Overlapping Interests: The Impact of Geographic Coordinate Assumptions on Limited-Area Atmospheric Model Simulations

ANDREW J. MONAGHAN, MICHAEL BARLAGE, JENNIFER BOEHNERT, CODY L. PHILLIPS, AND OLGA V. WILHELMI

National Center for Atmospheric Research,* Boulder, Colorado

(Manuscript received 5 December 2012, in final form 1 March 2013)

ABSTRACT

There is growing use of limited-area models (LAMs) for high-resolution (<10 km) applications, for which consistent mapping of input terrestrial and meteorological datasets is critical for accurate simulations. The geographic coordinate systems of most input datasets are based on spheroid-shaped (i.e., elliptical) Earth models, while LAMs generally assume a perfectly sphere-shaped Earth. This distinction is often neglected during preprocessing, when input data are remapped to LAM domains, leading to geolocation discrepancies that can exceed 20 km at midlatitudes.

A variety of terrestrial (topography and land use) input dataset configurations is employed to explore the impact of Earth model assumptions on a series of 1-km LAM simulations over Colorado. For the same terrestrial datasets, the ~20-km geolocation discrepancy between spheroidal-versus-spherical Earth models over the domain leads to simulated differences in near-surface and midtropospheric air temperature, humidity, and wind speed that are larger and more widespread than those due to using different topography and land use datasets altogether but not changing the Earth model. Simulated differences are caused by the shift of static fields with respect to boundary conditions, and altered Coriolis forcing and topographic gradients.

The sensitivity of high-resolution LAM simulations to Earth model assumptions emphasizes the importance for users to ensure terrestrial and meteorological input data are consistently mapped during preprocessing (i.e., datasets share a common geographic coordinate system before remapping to the LAM domain). Concurrently, the modeling community should update preprocessing systems to make sure input data are correctly mapped for all global and limited-area simulation domains.

1. Introduction

Computational advances and the need for high-resolution weather and climate data for numerous applications have led to increasingly fine spatial resolution in limited-area atmospheric model (LAM) simulations, now regularly at scales of 1–10 km (e.g., Im et al. 2010; Rasmussen et al. 2011; Monaghan et al. 2012). To enhance the fidelity of simulations, the implementation of finescale terrestrial fields (e.g., Chen et al. 2011) or assimilated point observations (Hahmann et al. 2010) into LAMs has grown. Such datasets can be regional or global and are mapped with a number of different coordinate systems that define locations on Earth to optimally preserve properties such as area, shape, distance, and direction (Bugayevskiy and Snyder 1995). Coordinate systems are geographic: based on a three-dimensional spheroidal Earth model and latitude–longitude coordinates; or projected: based on a two-dimensional coordinate plane derived from mathematically converting locations from the spheroidal Earth. Both types of coordinate systems employ datums consisting of the dimensions of a spheroid and how that spheroid fits onto Earth (orientation and origin). While many of the geophysical datasets assimilated into LAMs are based on spheroidal (i.e., elliptical) Earth models, most LAMs assume Earth’s surface is a perfect sphere in order to simplify calculations. At midlatitudes the location of a point can vary between spheroidal and spherical Earth models by more than 20 km (Fig. 1). In the past, when

* The National Center for Atmospheric Research is sponsored by the National Science Foundation.

Corresponding author address: Andrew J. Monaghan, National Center for Atmospheric Research, P.O. Box 3000, Boulder, CO 80307-3000.

E-mail: monaghan@ucar.edu

DOI: 10.1175/MWR-D-12-00351.1

© 2013 American Meteorological Society
computational limitations necessitated LAM simulations with coarser spatial resolution (i.e., $>25$ km). Earth model differences among input data to LAMs could often be neglected, and were, because the impact on simulations was assumed minor compared to other sources of uncertainty such as physical parameterizations and error growth. Accordingly, the built-in data preprocessing software packages that generate the initial and boundary conditions for LAM simulations may have simplified treatments of coordinate systems that in some cases neglect transformations between different Earth models. At the fine spatial scales now typical of LAM simulations, neglecting these differences may lead to significant geolocation discrepancies among input datasets that affect the fidelity of simulations or impact downstream users who require accurately mapped LAM output to drive other geophysical models or for GIS applications. For example, accurately geolocated LAM output is critical for driving high-resolution hydrological models because subtle geolocation errors can determine whether or not a small catchment is at risk for flooding (David et al. 2009).

We performed a series of high-resolution LAM simulations to explore the impacts of Earth model assumptions that occur when implementing alternative terrestrial fields in a LAM. How sensitive are LAM simulations to the assumptions made when ingesting these data? Methods are described in section 2, results are presented in section 3, and conclusions are drawn in section 4.

2. Methodology

The LAM simulations were performed with Version 3.3 of the Advanced Research Weather Research and Forecasting Model (WRF; Skamarock and Klemp 2008) and the accompanying WRF Preprocessing System (WPS; WRF Development Team 2011). WRF has multiple options for physical process parameterizations that are optimized for each unique application. The following parameterizations were chosen based on robust results for prior work for an overlapping region (Trier et al. 2011): Thompson cloud microphysics (Thompson et al. 2004); Rapid Radiative Transfer Model longwave radiation scheme (Mlawer et al. 1997); Dudhia shortwave radiation scheme (Dudhia 1989); Eta Model similarity for the surface layer (Janjic 2002); Mellor–Yamada–Janjic for the planetary boundary layer (Janjic 2002); and Kain–Fritsch for cumulus clouds (Kain 2004). Two-way land–atmosphere interactions in WRF were simulated with the Noah land surface model (Chen and Dudhia, 2001). The WPS software contains three programs that prepare terrestrial and meteorological data for input to WRF. Of particular interest here is geogrid.exe, which defines the model domains and interpolates terrestrial (topography, land use, etc.) fields to the model domains. The use of geogrid.exe is described below.

A 1-km domain was one-way nested inside a 3-km domain over a mountainous region in central Colorado (Fig. 2). This region was chosen for its large elevation gradients and diverse land use patterns, both being

![Fig. 1. Comparison of points with equivalent latitudes $\Phi$ on spheroidal (red) vs spherical (green) Earth models. The true shape of Earth’s surface that is approximated by the models is shown in blue. Latitude is defined as the location where a line with angle $\Phi$ to Earth’s equatorial plane intersects the spheroid (point “A”; solid line) or sphere (point “B”; dashed line) at a normal angle. Because lines with the same latitude intersect the equatorial plane at different locations [i.e., at variable locations along the plane for the spheroid (“geodetic”) vs at the center of the plane for the sphere (“geocentric”)], the location of point A when placed on the sphere (point “C”) will be different compared to point B. Stated another way, one would have to apply a coordinate transformation to point A in order to properly geolocate it on the sphere at point C, otherwise, point A will be improperly geolocated at point B if one assumes the spheroidal latitude is equivalent to the spherical latitude (as is usually the case when users remap WGS84 datasets onto LAM projections). The table in the bottom-right quadrant shows the difference in distance between points B and C that arise from this assumption. The maximum difference of $-21.4$ km occurs at 45° latitude [Eq. (3.12) from Pearson (1990)]; there is no difference at 0° and 90° latitude. An example of the difference between spheroidal and spherical coordinates is shown in Figs. 3b,c at $-39^\circ$N latitude. A mountain is located at point B in the WGS84 topographic dataset that is not shifted prior to projecting it onto the LAM sphere (Fig. 3b), but it is located at point C in the shifted dataset (Fig. 3c). Note that the point appears to be properly geolocated in Fig. 3b because the plotting software assumes WGS84 coordinates.]}
integral to our experiments. The initial and boundary meteorological conditions were derived from the 32-km North American Regional Reanalysis (NARR; Mesinger et al. 2006) which is mapped on a spherical Earth model. The WRF domains have 57 vertical levels from the surface to 100 hPa, with 13 levels in the lowest 1000 m. The experiments described below were performed for a 12-h daytime period 1200 UTC 4 July 2009–0000 UTC 5 July 2009, chosen because topography and land use play an important role in the development of the summer daytime boundary layer (e.g., Lu et al. 2012), thereby facilitating a comparison of topographic and land use impacts on simulations. A larger LAM domain would normally be used to minimize boundary effects (e.g., Warner et al. 1997); however, we employed a comparatively small domain to provide a focused study region. The simulations produced realistic results suitable for assessing the impacts of mapping assumptions.

Geogrid.exe currently recognizes geographic coordinate systems (GCSs; i.e., “latitude–longitude”), as well as three projected coordinate systems for input terrestrial datasets: Lambert conformal, Mercator, and polar stereographic. The same four coordinate systems can also be user specified when generating the WRF domain with geogrid.exe. When an input static dataset (e.g., topography, land use) is transformed from its native coordinate system to the chosen WRF coordinate system, geogrid.exe assumes that all coordinates lie on a perfect sphere (i.e., when geogrid.exe does the coordinate system transformation it does not do an Earth model transformation). For the simulations described below, all input datasets had GCSs, and the WRF domain onto which they were interpolated by geogrid.exe was a Lambert conformal projection.

Three simulations were performed, differing only by the input topography and land use datasets used to generate the WRF terrestrial fields within geogrid.exe. The first simulation (“CONTROL”) employed the default U.S. Geological Survey (USGS) Global 30 arc-s elevation dataset topography (GTOPO30; Gesch and Greenlee 1996) and land use categories (24 types) that come packaged with WPS. These data are based on the World Geodetic System 1984 (WGS84) GCS, whose spheroidal Earth model has a semimajor (equatorial) axis radius of 6 378 137 m; however, geogrid.exe assumes that they use a spherical Earth model with a radius of 6 370 000 m. The second simulation (“NEW”) employed two static datasets not available in WPS, version 3.3: the 3 arc-s (~90 m) National Geospatial-Intelligence Agency (NGA)/National Aeronautics and Space Administration (NASA) Shuttle Radar Topography Mission, version 2.1, dataset (SRTM; Farr et al. 2007; http://dds.cr.usgs.gov/srtm/) for topography, and the 3-arc-s USGS National Land Cover Database 2006 dataset (NLCD; Fry et al. 2011; http://www.mrlc.gov/nlcd06_data.php) for land use. The NLCD categories (33 total) were recategorized to the USGS 24-category system for consistency with the land use categories in the CONTROL simulation. Within the simulation domain, the NLCD categories present could be mapped directly to existing USGS categories, so no categorical mapping decisions were necessary. The native WGS84 GCS on which both SRTM and NLCD are mapped was retained for ingest to geogrid.exe. The third simulation (“NEW_SHIFT”) was the same as the NEW simulation, except that a transform (following Hedgley 1976) from the WGS84 spheroid to the default WRF sphere was applied to the SRTM and NLCD datasets prior to ingesting them into geogrid.exe. This transformation effectively validates the assumption by geogrid.exe of a spherical GCS, so that the latitudinal data for NEW_SHIFT were subsequently mapped “correctly” by geogrid.exe. Next, we discuss the topographic and land use mapping differences among the three cases and present results for the simulations.
3. Results

Figures 3 and 4 present the topographic and land use fields, and their differences, for the three cases. The geolocation of the topography and land use are the same for the CONTROL and NEW cases, since the native data for both used the WGS84 spheroid as their Earth model, but geogrid.exe assumed they were based on a sphere. Differences between the CONTROL and NEW datasets (Figs. 3d and 4d) are due to variation among their source datasets. While the differences in topography tend to be subtle (≤ 25 m over most of the domain), the differences among the land use datasets are marked. Urban areas are better defined in the NLCD land use data employed for the NEW case, as are the treeless mountainous areas lying above ~3700 m: these are correctly categorized as “barren/sparsely vegetated” in NEW, versus “mixed forest” in CONTROL.

The topography and land use for the NEW_SHIFT case (Figs. 3c and 4c) lie about 20 km to the south compared to CONTROL and NEW, because the data were transformed from WGS84 to the sphere before they were ingested and remapped to the WRF domains by geogrid.exe. As expected from the ~20-km shift, the differences between the topography and land use in
NEW versus NEW SHIFT are large (Figs. 3e and 4e). Note that interpolation techniques led to subtle differences between the NEW and NEW SHIFT topography and land use fields when mapped onto the WRF domain, despite being from the same input dataset, because the Earth model shift (20.88 km at 39°N) was not exactly proportional to the 1.00-km WRF grid spacing. These subtle interpolation differences are apparent in Figs. 3f and 4f, in which the difference between the topography and land use in NEW and NEW SHIFT is shown after realigning NEW SHIFT 20 grid boxes (i.e., 20 km) to the north, which is the grid increment for which NEW and NEW SHIFT best correspond when realigned. The “realigned” NEW SHIFT data facilitates interpreting the causality for differences among the simulations, as discussed below.

Figure 5 presents the results for 2-m temperature 10 h into the simulations, at 2200 UTC 4 July 2009 (1600 LST). The differences between CONTROL and NEW (Fig. 5d) are generally within ±1°C, suggesting a modest impact on simulations due to differences between the source datasets. The differences between NEW and NEW SHIFT (Fig. 5e) are substantially larger because of the ~20-km shift in topography, indicating that the manner in which the Earth models of terrestrial input data are treated has first-order impacts on high-resolution simulations. If the NEW SHIFT data are realigned so that they correspond with the NEW data (Fig. 5f), the differences due to shifting the topography by 20 km (Fig. 5e) are largely removed, isolating the other shift-related differences among the simulations caused by 1) the differential positioning of topographic/land use fields with respect to the NARR boundary conditions (which do not shift); 2) subtle topographic/land use interpolation differences (Figs. 3f and 4f); and 3) latitude-dependent physics (i.e., the Coriolis forcing and map factors that appear in the LAM’s momentum equations; this effect is likely comparatively minor). The collective impact of these influences is comparable to that of using different terrestrial

**FIG. 4.** As in Fig. 3, but for the (a)–(c) WRF land use index (categorical) and (d)–(f) differences (grid points with land use categories that differ between respective simulations are indicated with black).
input datasets (Fig. 5d). Therefore, even if LAM output are correctly remapped during postprocessing, there are still important localized impacts on the simulations due to the three factors noted above, which cannot be rectified.

To quantify the results shown visually above and examine whether they hold for other variables/levels, we performed a statistical comparison between the three WRF cases for near-surface and 6000 m AGL air temperature, specific humidity, and wind speed (Table 1). Statistics are calculated spatially for overlapping grid points of the three domains (see Fig. 2) for 2200 UTC 4 July 2009, therefore representing the domainwide differences between cases. For all variables and both levels, the RMSDs (correlations) are largest (smallest) for the “NEW-vs-NEW_SHIFT” comparison, indicating that the manner in which the Earth model of the terrestrial input data is treated has a consistently greater impact on simulations than changing the source of the terrestrial input data (i.e., the “CONTROL-vs-NEW” comparison). Additionally, comparing the CONTROL-vs-NEW results to the NEW-vs-NEW_SHIFT (realigned) results indicates that RMSDs and correlations of both CONTROL and NEW_SHIFT (realigned) with the NEW case are similar. This reinforces the finding that the collective impact on simulations of differential boundary forcing, terrestrial data interpolation, and latitude-dependent physics—all three due to the different Earth model treatments—is of the same order as for changing the terrestrial input datasets. The results in

![Figure 5](image-url)
Table 1 remain approximately the same for other levels and simulation time steps (not shown) despite model error growth (larger RMSDs) as time progresses.

4. Conclusions

The way Earth models are treated when mapping terrestrial datasets to a high-resolution projected LAM domain has important impacts on simulated meteorological fields at a midlatitude location. For the same terrestrial dataset, the ~20-km shift between spheroidal and spherical Earth models causes simulated differences larger than those due to implementing an alternative terrestrial dataset. Even if the geolocation of the LAM output is “corrected” during postprocessing, as it usually is, residual differences of shifting versus not shifting the coordinates are similar to those due to implementing an alternative terrestrial dataset.

Determining which Earth model to use when mapping input data to a LAM ideally requires knowledge of the mapping assumptions used in the global or regional atmospheric model output that provides the LAM’s initial and boundary forcing. Generally, these model data have GCSs with spherical Earth models. However, the terrestrial datasets over which these global or regional simulations were performed are usually derived from spheroidal models (e.g., WGS84), but are assumed to be spherical (e.g., Taylor 2012), just as for our LAM CONTROL simulations. In this common case, the best approach is to ensure that terrestrial and other input data (e.g., assimilated observations) are based upon the same spheroid as the forcing model’s terrestrial data, since the positioning of the terrestrial data has first-order impacts on the simulated fields. Fortunately, the user’s main task toward meeting this objective, which can be verified by checking whether all input data align for a common reference point such as a mountain, is usually just confirming that new input datasets are based on WGS84 spheroids because of the prevalent use of WGS84 for global geophysical datasets. Even though these terrestrial data will be incorrectly geolocated on the LAM sphere, this approach minimizes the impact on simulations of otherwise mismatching the terrestrial datasets between the forcing model and the LAM.

It is increasingly important to resolve mapping differences among input datasets in light of rapid growth of high-resolution LAM usage (Warner 2011). As noted above, LAM users should minimally ensure that terrestrial and meteorological input data are consistently mapped. Concurrently, to reduce simulation errors arising from improper geolocation of input datasets—related to the differential positioning of terrestrial fields among forcing models and LAMs, latitude-dependent physics, and improperly resolved gradients—the atmospheric modeling community should work to ensure input datasets are correctly mapped for all global and limited area model domains.

Acknowledgments. This work was funded by the National Aeronautics and Space Administration Grant NNX10AK79G. Michael Duda and two reviewers provided useful comments. Sources and citations for datasets are described within.

REFERENCES


