

# External data for lake parameterization in Numerical Weather Prediction and climate modeling

Ekaterina Kourzeneva

*Russian State Hydrometeorological University, Malookhtinsky pr. 98, RU-195196 St. Petersburg, Russia*

*Received 18 Feb. 2009, accepted 30 Aug. 2009 (Editor in charge of this article: Kai Myrberg)*

Kourzeneva, E. 2010: External data for lake parameterization in Numerical Weather Prediction and climate modeling. *Boreal Env. Res.* 15: 165–177.

Lake parameterizations in atmospheric modeling include a set of external data to indicate and to map physical properties of lakes. The main challenge is the need to consider all the lakes in the atmospheric model domain and to specify the corresponding parameters. For Numerical Weather Prediction (NWP), we also need the data to initialize the lake time-dependent variables (so-called cold start data). The first steps to make the set of lake parameters for the needs of atmospheric modeling are described in this paper. The mean lake depth was chosen to be the key lake parameter for which direct measurements were collected and processed. The Global Land Cover Characteristics (GLCC) dataset was used for mapping, and the mapping method was based on a probabilistic approach. Empirical Probability Density Functions were used to project the lake information onto the target grid of an atmospheric model. The pseudo-periodical regime of the lake model was used to obtain the initial fields of lake variables.

## Introduction

As the resolution of present-day atmospheric models increases, more and more fine-scale effects in the atmosphere from the underlying surface become apparent and should be taken into account. The appropriate parameterization schemes must be developed and applied. Parameterization of lakes as a specific type of underlying surface is among them. There are different aspects of the interaction between the lake surface and the atmosphere (Mironov 2008, Eerola *et al.* 2010).

As a rule, the parameterization of any type of the underlying surface includes: (i) a model or a method to simulate the physical processes, and (ii) a set of external parameters to specify and to map the physical properties of this surface type.

These two aspects are closely connected. For the parameterization scheme, we choose a model or a method, and the model (or method) operates with the specific external parameters.

The sensitivity of the parameterization scheme and hence of the atmospheric model is different to various external parameters. The most important surface characteristics should be specified properly to provide correct simulations. These characteristics should be specified in every grid box of an atmospheric model grid for every surface type. In case of the mosaic tiling approach, when the grid square is divided into several different surface types, we also should determine the fraction of every surface type in every grid box of an atmospheric model grid. These fields form the external parameter (physiographic) data set.

There is an additional aspect of the problem for operational NWP modeling: the need for initial fields for every time-dependent variable. Even if the surface-related data assimilation scheme is applied, we need data at least for the first (cold) start of the model.

All the issues mentioned above are relevant for the parameterization of lakes. The main aim of this study is to take some steps forward to facilitate progress in the second aspect of lake-surface parameterization, namely the development of the external lake parameter data set.

The main challenge with regard to the parameterization of lakes for atmospheric modeling is the necessity to consider all lakes in the atmospheric model domain together. The atmospheric model domain typically covers an extended area of many thousands of square kilometers. Different types of lakes (large, medium and small in size, deep and shallow) with different behavior can be located at the territory which all should be described together. There is a particular difficulty in providing the realistic values for external model parameters for all lakes within the atmospheric model domain.

To provide external lake-model parameters for different lakes we need specific measurements. Probably we also can estimate the necessary parameters indirectly (e.g. using geological information). It depends on practical and economical circumstances whether the measurements have actually been performed (and are available) or not. Different kinds of lake models can be used as a lake-surface parameterization scheme for an atmospheric model. We can use lake models that range from very sophisticated 3-dimensional (3-D) to bulk 0-dimensional (0-D) schemes. For different kinds of lake models, we must provide different sets of external parameters.

3-D lake models have been developed for lakes of any size, but are usually applied to large lakes (e.g. León *et al.* 2007, Long *et al.* 2007). These models need data on bathymetry and may even need the 3-D fields of water turbidity parameters. Measurements like these have been performed only for a few, well-studied lakes. Large and/or economically important lakes like Lake Victoria or Lake Ladoga are usually well-studied. There is no lack of data for them, at least for bathymetry.

One-dimensional (1-D) or 0-D models (e.g. Hostetler 1993, Blenckner *et al.* 2002, Mironov 2008) have been developed mainly for medium- and small-sized lakes. Such models need much fewer external data, but these are studied much less than large lakes and measurements for the lakes in this size range are rarely available. Besides, there are plenty of small- and medium-sized lakes, which makes the problem of external data much more pronounced.

For global atmospheric models, any physiographic data set should naturally be global. As a rule, fine-scale atmospheric models are regional, but they need to be applicable over any area of the globe. This means that ideally they need a global physiographic data set as well. This also applies to the data set for the lake parameters. So, our ambition and longer-term goal, is to obtain a single universal and global set of the parameters of lakes to be utilized in atmospheric modeling.

There are however several questions that should be answered. First, which lake parameters should be included into the lake data set, and which lake parameters are essential from the atmospheric modeling point of view? Second, there is a limited amount of direct measurements which are not enough to cover all lakes on the globe. Moreover, data are spread over different institutions in various states: water cadastres, national databases, etc. and should be collected. Can we obtain the estimates of lake parameters indirectly using geological information, or some other information? The next question is how to organize the data? Should we represent different sized lakes (large, medium and small) in the same way? How the lake data should be mapped, in other words, how we combine the lake parameters with the physiographic data sets commonly used in NWP and climate modeling? We need to maintain the final product, probably by adding new information. What data structure should we have to make this procedure easy? The last question is technical, related to data formats, programming tools for the code development, etc. This aspect may also appear to be important.

The development of an external parameter data set for the parameterization of lakes is a big task, it requires plenty of technical work. The first steps towards this goal were performed within

the framework of the INTAS innovation project 05-1000007-431 and in cooperation with Consortium for Small-scale Modeling (COSMO). As a result of these projects, an external parameter data set for lakes and mapping tools were developed. This data set was designed for the COSMO NWP model (Doms and Schattler 1997) with the Freshwater Lake model (FLake) (Mironov *et al.* 2007, Mironov 2008) used as a lake-surface parameterization. The experiments were also performed with the High Resolution Limited Area Model (HIRLAM) (Undén *et al.* 2002) and with some other models (Samuelsson *et al.* 2010). The experience obtained is presented below. In addition, a system generating initial fields for time-dependent lake variables was developed.

## Methods

### Lake parameters needed for an atmospheric model

An atmospheric model is usually interfaced with the underlying surface scheme in the following way: the atmospheric model provides the atmospheric state and needs the surface fluxes which are then used as boundary conditions for radiation or turbulence scheme. These fluxes can be computed by the atmospheric model using surface temperature and moisture provided by the surface model (including the lake model). Consequently, the surface temperature produced by the lake model is of special interest, being an interface between the lake model and the atmospheric model.

To find the external parameter(s) which have the strongest influence on the simulated lake surface temperature, sensitivity tests with the lake model FLake were carried out. The FLake model has been developed primarily for medium- and small-sized lakes, and atmospheric forcing for the sensitivity tests was provided by meteorological data from an observation campaign for Lake Erken, a medium-sized lake. Lake Erken is located in Sweden, to the north of Stockholm. It is more or less a typical, boreal lake, representative of the Scandinavian climate. Observations were available from May 1989 to October 1990, so the

time scale of one year was considered. The main FLake parameters to test were the mean depth of the lake and the optical properties (extinction coefficient). The results are presented and discussed in detail in Kourzeneva and Braslavsky (2005). The mean depth of a lake turned out to be the crucial parameter in the studied context, while the optical parameters of the water appeared to be less important. This conclusion is only reasonable for the conditions and time scale considered. In principle, for other geographical conditions, other time scales or considering other simulated lake characteristics like the lake bottom temperature, the conclusions may differ.

The main lake parameter therefore needed by the lake parameterization scheme in atmospheric models is the lake depth. For the FLake-based parameterization, it is the mean depth of a lake. It is reasonable to apply the 0-D lake model FLake for small- and medium-sized lakes, and it is natural to characterize them by the mean lake depth. Concerning large-sized lakes, they can be hardly characterized by their mean depth because of a possible significant variation in bathymetry. The model errors of 0-D or 1-D lake models applied for a large-sized lake may also be significant. But in this study we used the mean lake depth to characterize all lakes, even large-sized ones.

The lake fraction is another lake-related parameter necessary for atmospheric modeling. The lake fraction is the part of an atmospheric model grid box covered with lakes. Naturally, we need this if the tiling approach is used, but even without tiling, we use the lake fraction to construct the lake mask.

### Sources of lake-related information

Here, we rely on the mean lake depth data from hydrological institutions of different European countries. For lakes outside Europe, we use information of the International Lake Environment Committee. This information comes mainly from direct hydrological measurements and hydrological soundings, except for the data for Sweden, as it includes also rough estimations from geological conditions.

The data for Europe were collected from national lake databases and from water cadas-

tres of different European countries, including Norway, Sweden, Finland, Russia (and former Soviet Union), Poland, Germany, Austria, and Switzerland. Different organizations kindly provided the data, mainly through personal communication. In contrast to Europe, data for the rest of the world are much poorer.

The data were preprocessed, e.g. transferred from national coordinate systems to geographical coordinates. Many errors were corrected. Sometimes there were gaps in the data, e.g. information about mean lake depth was missing. For such lakes, a default value of 10 m was assigned by the mapping software package (*see below*).

Finally, the Hydrological Lake Dataset was created. For the lakes included, it provides the following information: geographical coordinates (of a point on the water surface), the mean depth of a lake, the maximum depth of a lake and the lake area when available, the lake name and the country where the lake is located. Lakes with the surface area of more than 1 km<sup>2</sup> are included. The dataset is presented in an ASCII format, it contains the data and metadata (metadata indicate the origin and sources of data for every country). No specific database software tools were used.

We characterized the lake location by the coordinates of only one arbitrary point on its surface. The form of the lake is considered by the mapping algorithm, which processes these data later (*see below*). Presently the Hydrological Lake Dataset comprises about 9500 lakes.

## Mapping the lake information

For mapping, a map *per se* is needed as well as a tool to combine the lake information with the map. The final purpose of mapping is to obtain a map with the lake depth specified. This information needs also to be projected onto an (arbitrary) target grid and a domain of an atmospheric model.

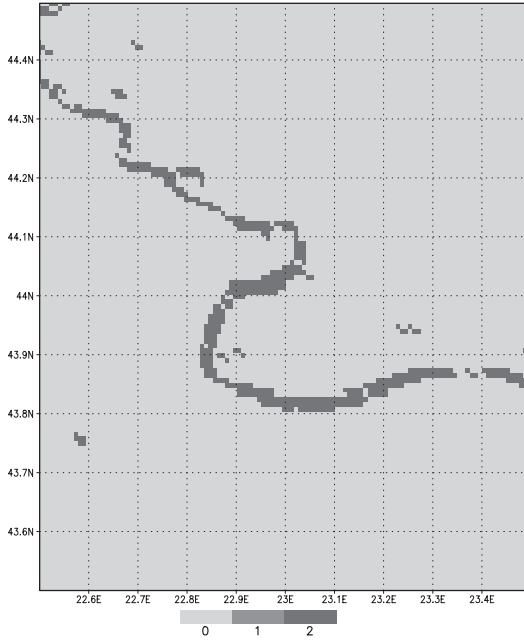
### Map — dataset for ecosystems

For this study it is natural to use for mapping such an ecosystem dataset which is already

widely used in the atmospheric modeling. We used the Global Land Cover Characteristics (GLCC) dataset of the U.S. Geological Survey. In principle, other datasets also may be used for mapping, e.g. the ECOCLIMAP dataset (*see Masson et al. 2003*). Usually the main source of information for the ecosystem datasets is space-born measurements. Also maps and atlases, including those which have been digitized, are used. Space-born measurements are processed automatically using statistical methods. Hence, there are errors and uncertainties in the final products. The dataset for ecosystems provides a map, where every pixel is classified according to its ecosystem. The resolution (size of the pixel) of GLCC dataset is 1 km, the coverage being global. To identify lakes, we used the ecosystem denoted as “inland water”. This was not entirely correct, as not only lakes, but also rivers are referred as “inland water” in GLCC. The same problem appeared with inland seas (like the Black Sea), and fjords, but these mistakes were corrected automatically by the mapping software package developed (*see below*).

We assumed that a lake on the map is a set of conterminal pixels with the “inland water” ecosystem type. In other words, the conterminal “inland water” pixels form one lake on the map which we call a “spot-lake” in order to distinguish it from real lakes and from lakes listed in the Hydrological Lake Dataset. Note that here we consider a map, and not yet the atmospheric model grid. The map represents the highest level of the horizontal resolution available. We actually combine the contaminant pixels into one large lake on the grid with the highest available resolution.

There are millions of “spot-lakes” in the global ecosystems dataset. Most of them correspond to real lakes. But rivers show up as chains of lakes (Fig. 1). Then, the erroneous small lakes (artifacts) may exist or some real small lakes may erroneously disappear, and the coastline may be erroneous. The situation may be illustrated by the comparison of the maps from two ecosystems datasets, GLCC and ECOCLIMAP. The versions of a map from GLCC and from ECOCLIMAP datasets for the Karelian Isthmus region are presented (Figs. 2–3). The coastline of Lake Ladoga differs in the two datasets, as



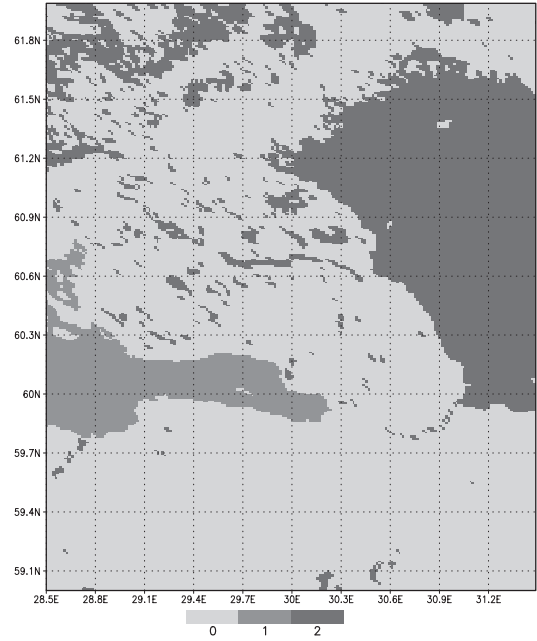
**Fig. 1.** GLCC dataset, the Danube region (0 = land, 1 = sea, 2 = inland waters).

well as the size and the form of small lakes and islands. Some small lakes exist on the one version of a map and do not exist on the other.

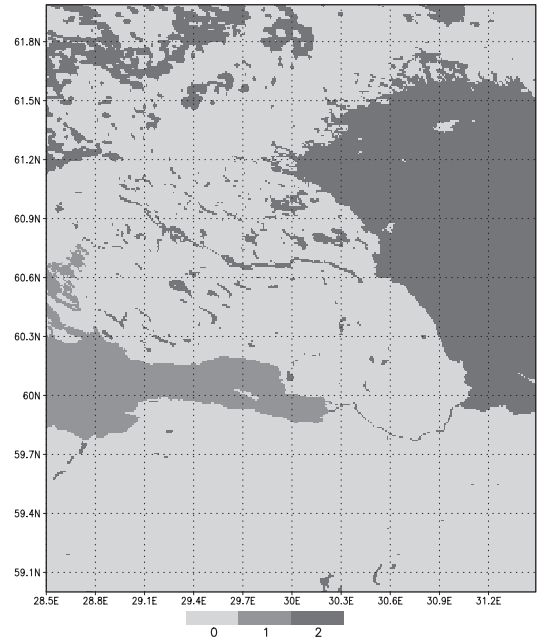
### Combining the Hydrological Lake Dataset with the map

Combining information from two datasets means that we should find the correspondence between lakes listed in the Hydrological Lake Dataset and “spot-lakes” from the ecosystem dataset. In practice, the problem has two aspects: the methodological aspect and that connected with code development.

The basic idea for choosing an appropriate methodology was that both the Hydrological Lake Dataset and the ecosystems dataset contain random errors. Hence, it was natural to use a probabilistic approach. In the present study, the probabilistic approach was used only intuitively, without strict quantitative justification. The procedure was reduced to scanning for some pixels on the map of the ecosystems dataset around the pixel which corresponds precisely to the coordinates of the lake from the Hydrological Lake

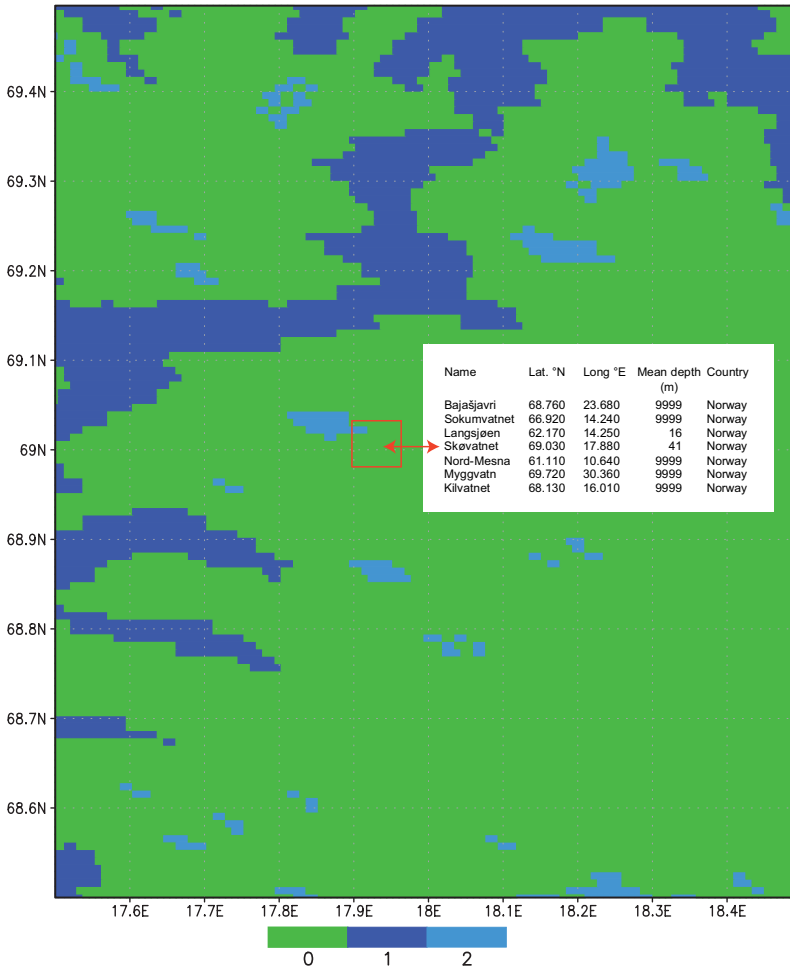


**Fig. 2.** GLCC dataset, the Karelian Isthmus region with Lake Ladoga (0 = land, 1 = sea, 2 = inland waters).



**Fig. 3.** ECOCLIMAP dataset, the Karelian Isthmus region with Lake Ladoga (0 = land, 1 = sea, 2 = inland waters).

Dataset. If at least one pixel in the scanned area belonged to some “spot-lake”, the correspondence between this “spot-lake” and this lake from



**Fig. 4.** The map fragment and the fragment of the Hydrological Lake Dataset. Several bits of the map within the scanned area (the red square) belong to a “spot-lake”, so the correspondence between this “spot-lake” and Lake Skøvatnet from the Hydrological Lake Dataset is established. After the analysis of the correspondences, the entire “spot lake” (not only “inland-water” pixels in the scanned area) will probably receive the mean lake depth of 41.0 m.

the Hydrological Lake Dataset was established (Fig. 4).

Furthermore, random errors in both datasets mean that, the task of finding correspondences may have no solution or may have more than one solution. It may happen that no “spot-lake” corresponding to the lake from the Hydrological Lake Dataset (even with the scanning procedure) is found. In practice, as we are interested in finding the depth data for “spot-lakes”, this case is not a problem. It may also appear that more than one lake from the Hydrological Lake Dataset refers to some “spot-lake” (especially considering the scanning procedure). In this case, information on lake depth from the Hydrological Lake Dataset was averaged. The averaging included weighting in proportion to appropriate lake areas. The highest priority was given to the lake which corresponded to the “spot-lake” without the scan-

ning procedure. It may appear that no lake from the Hydrological Lake Dataset refers to some “spot-lake” (this “spot-lake” is not recognized). This case is the most frequent. Such “spot-lakes” received the default depth (10 m at present).

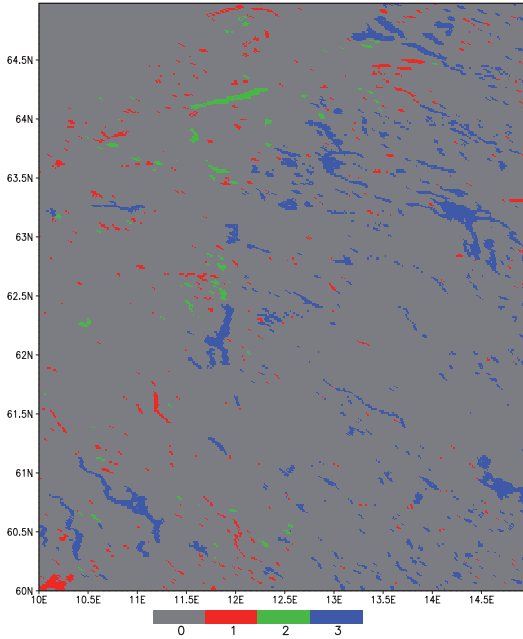
The code development for the mapping software package is the task which specialists in atmospheric modeling are rarely faced with. So, the task becomes non-standard, although it is basically a technical problem. In this study, FORTRAN90 was used for the code development. The main requirement for the software package is flexibility: the final product should be easily updated when new lakes are added to the Hydrological Lake Dataset. It was initially intended to be used by regional atmospheric models with a domain covering Europe. The algorithm for the interface software package is the following:

- Step 1 We number all the “spot-lakes” on the map, so every “spot-lake” receives an identification number (ID). This step is the most difficult in the code development. The specific programming recipes rarely used in the domain of atmospheric modeling are necessary here. In FORTRAN90, one can use either derived data types with self-pointers (a chain of pointers) or recursive functions. Here, we use the chain of pointers method (the explanation can be found e.g. in Barteniev 2000). As a result, we obtain IDs of the “spot-lakes” in every pixel of the map.
- Step 2 We establish links (correspondences) between the lakes from the Hydrological Lake Dataset and the “spot-lakes” identified by the IDs. For each lake from the Hydrological Lake Dataset, we find the appropriate pixel on the map using the coordinates of a point on its surface. If it is the pixel with non-zero “spot-lake” ID, we establish a link between this “spot-lake” and this lake from the Hydrological Lake Dataset. The scanning procedure is used at this step. As a result, for every “spot-lake” marked with the ID we obtain the link (or links) for the lake(s) from the Hydrological Lake Dataset. If no lake from the Hydrological Lake Dataset refers to the “spot-lake” in question, it means that this “spot-lake” is not recognized, and the empty link is established. So, every “spot-lake” marked with its ID receives at least one link, empty or not.
- Step 3 We analyze links for every “spot-lake” marked with its ID. If there is only one link to some lake from the Hydrological Lake Dataset, we use the mean lake depth value of this lake. If there is more than one link, we average the appropriate data (see above). If the link is empty, we use the default value for the mean lake depth. Then we set the mean lake depth value for every pixel of this “spot-lake”. As a result, we obtain the mean lake depth values for every pixel of the map. In the other words, we construct a “mean lake depth field” on the very fine 1-km grid.

The efficiency of the described mapping technology is illustrated in Figs. 5–6. Blue pixels are “spot-lakes” linked to the lakes from the Hydrological Lake Dataset. These “spot-lakes” received the real mean lake depth from there. Green pixels are “spot-lakes” linked to the lakes from the Hydrological Lake Dataset but the mean lake depth information was missing there. These “spot-lakes” were recognized, but they received the default value for the mean lake depth. Red pixels are zero-linked (not recognized) “spot-lakes”. Two examples are given, for the region in Norway and for the region in northern Russia. The efficiency of the mapping technology is quite high, but the result, of course, depends on data availability in the Hydrological Lake Dataset. Norway is the region with a great deal of lake data. Therefore, many lakes were recognized and received a real value of the mean lake depth (or the default value of the mean lake depth in the case of missing data in the Hydrological Lake Dataset). For northern Russia, the situation is less encouraging. Some large-sized lakes received a real value of the mean lake depth. However, for this region there is no information on small-sized lakes and even on one large-sized lake in the Hydrological Lake Dataset, hence, that those lakes were not even recognized.

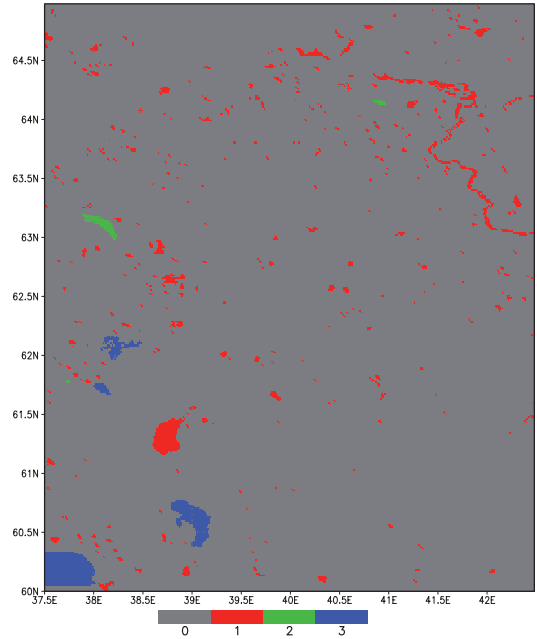
### Projecting the information onto a target grid and an atmospheric model domain

It is important that lakes which differ in surface area, mean depth — and hence in behavior — sometimes are located in one grid box of an atmospheric model grid. So, averaging the mean lake depth data is not desirable. But we can aggregate this information using a statistical approach, for example, using empirical Probability Density Functions (PDFs). An additional benefit from the statistical approach is a reduction of some random errors and uncertainties. Usually the resolution of an atmospheric model (target) grid is lower than the resolution of the map. We can, therefore, calculate the fractions of the lakes with the different depth values in every grid box of an atmospheric model grid and to make a histogram. These fractions can be interpreted



**Fig. 5.** Efficiency of the mapping technology: the region in Norway. Grey: land or sea area. Blue: “spot lakes” which received the real values for the mean lake depth from the Hydrological Lake Dataset. Green: “spot lakes” linked to the lakes from the Hydrological Lake Dataset but with missing information there. Red: not recognized “spot lakes”. “Spot lakes” marked with green and red received the default value for the mean lake depth.

as the lake-depth empirical PDFs. To construct the empirical PDFs we used the following depth classes (m): 0.0, 2.0, 4.0, 6.0, 8.0, 12.0, 16.0, 20.0, 24.0, 30.0, 36.0, and 42.0. For the grid box in question, the empirical Probability Density in each class is the ratio of the number of pixels with the lake depth in the appropriate class to the total number of pixels in this grid box. Four examples of the empirical PDFs for four specific grid boxes of the HIRLAM model (target) grid are presented in Fig. 7. The resolution of the atmospheric model grid is approximately 11 km, and the domain includes southern Finland and Karelia. The empirical PDFs in different grid boxes differ significantly. For example, in the grid box with the empirical PDF presented in Fig. 7a, 45% of the pixels have the lake depth values between 4.0 m and 6.0 m and 10% of the pixels have the lake depth values between 6.0 m and 8.0 m; other depth classes are absent, so in this grid box there are no lakes with other depths.



**Fig. 6.** Efficiency of the mapping technology: the region in northern Russia. Colour codes as in Fig. 5.

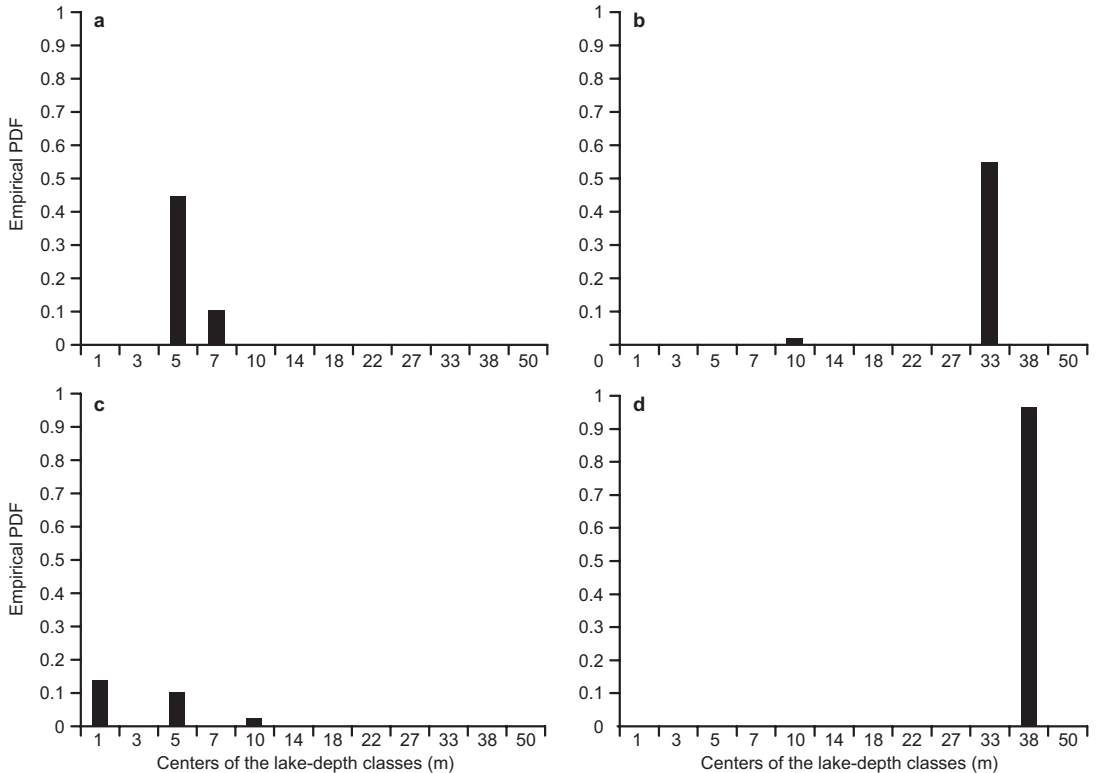
In the grid box shown in Fig. 7b, 4% of the pixels have lake depth values between 8.0 m and 12.0 m and 55% of pixels have lake depth values between 30.0 m and 36.0 m; lakes with the other depth values are absent. In the grid box depicted in Fig. 7c, there are only a few shallow lakes, and the grid box in Fig. 7d is almost totally covered by one deep lake. It is an open question as to which statistic should be used in order to prescribe the lake depth for the model grid box in question. We used the mode statistic (the value where the histogram reaches its highest peak and the empirical PDF attains its maximum value, in the other words, the most probable lake depth).

The lake fraction in every grid box of a target grid is calculated in a standard way using the information from the map.

## Results and discussion

The examples of the fields of mean lake depth and lake fraction are shown in Figs. 8–9, they were obtained using the aforementioned procedure. The mean lake depth has been mapped and projected onto the grid of the HIRLAM





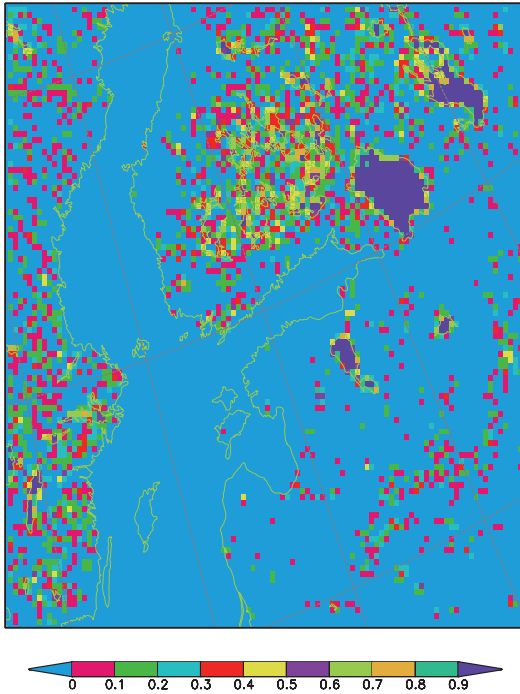
**Fig. 7.** The empirical Probability Density Functions (PDFs) for four specific grid boxes of the HIRLAM model grid. The resolution is approximately 11 km, domain includes southern Finland and Karelia.

model, rotated spherical coordinates are used. The domain covers Karelia with lakes Ladoga and Onega, and the southern part of Finland with hundreds of small-sized lakes. The model grid resolution is approximately 11 km. In HIRLAM, the mosaic tiling approach is used for the surface parameterization scheme, so that the grid square is divided into five surface types (tiles): water, ice, bare land, low vegetation and forest. This means that all lakes, even when the lake fraction is small, will be considered. The presented fields can be used by the atmospheric model (in this example case by HIRLAM) as the external parameters for the lake-surface parameterization based on the lake model FLake.

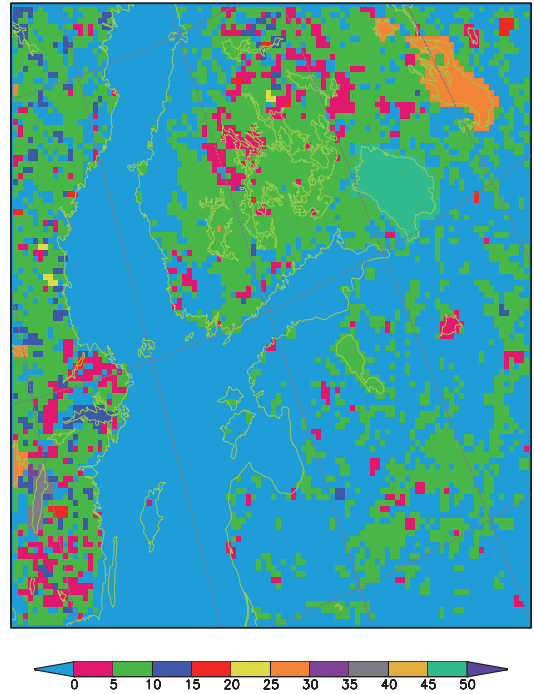
The presence of real information (in contrast to the default values) in the produced mean lake depth field is dependent on the completeness of the Hydrological Lake Dataset for the region in question. It was a substantial job to compile the Hydrological Lake Dataset for 9500 lakes. However, this information is far from sufficient,

as the amount of lakes in reality (and “spot-lakes” in the ecosystem dataset) is much larger. Although it is a difficult task to collect information (mainly from an organizational point of view), the next steps in this direction should be undertaken by a wider consortium. For many lakes in hard-to-reach regions (e.g. for Siberia) there are no measurements at all. However, we should use all information available from existing measurements.

It is rather risky to characterize the location of a lake by the coordinates of only one point on its surface, as it is done in the Hydrological Lake Dataset. Although the mapping algorithm considers the form of a lake, it only works well if the map resolution (the map of the ecosystems dataset) is sufficient. One can imagine the case when the lake consists of two parts connected by a channel which is so narrow that it not seen on the map. In this case, the mapping algorithm will provide the lake depth for only one part of the lake and the second part won't be recognized. In



**Fig. 8.** Lake fraction (0–1) for HIRLAM grid, the resolution is approximately 11 km, domain includes southern Finland and Karelia (see text for details). In this domain approximately 70% of grid boxes with non-zero lake fraction received the lake mean depth from the Hydrological Lake Dataset.



**Fig. 9.** Mean lake depth in metres for HIRLAM grid, the resolution is approximately 11 km, domain includes southern Finland and Karelia (see text for details). In this domain approximately 70% of grid boxes with non-zero lake fraction received the lake mean depth from the Hydrological Lake Dataset.

practice, this kind of case is not frequent, so that the one km resolution of the map is sufficient for our task. The most important lakes with complicated forms, like Lake Balaton, do not break down into small pieces and hence the whole lake receives the mean depth.

The Hydrological Lake Dataset contains mainly information from the direct measurements. As the direct measurements are often not available, indirect estimates are highly desirable. We can probably estimate the lake depth from the geological properties of its location: usually in mountain areas lakes are deep and in flat areas they are shallow. So, the estimate can be carried out using the variations in orography.

In principle, other datasets for ecosystems apart from GLCC might be used for mapping. For example, the ECOCLIMAP dataset may be used as well. This dataset is specially developed for atmospheric modeling. The problem

will remain in distinguishing between rivers and lakes, and wetlands also may be a problem. The ECOCLIMAP map correctly specifies many wetland pixels which indicate swamps and marshes. However, the aggregation method applied to the ECOCLIMAP database reduces the wetland information to a certain water percentage in the atmospheric model grid box, and this water could be erroneously referred to as lakes. Probably, when applying the lake depth mapping algorithm to the ECOCLIMAP map, it is also better to modify slightly the ECOCLIMAP aggregation method. The mapping method based on the probabilistic approach gives fair results. But the strict mathematical justification of the method will allow us to make its logic more clear and probably to make the algorithm more efficient.

The lake depth values may be very different, and different lakes with different behaviors may be located in one grid box of an atmospheric

model grid. So we recommend the approach based on the empirical PDFs. If other lake water parameters, like the extinction coefficient, vary significantly, the same approach may be recommended as well.

At present, only the mean depth of a lake is considered and no bathymetry data is included even for large-size lakes. But it is possible to include bathymetry data for the lakes when this information is available. This may be done after mapping, just by replacing the mean lake depth information on the 1 km grid by bathymetry data for the lake in question. Using bathymetry will make it possible to distinguish the different behavior of different parts of a large-size lake according to depth applying a 0-D or a 1-D model for different lake grid-boxes with their individual mean depth. The bathymetry even may allow applying a 3-D lake model.

Technical questions are still open. We used traditional instruments for the domain of atmospheric modeling (mainly based on FORTRAN coding). No special database software tools or Geo Informational System (GIS) technologies were used. They might probably help to cope with some problems, or at least could provide some practical solutions for using the collected lake information for applications other than atmospheric modeling (if any).

### Initial fields for prognostic lake variables

As it was already mentioned, in NWP we also need data to initialize all prognostic lake variables, which are known as “cold start” data. For instance, the prognostic variables for the lake model FLake are: mixed-layer temperature, mean water temperature, mixed-layer depth, bottom temperature, shape factor, etc. So, we need the “climatological mean” annual cycle of these variables for the lakes with different properties in different geographical locations.

In this study, the pseudo-periodic regime approach was used to develop such a dataset. To achieve this, first, we start the year-long lake model run with arbitrary initial values of prognostic variables of the lake model. We force the lake model by a given annual cycle of meteoro-

logical input (temperature, humidity, wind speed in the surface layer and radiation fluxes). After that, we start the next year-long lake model run with the initial values of the lake prognostic variables specified from the previous year-long lake model run. We force the lake model again with the same meteorological input. We repeat this procedure several times, and after a few model years (iterations) it is expected that running the model for one year more will not change the annual cycle of the lake-model variables. In the other words, it is expected that after a while the lake model will obtain a pseudo-periodic regime (a “perpetual year” solution). The model annual cycle will be the same for the two neighboring model years (if the iterative method converges). The annual cycle of the lake prognostic variables on the last iteration may be considered as a “pseudo-climate” of a given lake.

In our study, we used atmospheric forcing data (long-term monthly mean meteorological values) from the National Center of Environmental Prediction (NCEP) Reanalysis Project (Kalnay *et al.* 1996). A linear interpolation was carried out to obtain a 20-min temporal resolution. This procedure is somewhat questionable, as we suppress the daily cycle, but the average energy balance for a day is described correctly. The NCEP reanalysis grid is, in principle, more coarse than the grid of any regional NWP model. For every grid box of the NCEP reanalysis grid, pseudo-periodic solutions for several lakes that differ in terms of mean depth (in the depth range from 1 to 50 m) were obtained. The result is the “climatological mean” annual cycle of the lake prognostic variables for every grid box of the NCEP reanalysis grid for lakes with different mean depths (for an illustration, *see* Kourzeneva *et al.* 2008). This information was then interpolated to a finer regional atmospheric model grid considering the mean lake depth field.

The shortcoming of the pseudo-periodic solutions is the poor mathematical justification. Sometimes the lake model does not reach the pseudo-periodic regime, and the iterative method does not converge. It would be probably better to use the model lake climate values in usual sense instead of the model lake “pseudo-climate” values.

## Concluding remarks

One of the basic issues of lake parameterizations in atmospheric modeling, both for NWP and climate simulations, is the need for external lake parameters. The lake depth is the most important parameter, and this is the minimum required for all lake models. In some situations, the optical parameters may be important as well. In this paper, the experience of the development of the external lake parameters data set for a regional atmospheric model was described. We focused on the mean lake depth as the main lake parameter.

Despite the complexity of the problem, direct measurements of the lake depth could be collected and used. When bathymetry information is available it should be used as well. When the direct measurements are absent, the indirect estimates of lake depth such as geological information or some other approach could be used.

For mapping, it is natural to use the ecosystem dataset which is already widely used in atmospheric modeling. In principle, different ecosystems datasets may be used for mapping. The fact that none of them distinguish between lakes and rivers create some problems. An extra difficulty may be caused by wetlands. The mapping method based on the probabilistic approach proved to be efficient, and we recommend that it should be developed further. The use of the empirical PDFs for projecting the lake information onto the target grid of an atmospheric model is highly recommended. Additional software tools such as database management software or GIS might also be useful for mapping.

Also for NWP the initial fields for the prognostic lake variables, namely their “climatological mean” annual cycles, are necessary. This information may be obtained using the pseudo-periodical solutions of the lake model. The true lake model climatology is probably more appropriate.

We conclude that some useful steps forward were taken. But we still strive towards the proposed goal, which is the global universal database of lake parameters for the needs of atmospheric modeling, with as many lakes included as possible.

*Acknowledgements:* The author thanks Dr. Dmitrii Mironov from Deutscher Wetterdienst for inspiring and useful dis-

cussions (in fact, he was the project initiator), Ms. Gantuya Ganbat (Russian State Hydrometeorological University) and Dr. Laura Rontu (Finnish Meteorological Institute) for taking part in the software package development. The Hydrological Lake Dataset was made possible by people who kindly provided lake data, their names are listed in the data set metadata header. Three anonymous reviewers made useful comments which helped to improve much understanding and presentation of the results. Useful contacts with COSMO are gratefully acknowledged. The work was partially supported by the EU Commission through INTAS-05-1000007-431 project, the NetFAM/NordForsk (<http://netfam.fmi.fi>), and the Swedish Institute, project 01753/2004.

## References

- Barteniev O. [Барте́ньев О.] 2000. [*Modern FORTRAN*]. Dialog-MIFI, Moscow. [In Russian].
- Blenckner T., Omstedt A. & Rummukainen M. 2002. A Swedish case study of contemporary and possible future consequences of climate change on lake function. *Aquat. Sci.* 64: 171–184.
- Doms G. & Schattler U. 1997. *The nonhydrostatic limited-area model LM (Lokal-Modell) of DWD, part I: scientific documentation*. Deutscher Wetterdienst, Offenbach, Germany.
- Eerola K., Rontu L., Kourzeneva E. & Scherbak E. 2010. A study on lake temperature and ice cover in HIRLAM. *Boreal Env. Res.* 15: 130–142.
- Hostetler S., Bates G. & Giorgi F. 1993. Interactive coupling of a lake thermal model with a regional climate model. *J. Geophys. Res.* 98: 5045–5057.
- Kalnay E., Kanamitsu M., Kistler R., Collins W., Deaven D., Gandin L., Iredell M., Saha S., White G., Woolen J., Zhy Y., Chellian M., Ebisuzaki W., Higgins W., Janowiak J., Mo K.C., Ropelewski C., Wang J., Leetmaa A., Reynolds R., Jenne R. & Joseph D. 1996. The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteor. Soc.* 77: 437–471.
- Kourzeneva E. & Braslavsky D. 2005. Lake model FLake, coupling with atmospheric model: first steps. In: Undèn P. (ed.), *Fourth SRNWP/HIRLAM Workshop on Surface Processes and Assimilation of Surface Variables jointly with HIRLAM Workshop on Turbulence*, Workshop report, SMHI, Norrköping, Sweden, pp. 43–53.
- Kourzeneva E., Samuelsson P., Ganbat G. & Mironov D. 2008. Implementation of lake model FLake in HIRLAM. *HIRLAM Newsletter* 54: 54–61.
- León L., Lam D., Schertzer W., Swayne D. & Imberger J. 2007. Towards coupling a 3D hydrodynamic lake model with the Canadian Regional Climate Model: simulation on Great Slave Lake. *Environ. Modell. Softw.* 22: 787–796.
- Long Z., Perrie W., Gyakum J., Caya D. & Laprise R. 2007. Northern lake impacts on local seasonal climate. *J. Hydrometeorol.* 8: 881–896.
- Masson V., Clampeaux J.L., Chauvin F., Meriguet C. & Lacaze R. 2003. A global database of land surface

- parameters at 1-km resolution in meteorological and climate models. *J. Climate* 16: 1261–1282.
- Mironov D. 2008. *Parameterization of lakes in numerical weather prediction. Description of a lake model*. COSMO Technical Report 11. Deutscher Wetterdienst, Offenbach am Main, Germany.
- Mironov D., Heise E., Kourzeneva E., Ritter B. & Schneider N. 2007. Parameterization of lakes in numerical weather prediction and climate models. In: Umlauf L. & Kirillin G. (eds.), *Proceedings of the 11th Workshop on Physical Processes in Natural Waters*, Berichte des IGB vol. 25, Berlin, Germany, pp. 101–108.
- Samuelsson P., Kourzeneva E. & Mironov D. 2010. The impact of lakes on the European climate as simulated by a regional climate model. *Boreal Env. Res.* 15: 113–129.
- Undén P., Rontu L., Järvinen H., Lynch P., Calvo J., Cats G., Cuxart J., Eerola K., Fortelius C., Garcia-Moya J.A., Jones C., Lenderlink G., McDonald A., McGrath R., Navascues B., Nielsen N.W., Ødergaard V., Rodrigues E., Rummukainen M., Rõõm R., Sattler K., Sass B.H., Savijärvi H., Schreier B.W., Sigg R. & Tijn A. 2002. *The HIRLAM-5 scientific documentation*. SMHI, Sweden.