilizes the synoptic weather reanalysis data provided by the National Center for Environmental Prediction's (NCEP's) North American Mesoscale (NAM, formerly Eta) Data Assimilation System (NDAS, formerly EDAS). At a coarse modeling grid resolution, rivers and small inland water bodies are not well resolved and their effects on meteorological conditions cannot be simulated properly. For example, the high bias of surface temperature during daytime around the river basin is present in the present 4-km resolution MM5 simulations.

This problem is serious where the river basin is represented as a dry clay soil type with the gridded soil moisture distribution in MM5, initially provided through the downscaling of the EDAS data because of the disjoint resolutions. One of the reasons why clay river basins show high ground temperature is that dry clay conducts heat flux very poorly unless the soil moisture of the clay is higher than the threshold value of 0.25%. In such cases, heat will not be transported efficiently to the subsoil layers and the sensible heat will be trapped at the topsoil layer, resulting in high surface temperature during sunny midday. We hypothesize that the artificially high temperature areas are caused by difficulties in initializing the soil moisture field for the Noah LSM with EDAS, which is at ~ 32- km resolution and cannot resolve the river basin for the nest domain simulations.

Our original intention was to find a higher resolution soil data that represent wetness in the clay near the water basins correctly. It was suggested that the Natural Resources Conservation Service (NRCS) from the United States Department of Agriculture (USDA) would be the best data source. After the investigation of data suitability for MM5 modeling, we have realized that they have similar data sources as those currently utilized inside MM5. Instead, in order to account for the river basins in the current modeling resolution (~4-km), we propose to introduce two new clay soil types, clay 1 and clay 2 to accounts for the different "dampness" of clay soil in the river and lake basins.

Figure 8 shows the dominant soil type at 4-km resolution. The areas for the clay soil type in the original MM5 LULC dataset are identified. Figure 9 shows the areas with "water" LULC type from the UT-CSR LULC data at 30-m resolution. In Figure 10, the water fraction from UT-CSR LULC data is also layered on the clay subsoil distribution. Panel (a) and (b) are for 12- and 4-km domain respectively. The clay soil type areas with a water fraction greater than 1% but less than 50% are classified as new sub-soil type "clay 1", those with a water fraction larger than 50% but less than 90% as "clay 2." For these new clay soil types, the threshold values for the heat and moisture conductivities are reduced for the MM5 simulations. Some of the parameters for different soil levels (depths)—soil temperature, soil moisture, and water at four soil levels, canopy moisture, snow height, water equivalent snow depth, surface runoff accumulation depth, underground runoff accumulation depth, ground flux, and albedo—are not readily available and fixing these values accurately is beyond the scope of the present work. We have tried to match these parameters to those used in the current MM5 model as closely as possible. Figure 11 compares the soil type distribution between (a) the updated

database with the two new types defined by water fraction obtained from UT-CSR data and (b) the original MM5 database.

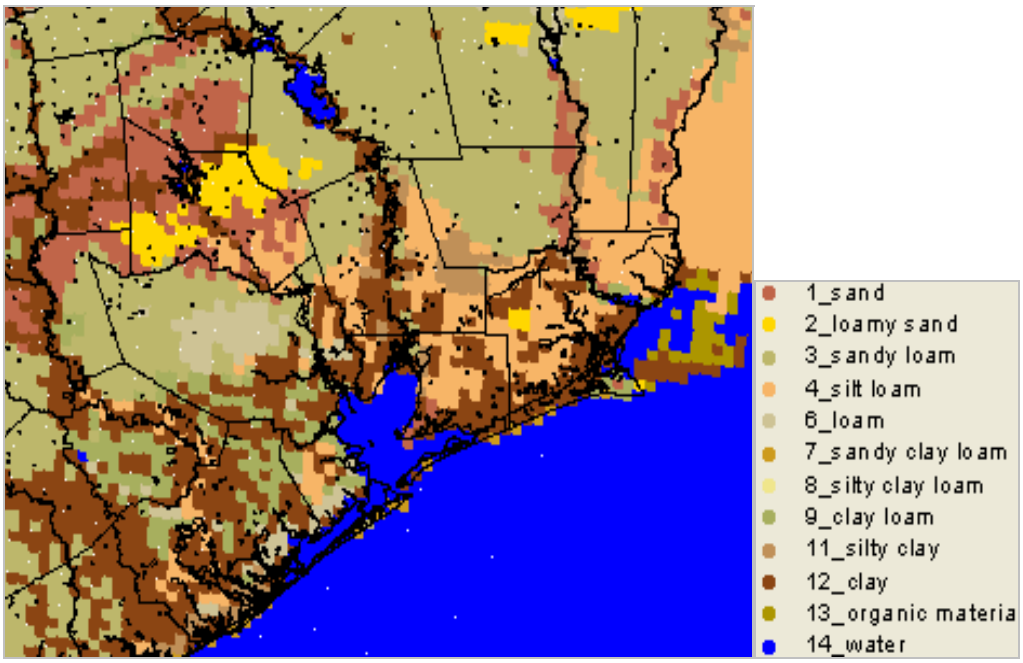


Figure 8. Dominant soil type at 4-km resolution.

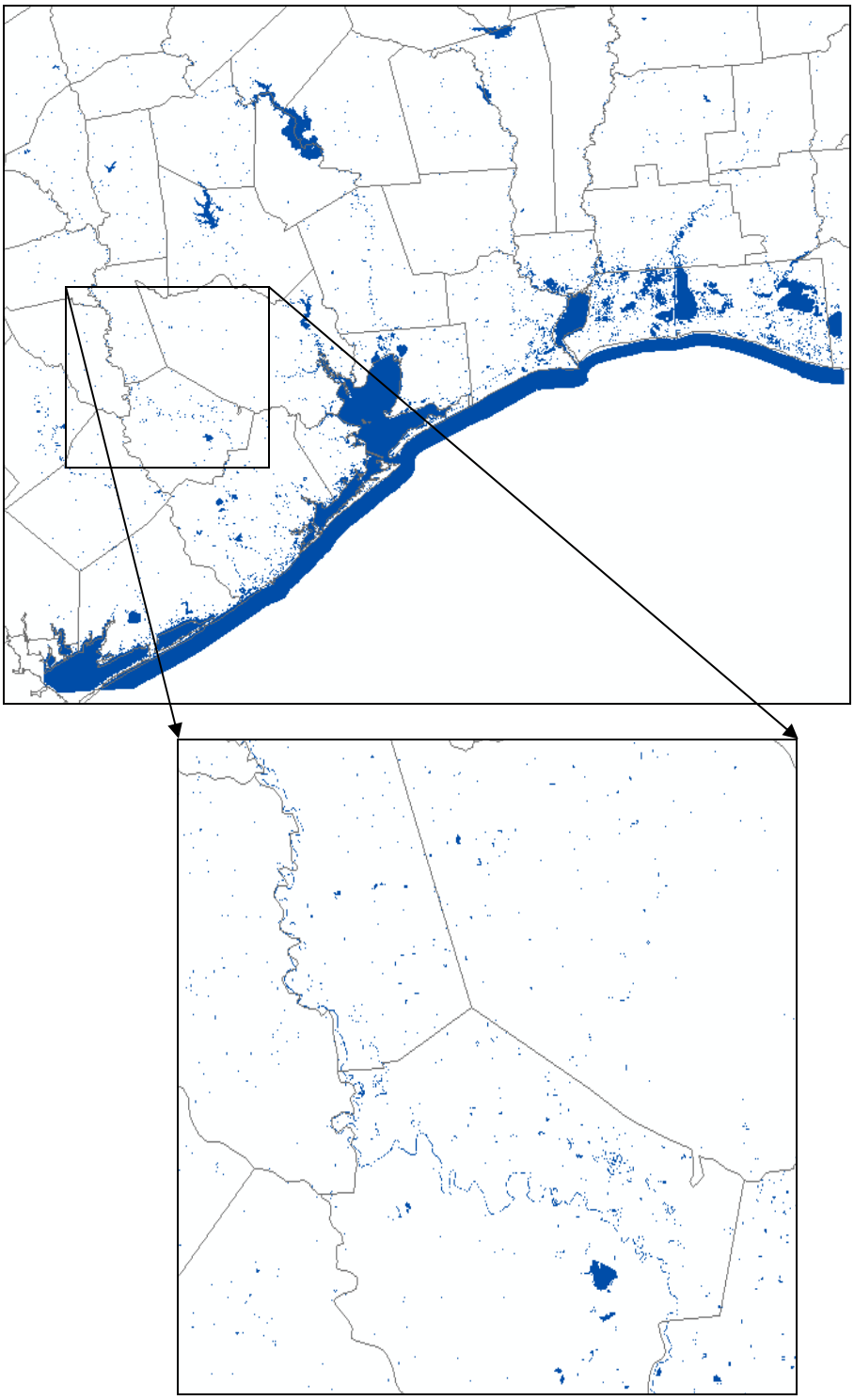


Figure 9. Blue color shows the areas with “water” LULC type from the UT-CSR LULC data at 30-m resolution.

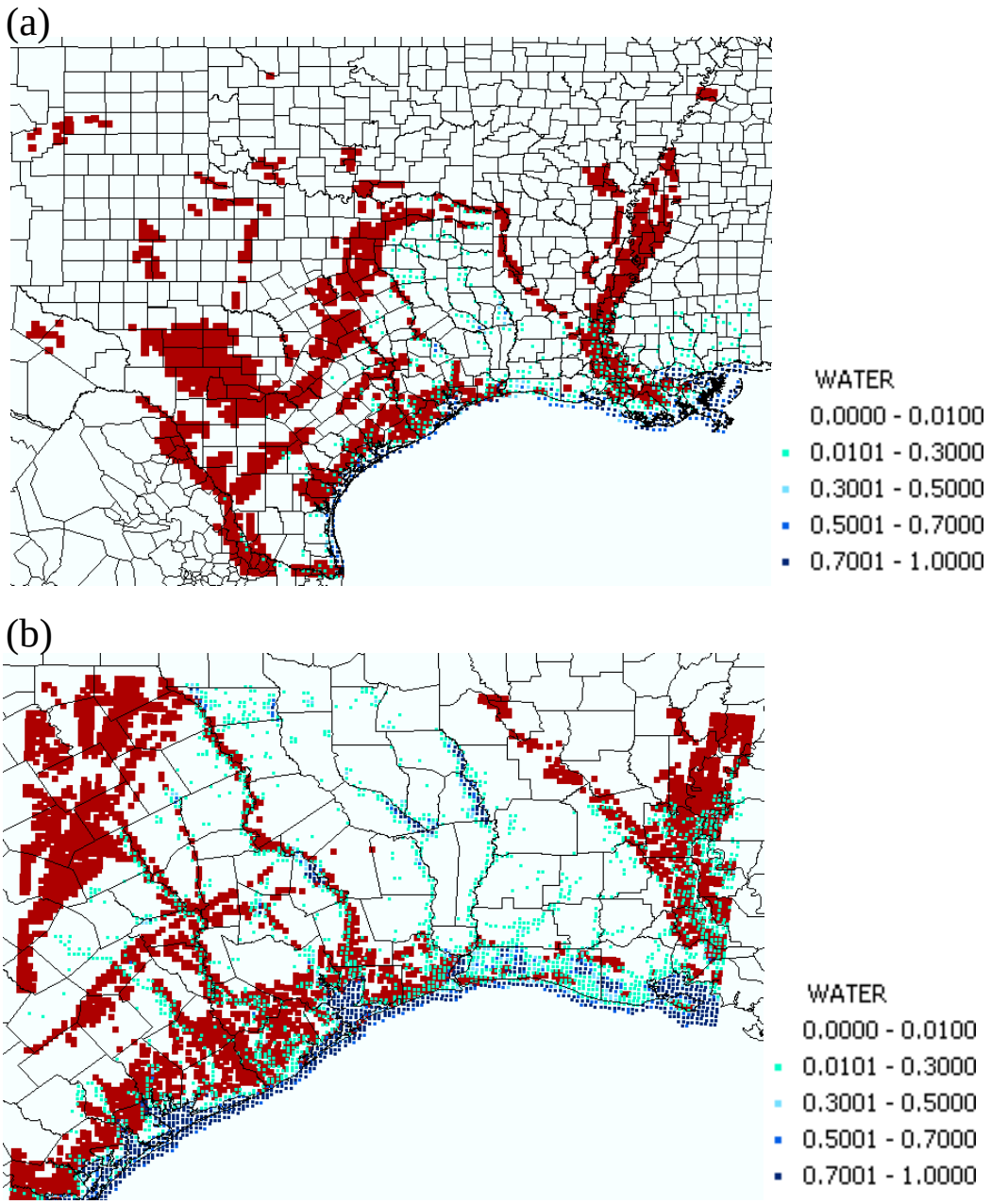


Figure 10. (a) Clay soil type (red color) layered with the fractional water land use type from UT-CSR data for the 12-km resolution domain. (b) is same as (a) but for the 4-km resolution domain.

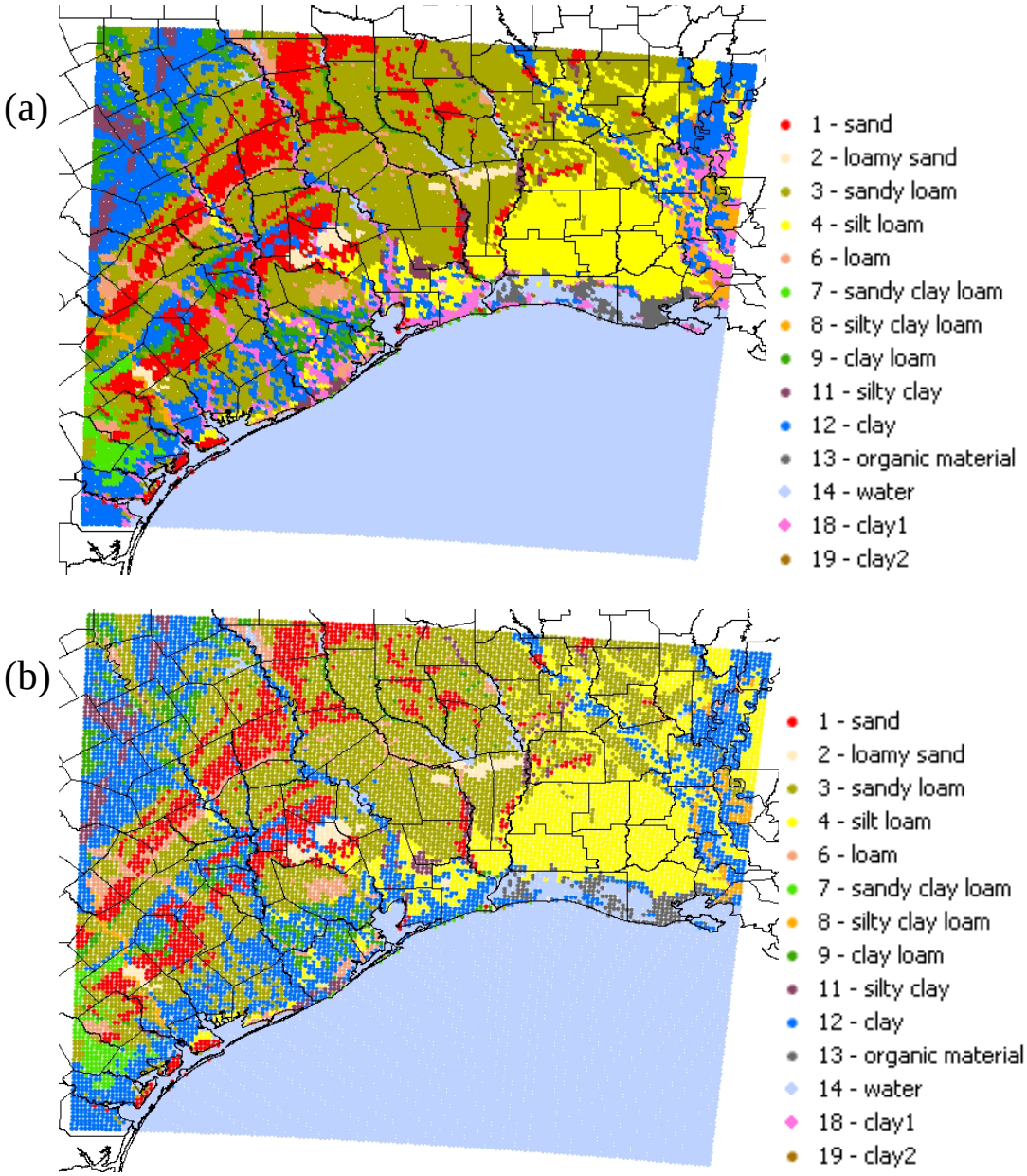


Figure 11. 4-km resolution soil type distribution from (a) the updated subsoil database with two new types defined by water fraction obtained from UT-CSR data and (b) the original MM5 subsoil database

3.5 Processing of LULC, vegetation fraction, and soil type datasets for use in TERRAIN program

The ASCII file from the previous step contains the row, column, LULC type, latitude, longitude, vegetation fraction, and water fraction, which are coordinated with the simulation domain (see the following sample).

Col	Row	LU	Lat	Lon	veg frac	water frac
63	35	52	27.6462	-99.7590	0.6242	0.0118
63	36	52	27.7529	-99.7586	0.6801	0.0002
63	37	52	27.8598	-99.7583	0.6698	0.0000
63	38	52	27.9666	-99.7579	0.6942	0.0006
63	39	52	28.0736	-99.7575	0.7032	0.0003
63	40	52	28.1806	-99.7572	0.6845	0.0003
63	41	52	28.2876	-99.7568	0.6835	0.0005
63	42	52	28.3947	-99.7564	0.6762	0.0003

The FORTRAN program called replace_lulc.f is used to replace the dominant LULC type, vegetation fraction and soil type in the TERRAIN_DOMAINx file with the new UTCSR data. After this step, the TERRAIN_DOMAINx file is ready for MM5 processing. The run script (so called “run.replace”, listed below) is where the FORTRAN program is compiled, parameters are assigned, and program is executed. The original TERRAIN file and UT-CSR data in ASCII format have to be placed in the same directory as the “replace” program. Column and row numbers are set in the run script for the user’s target domain. There are 3 “Y/N” flags to determine what parameters (such as land use/land cover, vegetation fraction, and soil type) will be updated. Note that the vegetation fraction data from the UT-CSR LULC are valid for summer months (July, August, and September) only.

```
#!/bin/csh -f
pgf90 -Mfreeform -pc 32 -byteswapio replace_lulc.f
set infile   = TERRAIN_DOMAIN2.E12                          # original TERRAIN file
set outfile  = TER_E12_UTCSR_LWmask_VEG_SOIL_FN      # output TERRAIN file name
set datafile = asc.lulc.MCIP_12_03.v2.000.rvs.txt                # LULC data file in ASCII format
set in_ii    = 135                          # y-direction (IX in mm5)
set in_jj    = 174                          # x-direction (JX in mm5)
set LU_FLAG       = 'Y'        # 'Y' update lulc, 'N' do not update
set VEG_FLAG    = 'Y'        # 'Y' update vegetation fraction, 'N' do not update
set SOIL_FLAG   = 'Y'        # 'Y' update soil type, 'N' do not update
./a.out << _EOF
$infile
$outfile
$datafile
$in_ii
$in_jj
$LU_FLAG
$VEG_FLAG
$SOIL_FLAG
EOF
```

Since the soil category in the TERRAIN file is updated by adding two newly defined soil types, a corresponding new soil table (SOILPARM.TBL) has to be used to run MM5. The updated soil table, called SOILPARM.TBL.UTCSR is delivered with this package. The original table in MM5 has to be renamed and a symbolic link needs to be made for the updated soil table:

```
mv SOILPARM.TBL SOILPARM.TBL.org
ln –s SOILPARM.TBL.UTCSR SOILPARM.TBL
```

4. MM5 Simulations with the Updated Surface Characteristics

Several MM5 simulations were performed for the TexAQS2000 episode for the 12-km and 4-km grid resolution domains with different combinations of input changes representing the improved surface characterization data.

4.1 Model configurations

The physical options and other model setup parameters used were identical to those used in Cheng and Byun (2008), which provides detailed analysis of the benefits of utilizing high resolution LULC data for the simulation of Houston-Galveston-Brazoria area. For the MM5 physical options, we generally followed the recommendations by John Nielson-Gammon's group at TAMU: Grell subgrid cloud convection, RRTM radiation scheme, UH modified MRF PBL and Noah LSM (Cheng and Byun, 2008). The only exception in the 4-km domain was that the sub-grid parameterization for clouds was turned off. Analysis nudging was used. The NCEP EDAS data at 40-km resolution (AWIP 212) was used to initialize 36-km simulations. The initial and boundary conditions of the 12-km domain for all the simulations were identical that came from the nest-down of 36-km results. Observational nudging in wind was performed in 4-km domain. The simulation cases shown in this report are listed below:

MML2: Base case, with UTCSR LULC
MML3: Use UT-CSR LULC with updated vegetation fraction
MML4: Use U-TCSR LULC with both updated vegetation fraction and new clay soil types/soil table

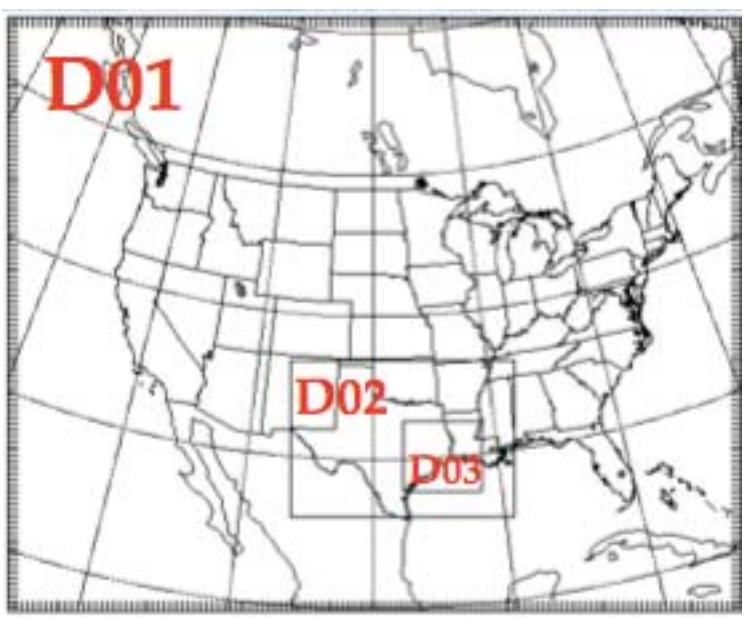


Figure 12. Domain configuration of the 3-nested domains used for MM5 simulations.

4.2 Time series analysis

Figure 13 shows regional average 1.5 m temperature, wind speed and wind direction for all CAMS sites available over the Houston area in the 12-km simulation domain. All three cases give similar predictions of average values with only slight differences shown in wind speed on 23 August and in wind direction on 30 August 2000. Behaviors of regional average 1.5 m temperature for different regions are very close but larger differences can be found for the ground temperature at individual sites (not shown here), which is directly affected by the changed vegetation fraction and soil parameters.

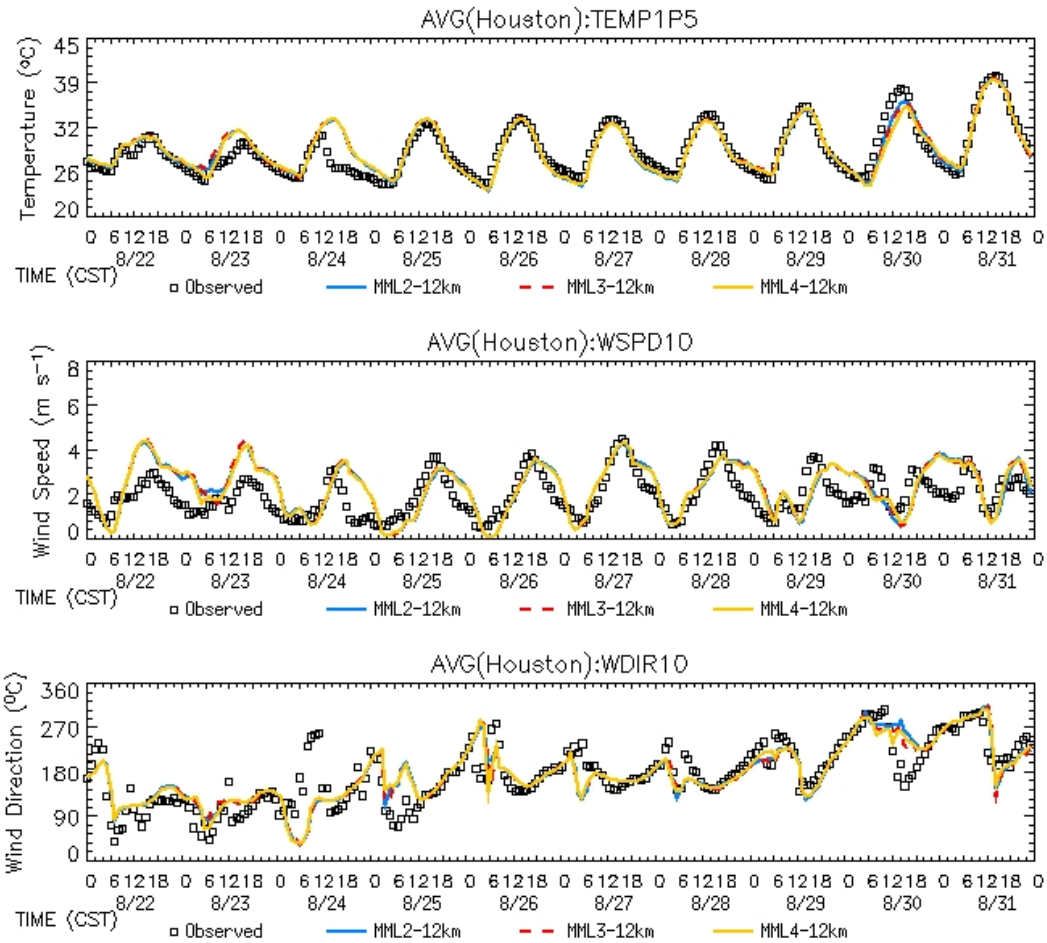


Figure 13. Time series comparison of regional average 1.5 m temperature, wind speed and wind direction at 12-km domain for the Houston area.

Figure 14 and Figure 15 are the time series regional averages for the Beaumont and Dallas area, respectively. More differences show up over the Beaumont area while very similar results are shown over the Dallas (and Houston) area among the three simulations. This is because a lot of updates in the vegetation fraction and soil types are located in the Beaumont area and near the coast of Louisiana. The direct comparisons with CAMS sites are not able to pick up the effects of changes in the simulations because their locations are away from the main areas of the changes in the input data. However, impacts of changes on MM5 simulations are more visible in the spatial plots.

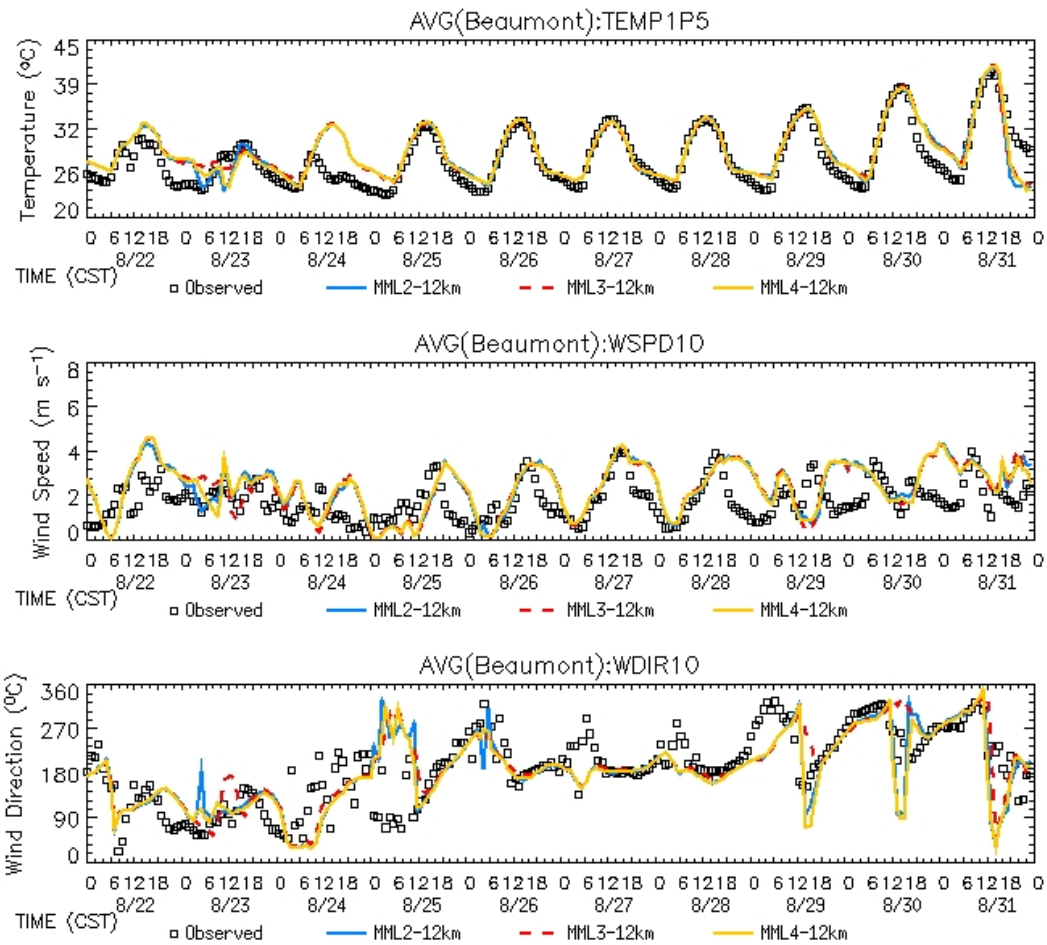


Figure 14. Same as Figure 13 but for the Beaumont area.

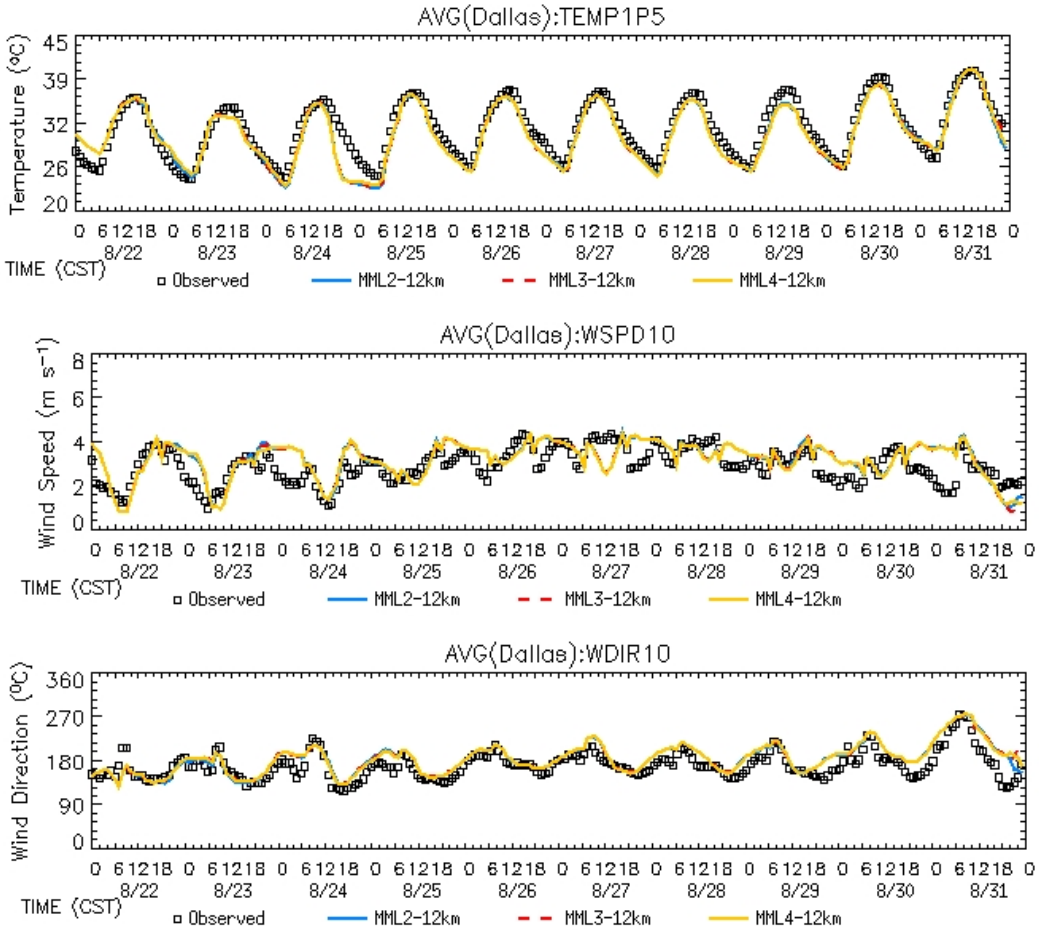


Figure 15. Same as Figure 13 but for the Dallas area.

Regional average plots in 4-km simulations (not shown) reveal that predictions of the three cases are similar to what have been shown for the 12-km resolution domain. Figure 16, Figure 17 and Figure 18 show 1.5 m temperature, wind speed and wind direction, respectively, at a few individual CAMS sites located in the Beaumont and Houston areas. The simulation with the updated vegetation fraction and soil types seems to produce slightly better results for the 1.5 m temperature on 23 August and the trend of temperature decrease on 31 August. In terms of wind speed and wind direction, slight improvement can be seen on 23 and 30 August. More detailed tests over other meteorological conditions would be necessary to verify the direct benefits of this change.

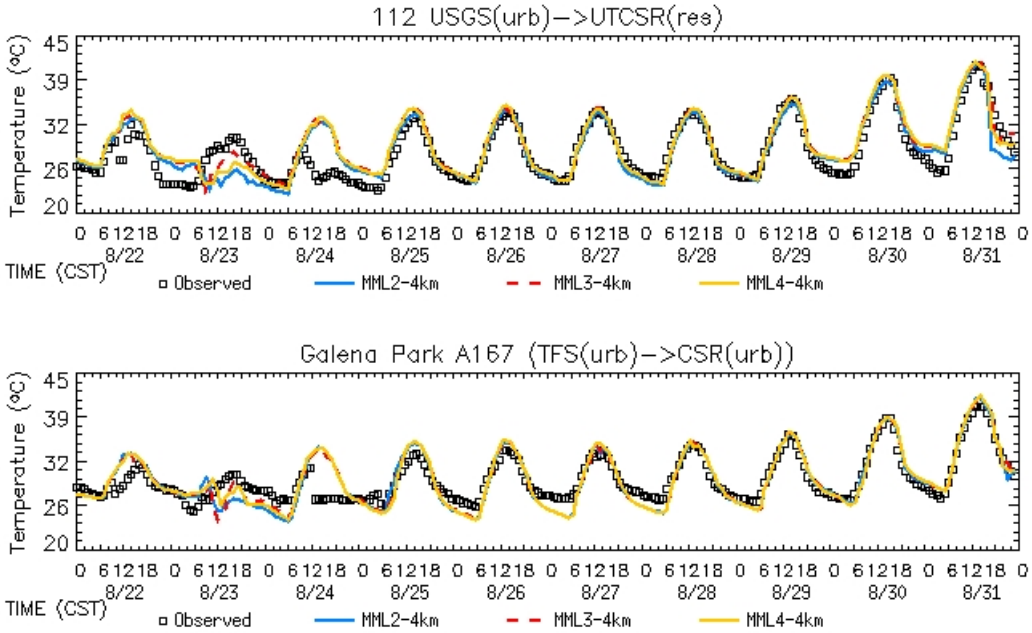


Figure 16. Time series comparison of 1.5 m temperature at 4-km domain at CAM site 112 (Beaumont) and 167 (Houston).

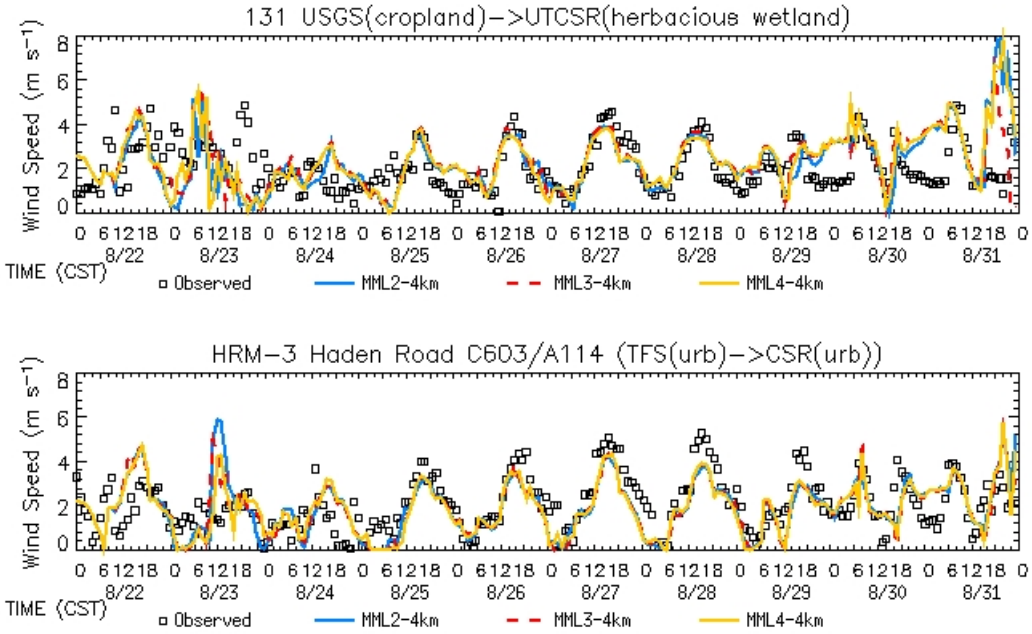


Figure 17. Same as Figure 16 but for wind speed at CAMS site 131 (Beaumont) and 603 (Houston).

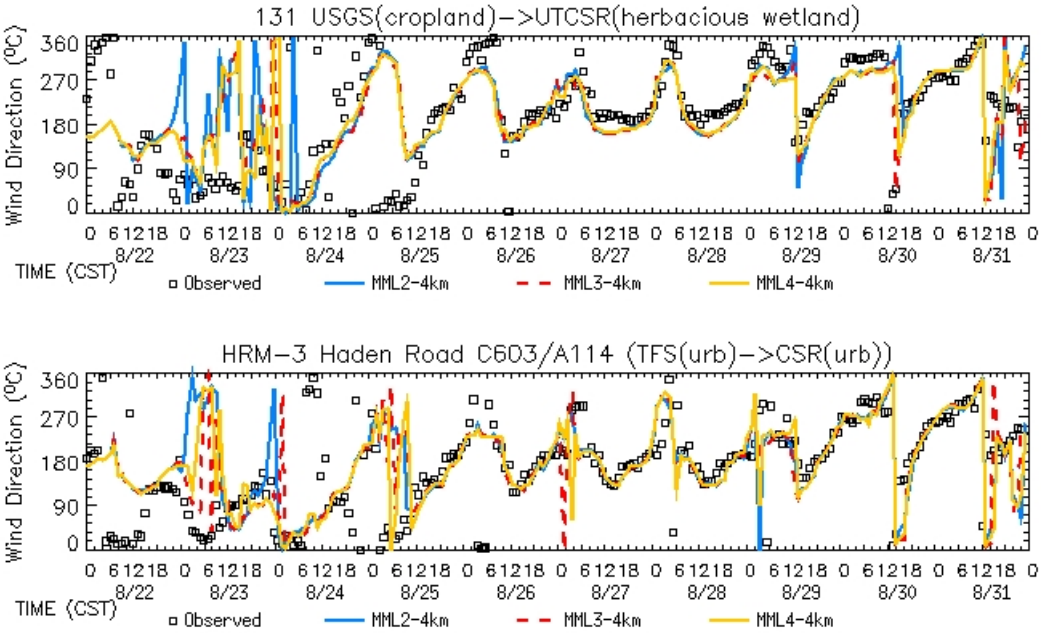


Figure 18**.** Same as Figure 16 but for wind direction at CAMS site 131 (Beaumont) and 603 (Houston).

4.3 Spatial analysis

Since the differences in the sensitivity cases are not readily distinguishable from the time series plots for the CAMs sites, we focus on the spatial-difference plots in this analysis. The corresponding series of figures of individual simulation cases are given in the Appendix. Figure 19 shows the spatial-difference plot among the three cases of ground temperature at 18 UTC 30 August 2000 for the 12-km and 4-km resolution domain. In general, larger differences in the afternoon and smaller differences at night can be seen. Using satellite-derived vegetation fraction data (MML3), the 12-km domain result shows higher ground temperature in northeastern Texas and the western and center part of Louisiana. That is due to the lower vegetation fraction value being used in simulation MML3. In southern Texas near Mexico and coastal areas of Louisiana, on the other hand, lower ground temperature is predicted because of increase of vegetation fraction used in the MML3 case (Figure 19b). For the 4-km resolution domain, because of new vegetation fraction data that represent the reality more accurately, the high ground temperature line present to the west of Galveston Bay is gone. Furthermore, the ground temperature along the wetland areas in the Louisiana coast is decreased (Figure 19e). The impact of using updated soil type and soil table in the MM5 simulation can be seen in Figure 19(a) and Figure 19(d). For the 4-km domain, the lower ground temperature is simulated with the new soil types. Other places such as Louisiana and west of Houston have temperature decreases as well.

Planetary Boundary Layer (PBL) height plots also show direct influences of using improved vegetation fraction and soil datasets. Reflecting the differences in the ground temperature, higher PBL height is simulated in northeastern Texas while lower PBL height is simulated in southern Texas (Figure 20b). PBL heights near Galveston and the eastern Texas and Louisiana coast also show large differences.

In terms of horizontal wind, we can only see slight differences among the three sensitivity cases. During the evening hours of 30 August, appreciable differences in winds are shown along the eastern Texas coast between the MML3 and MML2 cases (Figure 21 right panel). This is because the temperature gradient simulated by the different vegetation fraction and soil type data affects the developmental pattern of local flows like the sea breeze. Like other parameters, more differences can be seen with the MML3 case (with new vegetation fraction data) than MML4 (with new clay soil types). At 2 UTC 25 August, wind field difference is shown in both sensitivity cases in central Texas away from the areas where the surface characteristic data are changed. Figure 22 shows the thunderstorm event that occurred on 23 August. The changes in vegetation fraction and soil type data seem to affect the timing of thunderstorm development, probably due to the change in the way flow convergence develops. Thus, the strong downdrafts associated with the thunderstorm are simulated differently. It further results in changes in the location of rainfall. Although additional evaluation with high resolution in situ and remote sensing measurements are needed, it seems that use of improved surface data affects simulation of mesoscale phenomena such as sea breezes and thunderstorm formations.

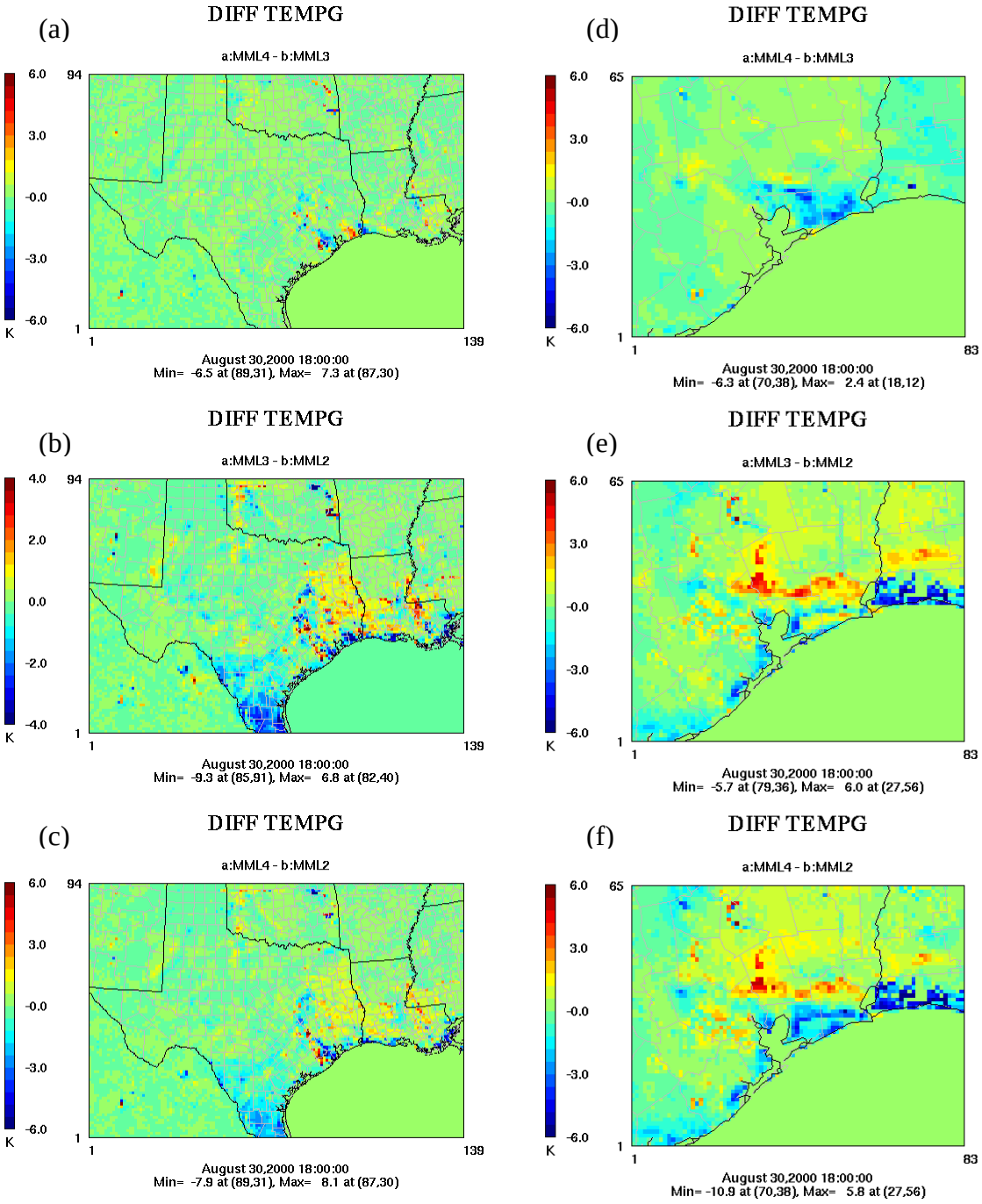


Figure 19. Difference plots of ground temperature for 12-km (left panel) and 4-km (right panel) resolution domain at 18 UTC 30 August 2000.

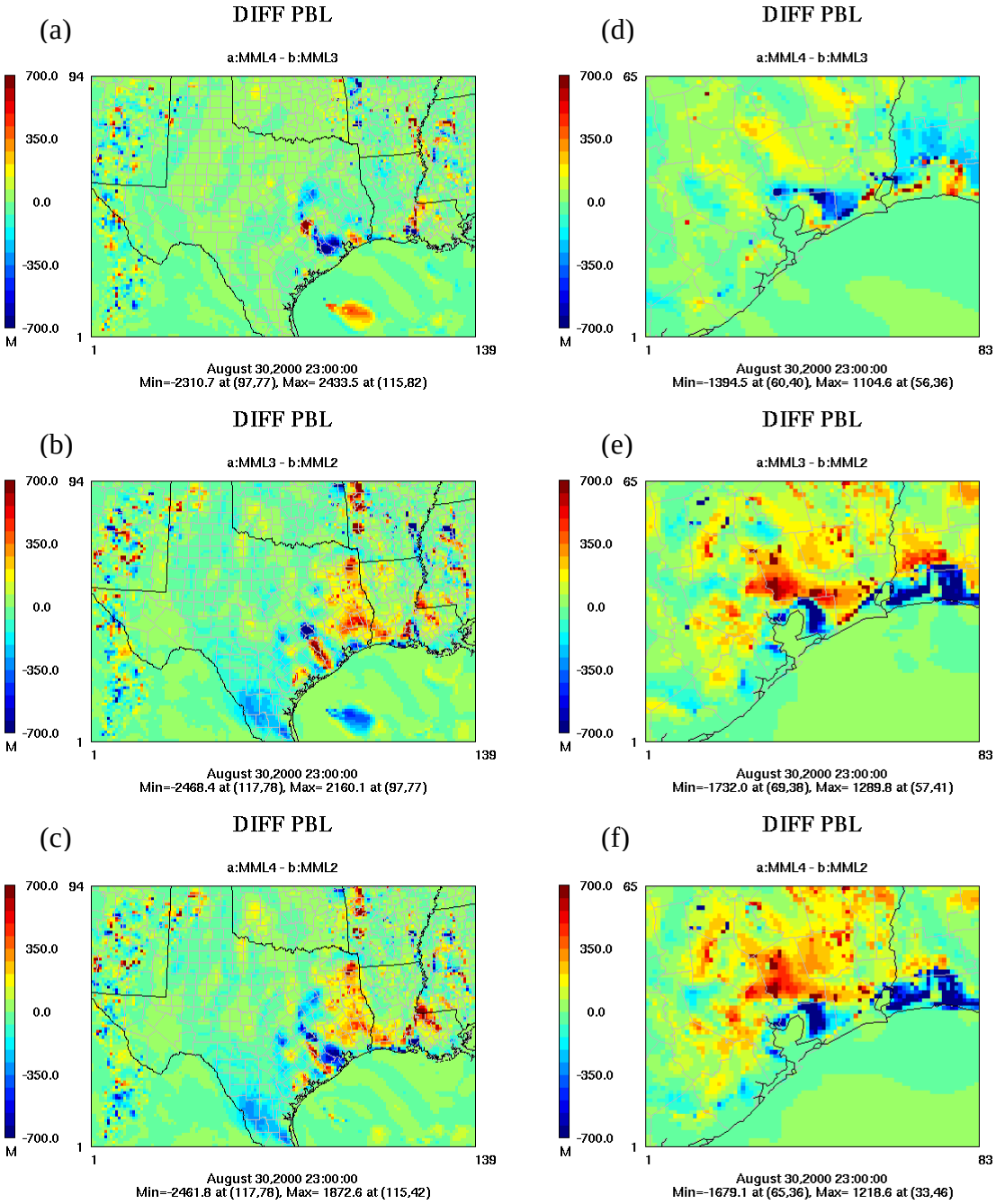


Figure 20. Difference plots of PBL height for 12-km (left panel) and 4-km (right panel) resolution domain at 23 UTC 30 August 2000.

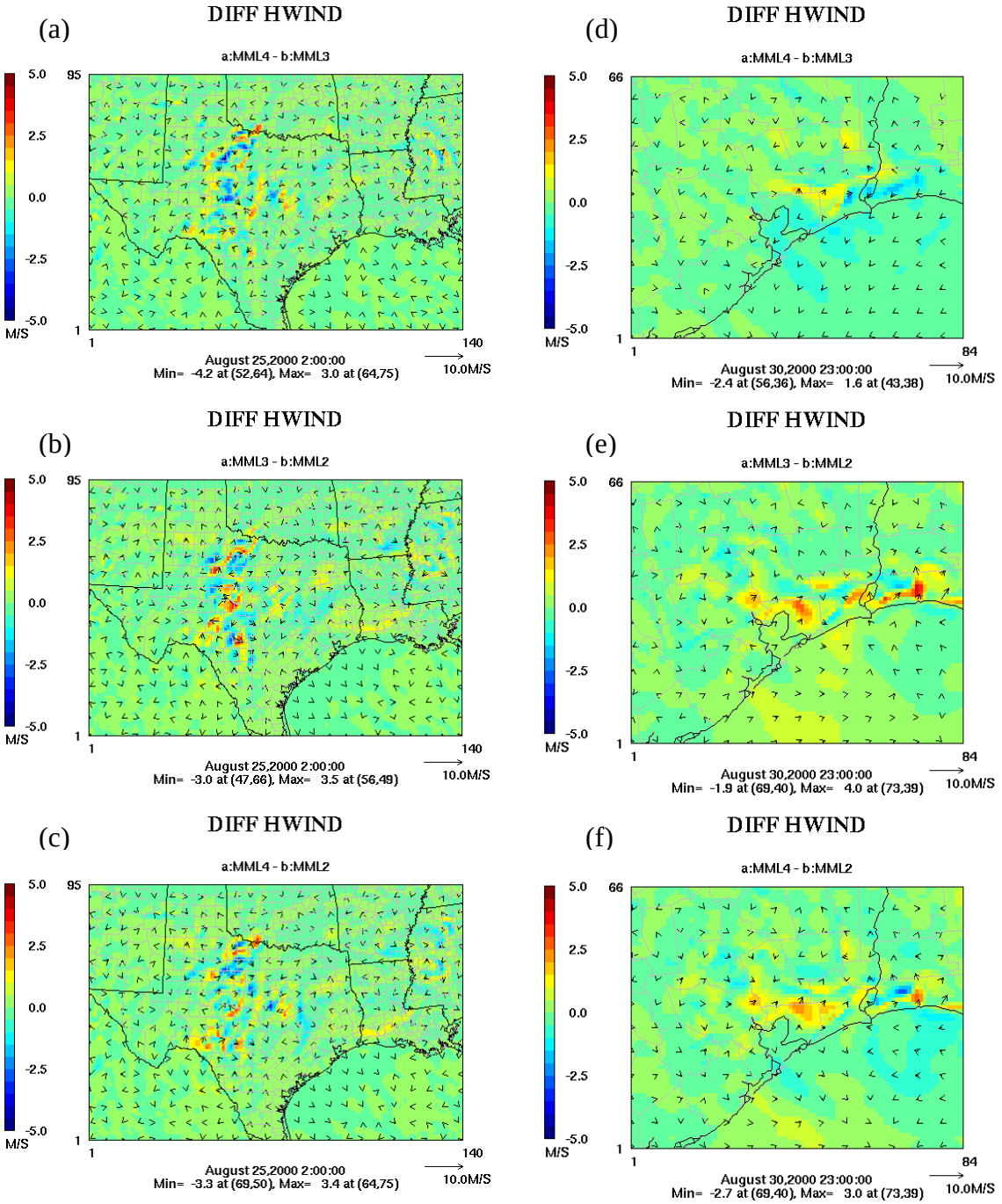


Figure 21. Difference plots of horizontal wind at 1st layer for 12-km at 2 UTC 25 August 2000 (left panel) and 4-km (right panel) resolution domain at 23 UTC 30 August 2000.

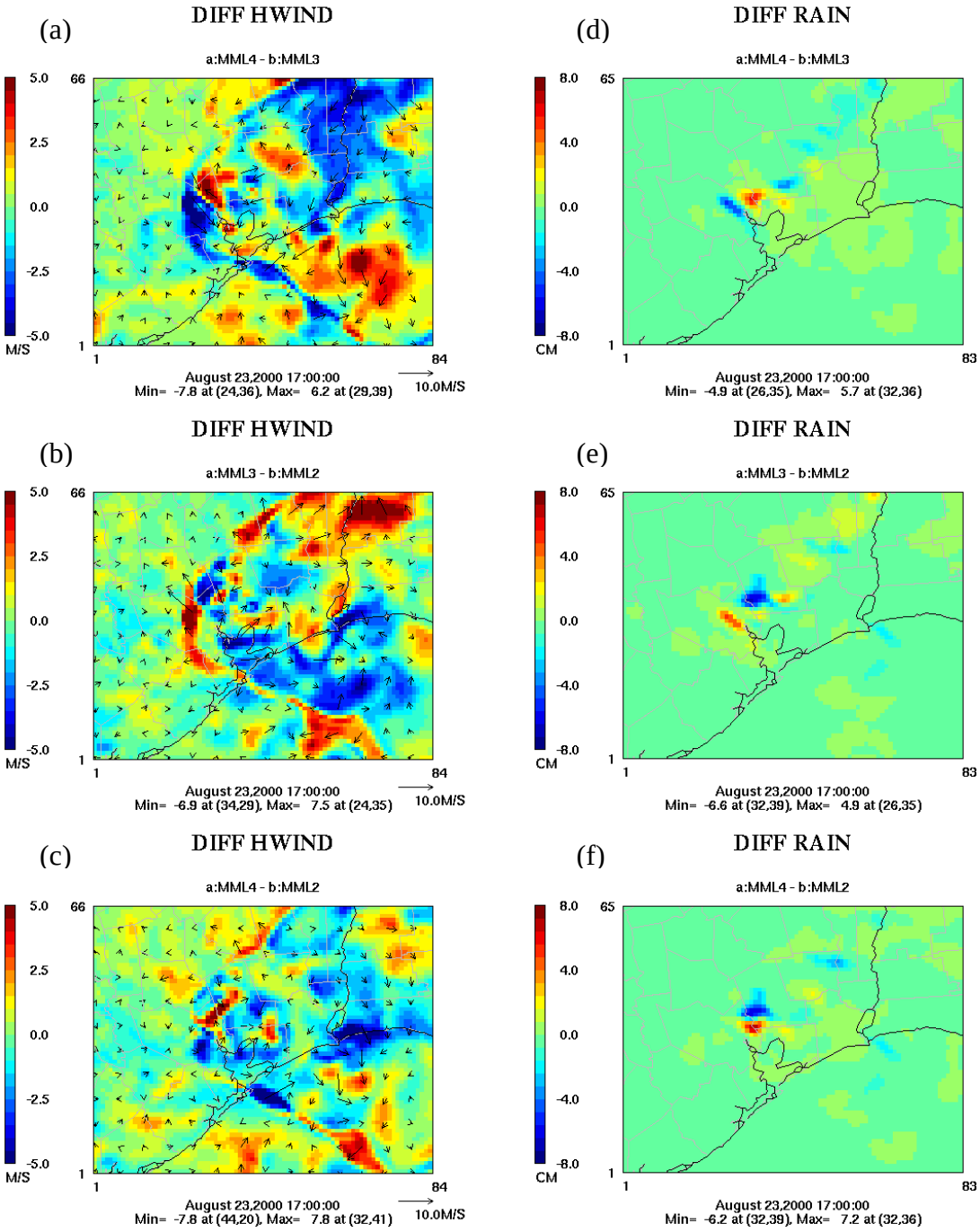


Figure 22. Difference plots of horizontal wind at 1st layer (left panel) and precipitation for 4-km resolution domain at 17 UTC 23 August 200

5. Processing of Sea Surface Temperature and MM5 Simulations

In the final report of the "Improved Modeling Inputs: Land Use and Sea-Surface Temperature" (Byun et al., 2007), we reported the procedure to interpolate the sea surface temperature (SST) measurement from geostationary satellites. Since then, we have made two improvements in the processing of SST for the MM5 simulations. One is expansion of the time span to fill in the missing satellite observation and the other is improvement of the spatial interpolation method to eliminate idiosyncratic patterns in the interpolated SST fields due to cells containing empty satellite data points.

5.1 Handling missing data

We have improved routines to utilize hourly GOES SST measurements. The most important issue in dealing with hourly SST data is how to handle the missing data, especially inside Galveston Bay where diurnal temperature variation is quite different from that of the outer sea. A simple horizontal interpolation method can lose local characteristics of diurnal variation; instead, we have set up a 3-step strategy to retrieve GOES SST and fill missing points.

(1) Get GOES SST at $t = t_0$
(2) Get GOES SST averaged values between $t = (t_0\text{-}2, t_0+2)$
(3) Get SST from monthly-averaged diurnal variation

First, we preferably use raw GOES SST values where measurements are available. Second, for missing points, we try to find available SSTs within 2 previous or next time steps. Available SSTs are averaged temporally, and are used to fill in the missing points. It should be noted that averaging is done only temporally, so averaged SSTs still have horizontal distribution in each grid cell. Choosing ±2 hours for averaging is just arbitrary, assuming SST doesn't change significantly during the 2-hour period. Third, if step 1 and 2 fail, we use 'monthly-averaged diurnal variation', which is described in the next section. Figure 23 is an example of SST retrieval using this 3-step method. Figure 23(a) shows GOES SST measurements on September 1, 2006 17:00, showing a lot of missing points. Averaged SSTs between 15:00 ~ 19:00 are added in Figure 23(b). Steps are completed by using 'monthly-averaged diurnal variation' in Figure 23(c), filling all missing points in the water body cells.

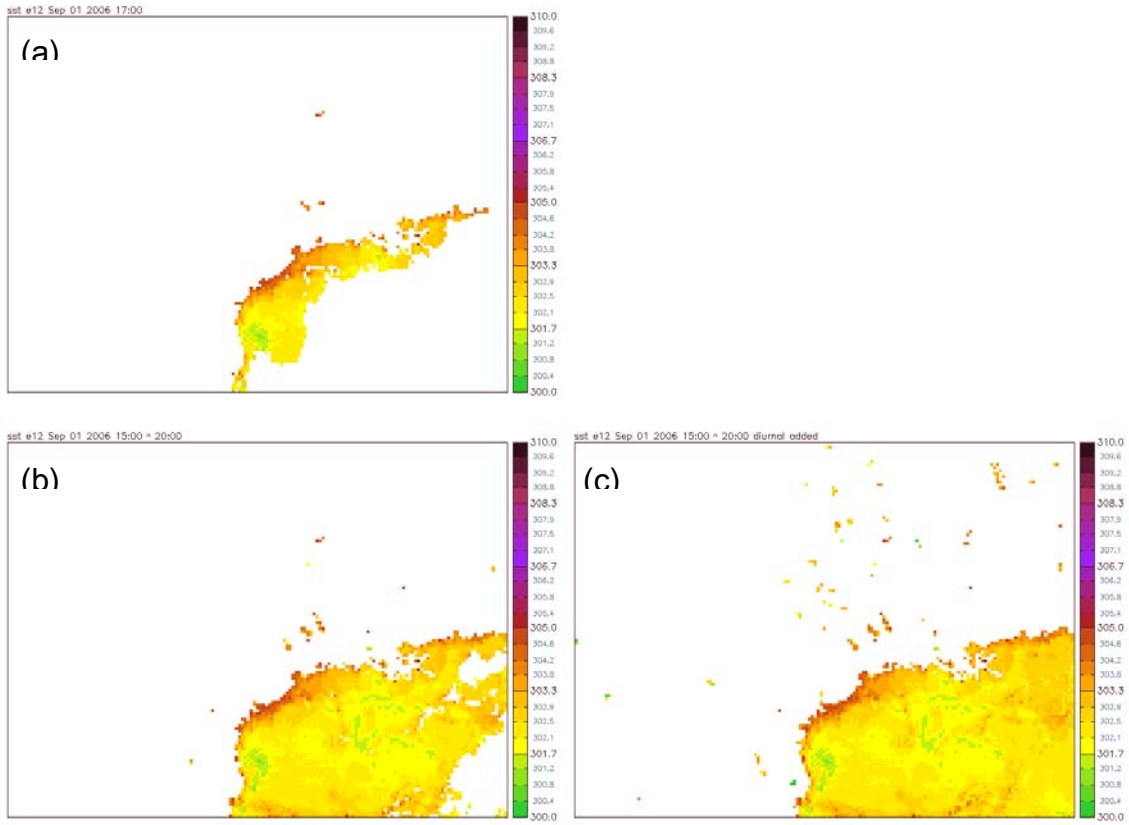


Figure 23**.** Steps of GOES SST retrieval. (a) Raw GOES SST measurement at Sep. 1, 2006 17:00, (b) SST averaged between 15:00 ~ 19:00 is added. (c) Monthly averaged diurnal SST variation is added.

5.2 Pre-interpolation of SST

Another technical issue we have encountered during the SST preparation is shown in Figure 24. In a fine resolution grid like 4km, we have seen some artificial patterns occurring in the SST retrieval field as shown in Figure 24(c), which is related to the resolutions of SST measurements and the grid resolution of the domain used for modeling. Since spatial resolution of GOES SST is around 6km (Figure 24(a)), there are grid cells without measurements. Those points are always treated as missing points, causing a weird texture pattern. To avoid this, we have increased SST measurement resolution by horizontal interpolation (doubling resolution in the 4km domain) as shown in Figure 24(b). Finally, gridded SSTs in Figure 24(d) show continuous distribution without patterned missing points. This pre-interpolation method works fine down to 1km resolution.

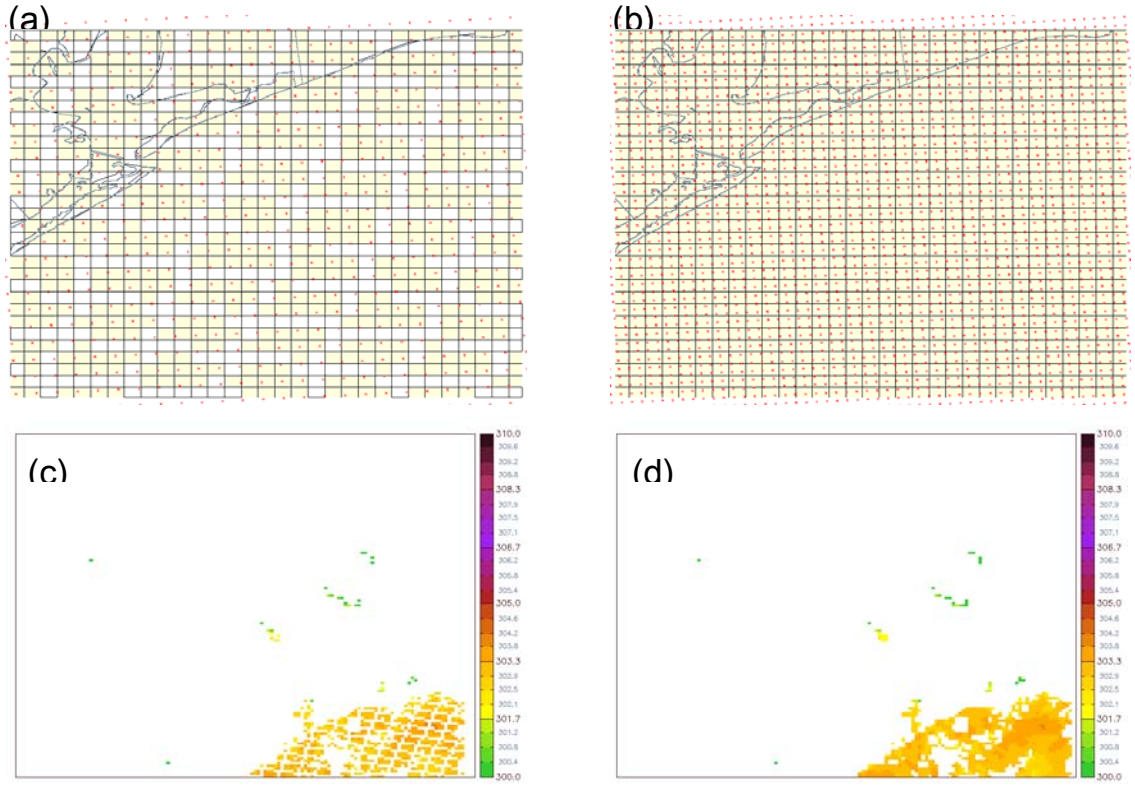


Figure 24**.** Example of SST pre-interpolation for fine resolution grid cells. Left panels show (a) 4km CMAQ domain grids and locations of GOES SST measurements, and (c) gridded SST. Shaded cells in (a) indicates grid cells with available SST measurements. Right panels show (b) doubled SST points by interpolation, and (d) corresponding SST retrieval in grid cells.

6. Conclusive Remarks

In this project, UH has improved key model inputs for MM5 characterizing the land surface conditions, such as land-sea masks, vegetation fractions, and clay subsoil types, to be consistent with the highly accurate LANDSAT-based land use land cover data for the Eastern Texas region. We also improved the method of handling missing data in satellite observations to generate more consistent sea-surface temperature inputs for the MM5 simulations.

Several MM5 simulations were performed for the TexAQS2000 episode for the 12-km and 4-km grid resolution domains with different combinations of input changes representing the improved surface characterization data. The sensitivity simulations show only small differences in the regional average time series plots of temperature and wind at the CAMS locations, but larger differences exist for the ground temperature at individual sites where vegetation fraction and soil classifications are changed. Spatial

analysis of the sensitivity simulations reveals distinguishable effects of the input changes such as the convergent flow patterns and timing of sea breeze and thunderstorm events as well as distribution and magnitudes of daily maximum temperatures. Although additional evaluation with high resolution in situ and remote sensing measurements are needed, it seems that use of improved surface data affects simulation of mesoscale phenomena that are important in characterizing air pollutant source-receptor relations.

One of the future projects would be incorporating the additional high resolution LULC data currently being assembled for the north-eastern Texas region by TAMU. Also, as NCAR has stopped further development of the MM5 system, it would be necessary to prepare for the transition to the Weather Research and Forecasting (WRF) model as the meteorological driver for air quality modeling. The WRF modeling system is a next-generation mesoscale meteorological model that has started to be used widely in operational and research weather simulations. We need to develop a localized WRF modeling system that inherits and further improves the current MM5 versions used by TCEQ and UH for modeling the Eastern Texas region. A future task extending the present work could be implementation of the LULC dataset into the WRF meteorological modeling system. For the last several years, University of Houston (UH) has been comparing the performance of WRF and MM5 simulation results. The main objectives of the new project would be to improve the land-surface and vegetation databases in WRF, and to perform comparative evaluation between the WRF and MM5 results.

7. References

Byun, Daewon, S. Kim, F.-Y. Cheng, H.-C. Kim, and F. Ngan, 2007: Improved Modeling Inputs: Land Use and Sea-Surface Temperature, Final Report, Texas Commission on Environmental Quality, August, 2007, 33 pp

Cheng, F.-Y., and D. Byun, 2008: Application of high resolution land use and land cover data for atmospheric modeling in the Houston-Galveston metropolitan area; Part I, Meteorological simulation results, Atmospheric Environment (in print)

Cheng, F.-Y., S. Kim and D. Byun, 2008: Application of high resolution land use and land cover data for atmospheric modeling in the Houston-Galveston metropolitan area; Part II, Air quality simulation results, Atmos. Env. 42, Issue 20, 4853-4869.

Coats, C. J., M.T. Odman, and S. Datta, 2007: Effects of Sub-Grid Scale Terrain upon CAMx Air Quality Simulation, 6th Annual CMAS Conference, Chapel Hill, NC, October 1-3, 2007

Gutman, G. and A. Ignatov, 1998: Derivation of green vegetation fraction from NOAA/AVHRR for use in weather prediction models, Int. J. Rem. Sens., 19: 1533-1543.

Wells, G., 2006: The New Eastern Texas Land Use Land Cover Classification Project, HARC Project Contact number: H-46-T28-2004-T2, UT-Austin Center for Space Research, Austin, Texas