1.1 Impact of High Resolution Land-Use Data in Meteorology and Air Quality Modeling Systems

Limei Ran[1], Jonathan Pleim[2], and Robert Gilliam[2]

[1]Institute for the Environment, University of North Carolina, Chapel Hill, NC, USA

[2]United States Environmental Protection Agency, Atmospheric Model Development Branch, Atmospheric Modeling and Analysis Division USEPA/ORD/NERL Research Triangle Park, NC 27711

Abstract Accurate land use information is important in meteorology for land surface exchanges, in emission modeling for emission spatial allocation, and in air quality modeling for chemical surface fluxes. Currently, meteorology, emission, and air quality models often use outdated USGS Global Land Cover Characterization (GLCC) 30-s (around 1 km) resolution land cover data. With the release of the 2001 National Land Cover Data (NLCD) products at 30 m cell resolution for the United States and 2001 NASA Moderate Resolution Imaging Spectroradiometer (MODIS) land cover data at 1 km cell resolution for the globe, meteorology and air quality modelers want to use these more current and accurate land cover data sets. In the Spatial Allocator, C++ programs were developed with the Geospatial Data Abstraction Library (GDAL) to compute modeling domain gridded land cover information based on input image data of the 2001 NLCD and MODIS land cover data. The programs output gridded fractional coverage of each land category for use in the Weather Research and Forecast (WRF) and Community Multiscale Air Quality (CMAQ) models. The land use data are used to specify vegetation and surface related parameters that are needed in land surface models (LSM) and dry deposition models. We have incorporated the gridded 2001 NLCD and MODIS land cover data in the WRF and CMAQ modeling for the CONUS and east US 12 km resolution domains. Preliminary WRF results show slight improvement and CMAQ runs show largest difference in the bi-directional NH_3 surface flux. We believe that these new land cover data should have more effects on both meteorological and air quality model simulations for higher resolution modeling.

Keywords NLCD, MODIS, WRF, CMAQ, LSM, landuse, land surface model, meteorology model, air quality model

1. Introduction

Land cover data are used in meteorology, emission, and air quality modeling. In meteorology modeling land surface exchange processes are based on land cover categories within each modeling grid (e.g. Xiu and Pleim, 2001). In emission modeling, land cover data are used to spatially allocate some county-based emission inventories to modeling grids and to compute biogenic, dust, and fire emissions. In air quality modeling chemical surface fluxes are modeled based on different land cover categories (e.g. Pleim et al., 2001). Meteorology, emission, and air quality models often use land cover data created in the early 1990s. The land cover data that comes with WRF are USGS Global Land Cover Characteristics (GLCC) 30-s (~1 km) land use data which were developed from 1 km Advanced Very High Resolution Radiometer (AVHRR) satellite images obtained in 1992–1993. Currently, there are two new land cover data sets available for the US and Globe. One is the 2001 National Land Cover Data (NLCD) at 30 m resolution generated from Landsat 7 and 5 Thematic Mapper (TM) images for the entire U.S. (Homer et al., 2004) and the other is the 2001 NASA Moderate Resolution Imaging Spectro-radiometer (MODIS) land cover data at 1 km resolution generated from TERRA MODIS satellite images for the Globe (Friedl et al., 2001). We have developed programs to process the 2001 NLCD (for U.S. areas) and MODIS data (for areas outside U.S.) into fractional coverage of each land category for use in the Weather Research and Forecast (WRF) and Community Multiscale Air Quality (CMAQ) models. The gridded land use data are used to specify vegetation and surface related parameters that are needed by land surface models (LSM), dry deposition models, and biogenic emission model.

2. Data and Methods

The NLCD data base was developed by a group of federal agencies under the Multi-Resolution Land Characteristics (MRLC) Consortium since 1992. The NLCD 2001 data from MRLC includes 21 classes of land cover, tree canopy percent, and imperviousness percent at each 30-m pixel. NOAA Coastal Services Center Coastal Change Analysis Program (C-CAP) NLCD data has 30 land cover classes which includes detailed costal wetland classes in addition to MRLC NLCD 21 classes. NASA MODIS land cover data includes 20 International Geosphere-Biosphere Programme (IGBP) classes.

Those land cover data are stored in pixel-based format, which is often called raster data. We developed C++ raster tools using the Geospatial Data Abstraction Library (GDAL) in the Spatial Allocator (SA, http://www.ie.unc.edu/cempd/projects/mims/spatial/) to compute gridded land cover information from MRLC 2001 NLCD (including land cover, imperviousness, and canopy), NOAA coastal 2001–2006 NLCD land cover, and NASA 2001 MODIS land cover data (for areas

outside the US). There are two steps (run by two script files) in processing those new land use data sets. The first step is to pre-process downloaded NLCD data sets to get rid of spatial overlaps among data sets. The second step is to compute the gridded land cover fraction of each land use category for user-defined modeling grids from pre-processed NLCD data sets and MODIS data. The program outputs gridded fractional coverage of each land category stored in a WRF-ready NetCDF file. Figure 1 shows the 30 m NLCD (top) and the WRF GLCC land cover data (bottom) for deciduous forest processed for 1, 4, and 12 km resolution grids. For the higher resolution grid domains (1 and 4 km), gridded NLCD data have much finer gradation and better representation of deciduous forest distribution. Gridded GLCC land use data tend to underestimate or overestimate percent of deciduous forest coverage.
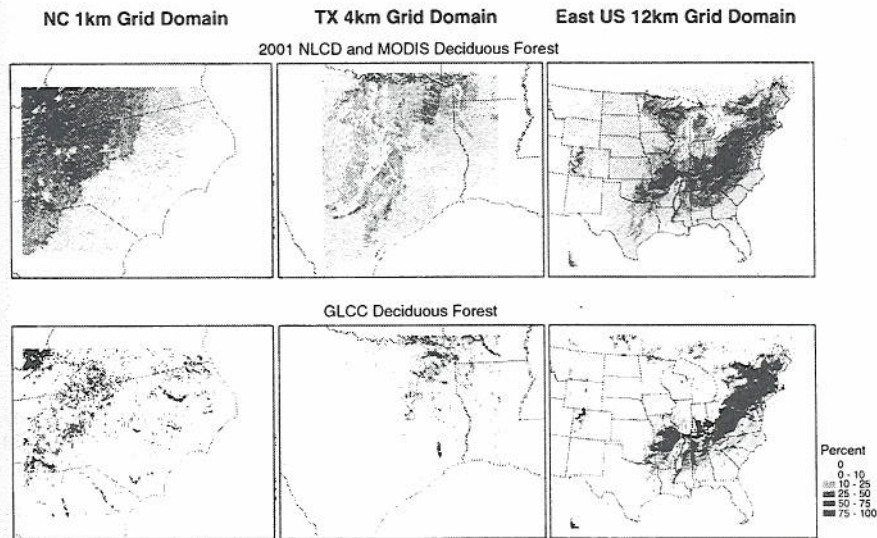

Fig. 1. Gridded land cover – deciduous forest at different grid scales

Vegetation and surface related parameter tables for the Pleim-Xiu land surface model (PX SLM) were modified based on NLCD and MODIS land cover categories. Then, we ran the WRF and CMAQ models for August 2006 on the continental US 12 km grid domain using both USGS GLCC and the new NLCD/MODIS land cover data sets. We used WRF – ARW v3.0 with analysis nudging winds for all levels, and T, and q_v above PBL and indirect soil moisture and temperature nudging in PX LSM. CMAQ v4.7 with CB05 gas-phase chemistry, AE5 modal aerosols, M3dry dry deposition w/ preliminary bidirectional NH3, and ACM2 were used for this study. 2006 emissions and Biogenic Emissions Landuse Database, version 3 (BELD3) data were emission inputs to CMAQ.

3. Results and Analysis

Preliminary WRF output comparisons show small differences in mean absolute error (MAE) at the 12 km grid resolution domain (Fig. 2) with the WRF/NLCD run slightly better in most of the eastern part of the domain but slightly worse in the middle part where shrublands and grasslands dominate. Note that these differences are very small due, in part, to the soil moisture nudging scheme in the PX LSM that minimizes biases in 2-m temperature and mixing ratio. Largest difference between CMAQ outputs is in the bi-directional NH3 surface flux because this is closely related to landuse, especially crops which results in a significant difference in NH_3 air concentrations (Fig. 2 right panel).
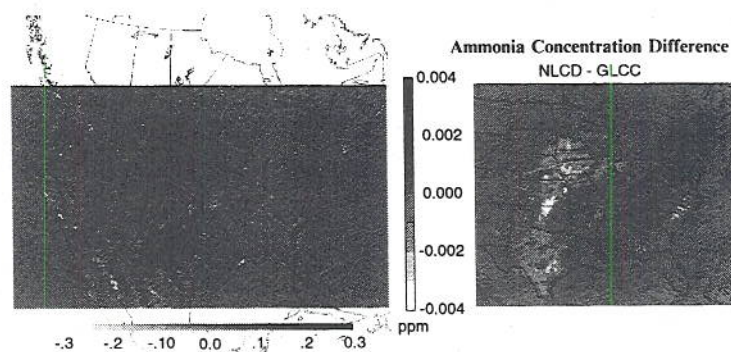

Fig. 2. WRF and CMAQ comparisons. Difference in 15-day mean absolute error in 2-m temperature for WRF/NLCD – WRF/GLCC (left) and difference in ammonia concentration at 20 UTC on August 20, 2006 for CMAQ/NLCD – CMAQ/GLCC (right)

4. Conclusions and Future Studies

The preliminary model results from WRF and CMAQ with the new high resolution land cover data are encouraging. We expect much greater improvement for higher grid resolution model runs (4 and 1 km). We are also developing Biogenic Emission Landcover Data version 4 (BELD4) land use data from 2001 30 m NLCD, 2001 1 km MODIS, USDA FIA data, and USDA National Agricultural Statistics (NASS) crop data, which are planned to be the input to estimate biogenic emissions from both the Model for Gaseous and Aerosols from Natural (MEGAN) sources and the Biogenic Emission Inventory System (BEIS).

References

Friedl, M.A., D.K. McIver, X.Y. Zhang, J.C.F. Hodges, A. Schnieder, A. Bacinni, A.H. Strahler, A. Cooper, F. Gao, C. Schaaf, W. Liu, 2001: Global land cover classification results from MODIS. Geoscience and Remote Sensing Symposium, IGARSS apos; 01. IEEE 2001 International 2, pp.733–735.

Homer, C, Huang, C, Yang, L, Wylie, B. and Coan, M. 2004: Development of a 2001 National Land-cover Database for the United States. *Photogrammetric Engineering and Remote Sensing*, 70:829–840.

Pleim, J. E., Xiu, A., Finkelstein, P. L., Otte, T. L., 2001. A coupled land-surface and dry deposition model and comparison to field measurements of surface heat, moisture, and ozone fluxes. Water, Air, & Soil Pollution: Focus 1, 243–252.

Xiu, A., and J.E. Pleim, 2001: Development of a Land Surface Model. Part I: Application in a Mesoscale Meteorological Model. *J. Appl. Meteor.*, **40**, 192–209.

5. Questions and Answers

Question 1: How is the bi-directional ammonia surface flux modeled in relation to land use?

Answer 1: Currently, the ammonia bi-directional model specifies the gamma values ($NH4^+/H^+$) for soil and stomata according to land-use category such that cropland uses gamma = 1,000 and all other vegetated LU categories use gamma = 100. This is a prototype. We are developing methods to derive more realistic gamma values from data on crop types and fertilizer application.

Question 2: Nowadays land use data is available on a very detailed scale, but it depends on the algorithms used to extract information from the signal. Do we know the accuracy of this procedure and how it has been tested?

Answer 2: Land use data is normally generated from satellite images with digital signals based on classification algorithms, such as cluster algorithms. The 2001 National Land Cover Database (NLCD) has been developed under the Multi-Resolution Land Characteristics (MRLC) Consortium which is a group of US federal agencies. It is well documented database with detailed accuracy assessments. 1992 NLCD land cover dataset was first developed in the early 1990s from Landsat 5 TM images. 2001 NLCD data was developed in the recent years mostly from Landsat 7 TM images. This new 30 m resolution land cover data set is well documented and evaluated based on ground truth field information by different agencies. The data set has been created from an improved classification algorithm in comparison with the 1992 NLCD land cover data. The technical methods and algorithms used to generate the 2001 NLCD data set is published in a journal article.