Geospatial datasets for use in NWP (and climate) models

Jeanette Onvlee, Laura Rontu, Ekatharina Kourzeneva, 20140116

1. Introduction

NWP and climate models all use external databases containing geospatial information:

- topography: e.g. GTOPO30 (global, 1km resolution), GMTED2010 (global, 250m), SRTM (~100m), ASTER (~30m), national databases (down to O(1-10m) resolution)
- soil texture (e.g. HWSD)
- land cover or land use (e.g. Corine, Globcover, Ecoclimap)

These geospatial datasets are based on very high resolution satellite information (from e.g. MODIS or Landsat). Normally raw data from quite a few satellite passes have to be combined to obtain a more or less complete data coverage. Nonetheless, in particular for data based on optical (cloud-affected) sensors, some gaps may remain, and in regions with frequent clouds data can be of less quality.

Several types of classification systems can then be used to derive from the satellite composites land use/land cover information. Differences in the categories adopted can make it difficult to compare or combine data generated by different classification systems. To obtain large-scale datasets, sometimes various national geospatial datasets are combined. This introduces a serious risk of inhomogeneities in resolution and/or classification.

The final step to create the orographic and physiographic data which are actually used in an atmospheric model, is to select the data over the model domain and filter them to the resolution and parameters used by the model. The tools used for this are specific to each model. Generally, each of the European short-range NWP consortia has about 1 expert who is truly knowledgeable on this final production step.

In the 2013 EWGLAM meeting, a general feeling was expressed that the issue of geospatial data could be an interesting topic for further cooperation. This document is a discussion paper providing a rationale and possible options for joint activities in this field, and aiming to assess the interest and ideas of the SRNWP consortia for such a new activity.

2. The quality of geospatial data: sufficient or ???

It is to be expected that geospatial datasets suffer from quality problems like any other dataset: data gaps, artifacts, misclassifications, etc. NWP and climate modelers usually have very little insight in how good or bad these external data really are. Most NWP developers simply assume that the data are sufficiently accurate (in the sense that they are not limiting for model performance), but few efforts have been made to truly assess this. One complicating factor in that respect is that such an assessment is often not trivial to make for someone who is not experienced in the use of geospatial data handling tools.

So how reliable are the orographic and physiographic data used in presently operational SR NWP models, really? And to what extent does their quality affect NWP model performance? Very little is generally known about this model sensitivity to its geospatial input information. There are indications that well-known and widely used geospatial datasets do indeed suffer locally from errors that are capable of impacting model performance quite negatively. Within the Hirlam community, there have been some efforts in the past years to assess and inter-compare several geospatial databases. Some of their experiences (see also the presentation http://srnwp.met.hu/Annual_Meetings/2013/download/wednesday/Ekaterina_Kurzeneva.pdf):

- The database used for physiographic information in the Harmonie model is Ecoclimap. When first testing Harmonie on various domains, it was noted that the

Ecoclimap-1 database used at that time showed some serious defects and artifacts. For example, in Byelorussia several large artificial lakes were present which do not exist in reality. In a second version, Ecoclimap-2, many of these artifacts were removed. This led among others to the strong reduction of a wind bias (from 3 m/s to ~0.7 m/s) which had been seen in the Lithuanian Harmonie model until then.

- In tests over various other Harmonie domains, the use of Ecoclimap-2 generally also resulted in a better meteorological performance than Ecoclimap-1. However, the new Ecoclimap-2 database also contained some potentially serious new errors with land-water transitions. No less than 40% of the (large) Curonian lagoon in the Baltic appears as land. Land-sea and land-lake transitions appear to be shifted with respect to higher-resolution digital elevation models. How damaging these deficiencies are in terms of meteorological impact remains to be tested. Similar types of error have also been seen in other comparable global databases.
- In Harmonie Cy38, a new sand and clay database, HWSD, has been tested. This database is generally of higher quality and greater detail than its predecessor, but local anomalies can be seen. E.g. the level of spatial detail in Scandinavia has been seriously reduced, and the border between Sweden and Norway clearly shows up in it, suggesting a problem with the homogeneity of classification in that area.
- The topography database routinely used in Harmonie is GTOPO30. Experiences have been gained with using alternative (newer, higher resolution) DEM datasets like GMTED2010, SRTM, ASTER and national datasets. In mountainous terrain, the highest resolution DEM data provide the most accurate information on elevation and slopes. However, not enough verification information is presently available yet to assess the added value of the alternative datasets in terms of NWP forecast quality.

These experiences suggest that deficiencies and inhomogeneities exist in several widely used geospatial datasets which sometimes have a quite significant negative (local) impact on model quality. A database which is of high quality over one region can be relatively poor over another. Quality problems may be higher for high-latitude areas, due the greater difficulties to get sufficient (cloud-free) satellite coverage there. Waiting for the next generation database to solve such problems does not appear a very satisfactory response; each new database, while presumably of better quality than its predecessors, will have problems of its own to discover and solve.

With the growing level of experimentation with models on sub-km resolutions and with the increasing sophistication of model (surface) parametrizations, the demands on geospatial data availability, resolution and quality are likely to grow further. In the coming years there is likely to be a strong drive to push horizontal resolutions down towards several hundred meters. Models on such hectometric scales will have need of high quality geospatial data of great spatial detail, which are gradually becoming available. It will be important for the NWP community is well aware of which databases they can (or will soon be able to) employ, and what their strong and weak points are in terms of quality. Very detailed national databases often do exist, but model domains usually do not follow national borders. Also, national databases often are not easily accessible or freely available for international use.

Additionally, there is an increasing trend towards more complete earth system modeling. This will require the use of a wider variety of geospatial data (e.g. on vegetation and biology) to describe new non-atmospheric model components.

3. Challenges for the NWP community

A prime challenge for the NWP community is to enhance the awareness of modelers that the geospatial information used by the model may sometimes significantly affect forecast quality. Otherwise modelers could for example run a real risk of attributing a poor model performance

which is actually due to deficiencies in the geospatial data, to model error instead. This could lead to a misdirection in research efforts.

A second crucial issue to tackle is to assess the sensitivity of the atmospheric model to the quality of the external data used by it. How to assess the quality of existing orographic, soil texture or land cover databases? And how to assess the impact on an atmospheric model of the choice for, or the deficiencies in, a certain data source? Stronger efforts to analyze this, and share experiences on it, are clearly needed. One problem for modelers is that even when they do suspect that a certain model behaviour may be due to the underlying geospatial data, this may not be trivial for them to check. Insight in the origins and characteristics of the database and some expertise with visualization and manipulation tools for geospatial information are often needed for the inspection of orographic or physiographic data.

Experiences such as described above show the desirability to more stringent evaluate the quality and impact of geospatial data used in our models, and possibly to consider other sources or other ways of handling these data. However, the NWP community clearly suffers from a lack of expertise in this respect. Within each of the SR NWP consortia in Europe there are generally at most two or three people with good knowledge on handling such data. At present we are highly dependent on the geospatial community to provide, analyze, and solve problems in, geospatial data for us. Given their potentially significant impact on NWP model performance, that is not a situation which seems attractive to continue in the future.

Actions that the NWP community could consider in this respect, are:

- Make better use of the expertise that we already have, by starting to share experiences between the present "geospatial experts" of the SR NWP consortia: what do we presently use, what are the experiences and known problems, what are future needs and plans for geospatial data, how could cooperation help?
- Establish closer links with the "real experts" in the geospatial community: with the community providing relevant geospatial databases (exchange of experiences, discussion of NWP/climate modeling requirements), and more widely also with experts in handling and manipulating geospatial information (university groups?), to help us to analyze the datasets that we use.
- Ultimately, we may wish to strengthen our capability for identifying and solving problems with geospatial data ourselves: either obtain new staff or train a wider group of existing staff how to at least analyze existing databases in relation to NWP needs. Possibly share tools and knowledge (training?) for handling geospatial data.

4. Proposed way forward

It is proposed that the SRNWP community explores the possibilities of cooperation in the field of geospatial data in a two-step procedure:

- First assess, by means of a questionnaire, the present situation and ideas in the SR NWP consortia on this issue:
    - For the consortium and the NMS's participating in it: which geospatial databases are available, used or planned to be used in the future? What are the existing expertise, available tools, experiences with quality assessment?
    - What are the ideas on future directions and needs for geospatial information, possible future partners, the perceived need for capacity building?
    - Who would be the contact points able to, and interested in, work on this together? What are the contacts with partners in the geospatial community?

  A draft questionnaire is attached hereby.

- Then get the relevant proposed experts together in a web meeting or workshop to discuss ideas and possibly shared ambitions, e.g. to:
    - Make an inventory of available databases and tools
    - Share experiences on the impact of geospatial data on model quality
    - Make an inventory of contacts with the geospatial data providers and experts, and define steps to enhance these contacts
    - Assess needs and possibilities for capacity building/ training activities
    - Assess future needs for geospatial data, and identify promising data sources for them
    - Consider to engage other relevant partners from e.g. the climate or earth system modeling communities in similar activities
    - Write a proposal for a common action (externally funded, or e.g. COST)?

Based on the outcome of these two steps, decisions can be made on the desirability for further specific SRNWP activities in this field.

The SRNWP consortia are hereby asked to consider:

- whether or not they believe the issue of geospatial data to be an interesting and useful new topic for cooperation within SRNWP;
- whether or not they agree with the proposed two-step approach to initiate such an activity;

and are cordially invited to provide any ideas and suggestions of their own concerning this topic and how it should be handled within SRNWP (or not).

Questionnaire:

1. **Present situation: general information**
   - *(a)* Which geospatial databases (topography, soil texture, land cover/land use) are presently used in your consortium's modeling system routinely? Are these databases freely available? Are you willing/able to share them with other consortia?
   - *(b)* Do you have experience with using alternative geospatial databases in experimentation or operations? If so, which? Are these databases freely available? Are you willing/able to share them with other consortia?
   - *(c)* Do you have plans of changing to different geospatial databases within the next two years? If so, which databases and why?
   - *(d)* Is the geospatial database (or databases) which you employ used only for NWP purposes, or also for climate modeling or other purposes? If so, which?
   - *(e)* Which tools are used by your consortium expert(s) to create, inspect and manipulate the geospatial database used in your model? Are these tools available for sharing?

2. **Existing expertise and contacts**
   - *(a)* who is/are the expert(s) responsible for the maintenance of the geospatial data in your modeling system? Are there other people knowledgeable in the handling of geospatial data as well? If so, who?
   - *(b)* do you have contacts in the geospatial community (database providers, or more general experts in geospatial data handling)? Do you think that the present level of contact is sufficient, or that it should be enhanced?

3. **Quality aspects**
   - *(a)* what experience do you have with assessing the quality of geospatial data? If so, please describe your method of assessment (against which data) and experiences.
   - *(b)* What, if any, are known or suspected weaknesses in the geospatial datasets that you use? Have you assessed the impact of deficiencies in the geospatial data on your model performance? If so, how?
   - *(c)* In case of noted or suspected deficiencies: How do you plan to address them? What actions do you undertake?
   - *(d)* Are you interested in intensifying efforts, within your consortium or within the NMS's partaking in it, on assessing the quality of geospatial data and its impact on model quality?
   - *(e)* Are you interested in sharing experiences and common efforts on this in SRNWP context? If so, do you have suggestions for how such a cooperation could take shape?

4. **Capacity building**
   - *(a)* do you feel that in your consortium there is sufficient expertise in (i) the construction of a geospatial dataset as input to your model, (ii) visualization and manipulation of these data in the context of model development and evaluation?
   - *(b)* do you feel a need for strengthening the capabilities of your main expert(s)?
   - *(c)* do you feel a need for widening the expertise on geospatial data handling and manipulation to a wider group of people (model developers)?
   - *(d)* Can/do you make use of expertise provided by geospatial data experts (cooperating with research institutes, state agencies, universities, etc)? If so, which partners?

5. **Future needs**

(a) How do you think that the requirements for geospatial data for your model will develop in the coming years (in terms of resolution, range of physical parameters, …)?

(b) Do you believe that these needs can be met by existing geospatial data sources? Or that data sources for this will become available in the near future?