



## Status and progress in global lake database developments

Olga Toptunova<sup>1</sup>, Margarita Choulga<sup>2</sup>, and Ekaterina Kurzeneva<sup>3</sup>

<sup>1</sup>Meteorological forecasts Department, Russian State Hydrometeorological University (RSHU)  
Malookhtinsky pr. 98, 195196, Saint-Petersburg, Russia

<sup>2</sup>Research Department, European Centre for Medium-range Weather Forecasts (ECMWF),  
Reading, RG2 9AX, UK

<sup>3</sup>Research Department, Finnish Meteorological Institute (FMI),  
00560, Helsinki, Finland

**Correspondence:** Olga Toptunova (olgakolp@yandex.ru)

Received: 14 February 2019 – Revised: 17 April 2019 – Accepted: 24 April 2019 – Published: 21 May 2019

**Abstract.** Lakes affect local weather and climate. This influence should be taken into account in NWP models through parameterization. For the atmospheric simulation, global coverage of lake depth data is essential. To provide such data Global Lake Database (GLDB) has been created. GLDB contains information about lake location (latitude, longitude), water surface area, and lake mean and max depths. The mean depth is provided as a gridded data set.

### 1 Introduction

According with the latest research in the world there are ~ 117 million lakes with an area of more than 0.002 km<sup>2</sup>. Totally, they occupy about 5 × 10<sup>6</sup> km<sup>2</sup>, which is about 3.7 % of the earth's surface (Verpoorter et al., 2014). Lakes affect local weather and climate (Eerola et al., 2014; Samuelsson et al., 2010). In addition, lakes can affect global climate (Bastviken et al., 2011; Raymond et al., 2013; Stepanenko et al., 2011).

Lakes influence the structure of the atmospheric boundary layer by affecting the surface fluxes, influence temperature, amount of precipitation, generate night convection and intensive thunderstorms and winter snowstorms, increase wind speed and change energy balance between atmosphere and surface.

To take into account lake influence in NWP models GLDBv1 has been developed (Kourzeneva et al., 2009, 2012). It contained in situ lake mean depth gridded data and includes slightly more than 13 000 lakes. However total number of lakes on the Earth's surface is much larger – around 117 million according to the latest estimates. To take some extra lakes into account GLDBv2 has been created. The new version contained in situ information about more than

13 500 lakes, and indirect estimates of the mean depth for the boreal zone lakes based on their geological origin (Choulga et al., 2014). GLDBv2 has been upgraded to GLDBv3 with indirect mean depth estimates for the rest of the world and with some depth corrections for unidentified rivers. GLDBv3 contains in situ information about 14 960 lakes and consists of several data sources: lists of lakes with in situ data, indirect depth estimates, global lake cover and digitized bathymetry. This third version of GLDB is a global lake depth data set with in situ and estimated values on the ~ 1 km grid.

### 2 Towards GLDB version 4

#### 2.1 Introduction of new data sources

The aim of the actual upgrade of GLDBv3 is to add new reliable in situ data from different sources. New data will be used to verify and update indirect depth estimates and to calculate and add new indirect estimates for the region where they were absent before. New in situ data is collected from different sources: Limnology institute global database (St.-Petersburg, Russia), Global Reservoir and Dam Database (GRanD), national databases with open access.

Only natural lakes can be used for depth estimates based on geological origin of lakes. That is why Limnology institute global database was especially important. It has vast dataset (~ 58 000 records) with mainly natural lakes. Global Reservoir and Dam Database (GRanD, ~ 14 500 records) has only man-made lakes and reservoirs, which are locally managed, and should be treated separately in NWP models. In addition, national databases, articles and other scientific or semi-scientific open sources have been checked in order to verify or complete lake depth information in GLDB.

## 2.2 Cross-check of the list of lakes

Data cross-checking is extremely painstaking and time-consuming process, that has to be done in order to upgrade GLDB with reliable in situ data. It is acceptable that all sources of in situ data may have inaccuracies – limitation of measuring instrument. Random errors and systematic outliers should be eliminated from new data. A special semi-automatic procedure is developed for data cross-checking to either add or reject new data to GLDB. Preliminary random check of GRanD and Limnology institute datasets showed inaccuracies in such cases:

- Coordinates – location error: incorrect conversion, sign, decimal separation point.
- Water surface area/volume – measurement unit error.
- Mean depth data – incorrect lake depth units or instead of mean lake depth is presented max lake depth (or family lake mean depth).
- Duplicates – same lake is mentioned several times with different information. For example, Chinese reservoir has been mentioned three times with depths varying from 10 to 20 m in Global Reservoir and Dam Database (GRanD) (Fig. 1).

Indirect depth estimates from GLDBv3 have been verified against newly collected in situ data. Information about 533 of the lakes all over the globe have been used. Although RMSE is less for indirect estimates than for default depths of 10 m, BIAS has the same absolute value and is less than 1 m. Table 1 shows importance of natural lakes and reservoirs distinguishing. This shows that indirect depth estimates based on geological origin of lakes should be used only for natural lakes (not man-made ones!) (Table 1).

## 2.3 Comparison with ECOCLIMAP lake coverage

In total, more than 3000 in situ lake depths all over the globe have been added. But only 17 % of them were found on the global ecosystem map ECOCLIMAP2 (Champeaux et al., 2004) that is used as lake cover in GLDBv3. Almost half of unfound lakes have area less than 1 km<sup>2</sup>. Though 2.5 % of unfound lakes are larger than 50 km<sup>2</sup>.



**Figure 1.** Duplicate example in Global Reservoir and Dam Database (GRanD). (Google Earth, <https://earth.google.com/web/>, last access: May 2019).

For example, Burullus el Nahr lagoon (31°29' N 30°52' E) in Egypt with surface area 566 km<sup>2</sup> is brackish shallow waterbody. On the ECOCLIMAP2 it is presented as a part of Mediterranean Sea (Fig. 2).

Oder Bay Lagoon (53°48'16" N, 14°8'25" E) on the border of Germany and Poland with water surface area of almost 700 km<sup>2</sup>, has mean depth less than 4 m. So, on the lake cover the lagoon is presented as a part of Baltic Sea (Fig. 3). It should be kept in mind that sometimes inland coastal water bodies get be merged with ocean waters.

For man-made lakes it should be taken into account that new reservoirs emerge all the time all over the globe. For example, Indira Sagar Reservoir (22°17'01" N, 76°28'28" E) in India with water surface area of more than 900 km<sup>2</sup> and mean depth around 13 m was built in 2005 (Fig. 4). This reservoir is omitted on ECOCLIMAP2, which is based on 1999–2003 satellite data (Fig. 4b).

The last example is Egyptian saline lake Mariout (31°9'11" N, 29°53'55" E) with surface area of less than 65 km<sup>2</sup> and mean depth 1 m (Fig. 5). It is completely missing from the lake cover due to ECOCLIMAP2 algorithm has not been recognized heavy eutrophication water.

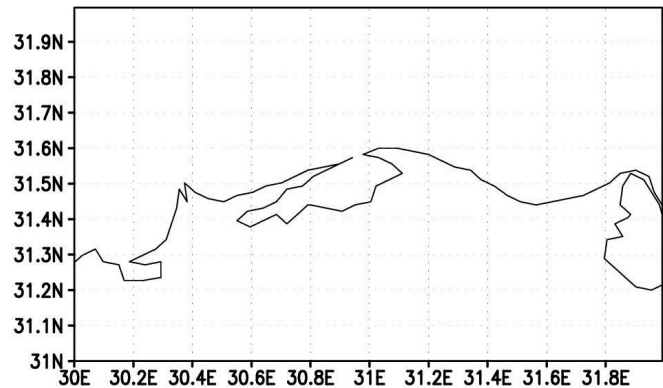
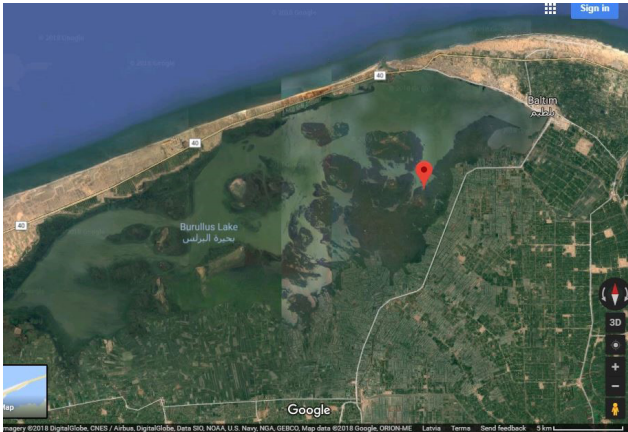
## 3 Conclusions

GLDB quality is determined by its major information sources – in situ measurements (are used directly and for indirect depth estimates). Several weather centers, like ECMWF, HIRLAM and COSMO, already use GLDB for their research and operative issues, so it is very important to maintain dataset quality on the same level or higher, so all new in situ data have to be carefully checked in advance.

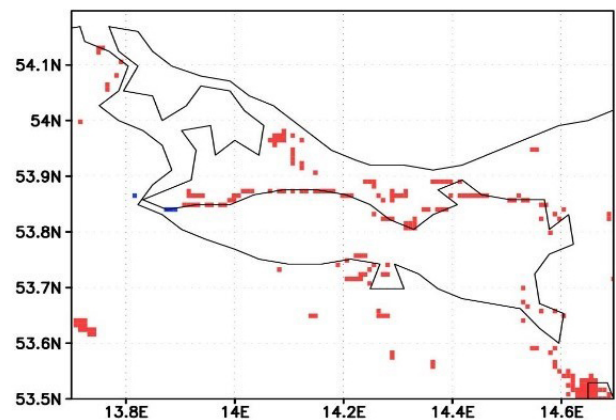
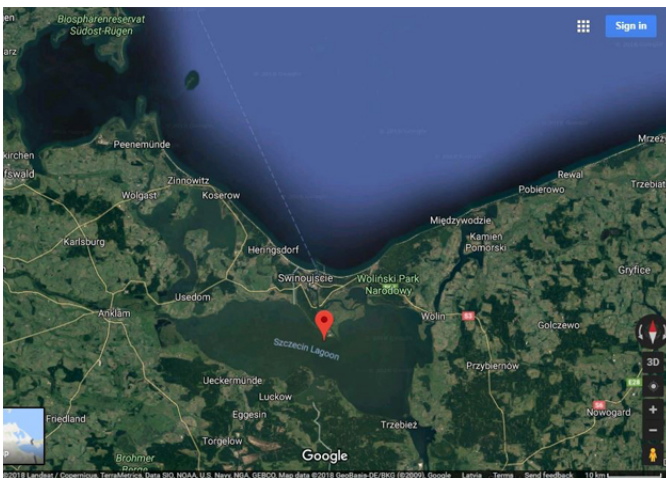
New in situ data have been collected from major global sources, where unfortunately preliminary random check identified some significant errors in location, measure-

**Table 1.** Verification of indirect estimates against new lake depth observations.

Sample size	533		339		194	
	natural lakes and reservoirs		natural lakes only		reservoirs only	
	GLDBv3	Default 10 m	GLDBv3	Default 10 m	GLDBv3	Default 10 m
BIAS	-0.9	0.9	0.7	3.4	-3.8	-3.3
RMSE	11.6	13.6	11.1	13.9	12.4	13.0



**Figure 2.** Burullus el Nahr Lagoon on Google Earth (to the left) (Google Earth, <https://earth.google.com/web/>, 2019) and on ECOCLIMAP2 (to the right).



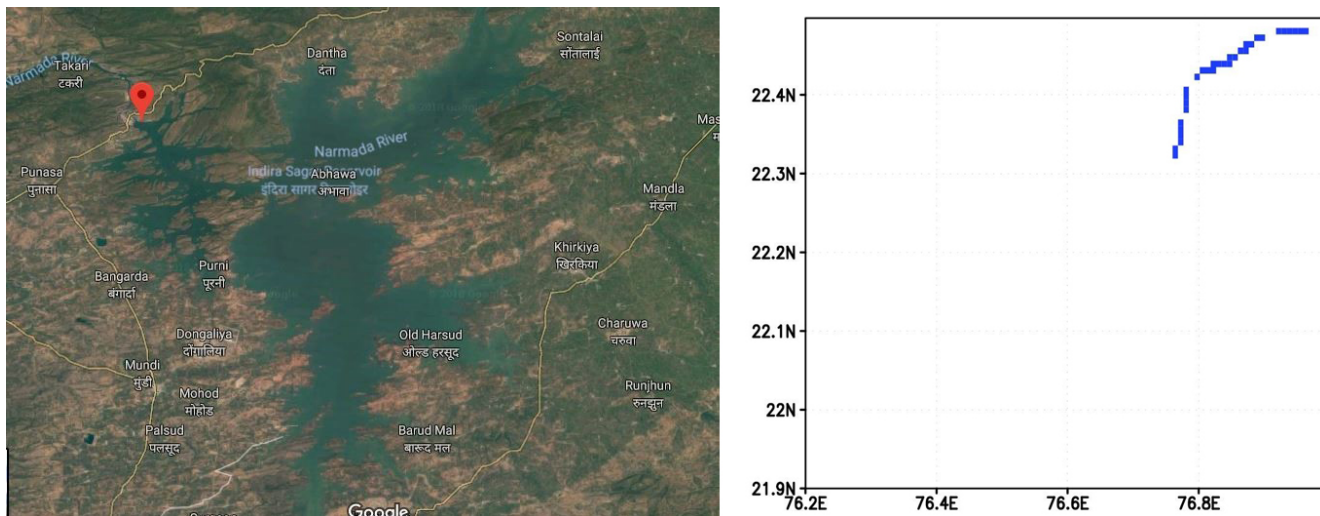
**Figure 3.** Oder Bay Lagoon on Google Earth (to the left) (Google Earth, <https://earth.google.com/web/>, 2019) and on ECOCLIMAP2 (to the right).

ment units and even lake identification. In total more than 3000 new in situ lake depths have been added to Global Lake Database GLDB. Dataset quality control is very important and all new in situ data has to be carefully checked in advance, because GLDB is already used by several global weather centers (e.g. ECMWF) and limited-area modelling consortia (e.g. HIRLAM and COSMO) for research and operative issues.

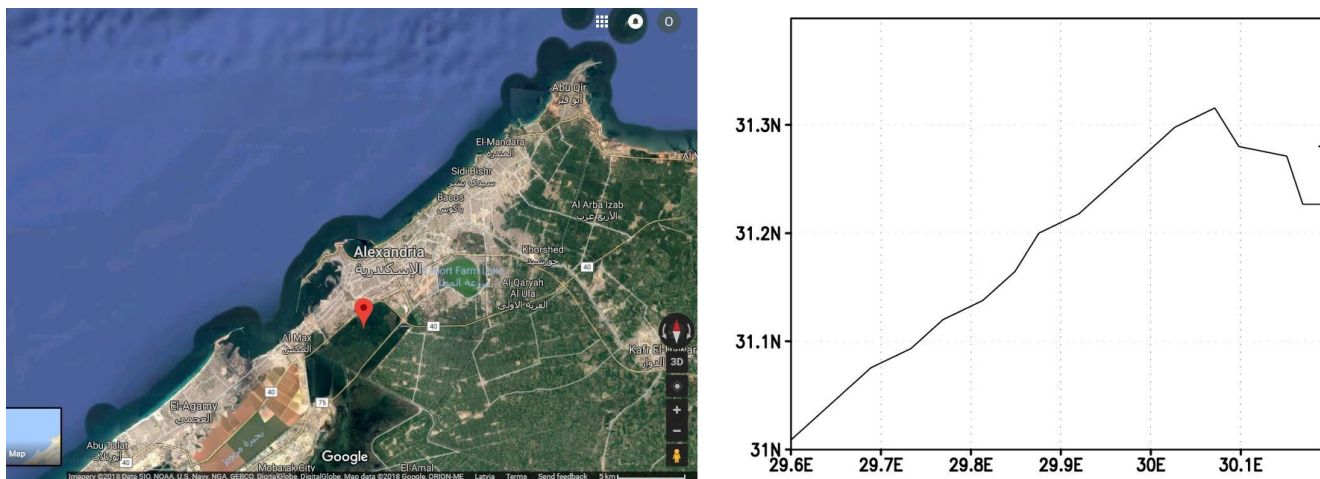
Over 83 % of newly added data have not been found on global ecosystem map ECOCLIMAP2. Main reasons for data set mistakes are:

- inland coastal waters are merged with ocean,
- map ECOCLIMAP2 does not contain all lakes amount,
- inaccuracies in water detection lake cover algorithm,





**Figure 4.** Indira Sagar Reservoir on Google Earth (to the left) (Google Earth, <https://earth.google.com/web/>, 2019) and on ECOCLIMAP2 (to the right).



**Figure 5.** Lake Mariout on Google Earth (to the left) (Google Earth, <https://earth.google.com/web/>, 2019) and on ECOCLIMAP2 (to the right).

- lake is simply too small for the horizontal resolution, realized in ECOCLIMAP2.

In the future newly added in situ data will be used for verification and upgrade of indirect depth estimates.

In future, it is planned to increase GLDBs horizontal resolution. To solve these problems with unfound lakes it is supposed to use continuous depth fields concept. In this case, it will be possible to use a typical mean depth value for the region.

**Data availability.** GLDB dataset and its full technical documentation can be found here: <http://www.flake.igb-berlin.de/ep-data.shtml> (GLDB, 2019). GLDB is available under Creative Commons license with Attribution (CC-BY).

**Author contributions.** OT composed this article and added new data for verification, MC analysed the results, EK provided guidance.

**Competing interests.** The authors declare that they have no conflict of interest.

**Special issue statement.** This article is part of the special issue “18th EMS Annual Meeting: European Conference for Applied Meteorology and Climatology 2018”. It is a result of the EMS Annual Meeting: European Conference for Applied Meteorology and Climatology 2018, Budapest, Hungary, 3–7 September 2018.

**Acknowledgements.** Thanks are extended to EUMETNET as funder of the work.

**Review statement.** This paper was edited by Emily Gleeson and reviewed by two anonymous referees.

## References

- Bastviken, D., Tranvik, L. J., Downing, J. A., Crill, P. M., and Enrich-Prast, A.: Freshwater methane emissions offset the continental carbon sink, *Science*, 331, 50 pp., <https://doi.org/10.1126/science.1196808>, 2011.
- Champeaux, J.-L., Han, K.-S., Arcos, D., Habets, F., and Masson, V.: Ecoclim2: a new approach at global and European scale for ecosystems mapping and associated surface parameters database using SPOT/VEGETATION data – First Results, *Int. Geosci. Remote Sens. Symp.*, 3, 2046–2049, 2004.
- Choulga, M., Kourzeneva, E., Zakharova, E., and Doganovsky, A.: Estimation of the mean depth of boreal lakes for use in numerical weather prediction and climate modelling, *Tellus A*, 66, 21295, <https://doi.org/10.3402/tellusa.v66.21295>, 2014.
- Eerola, K., Rontu, L., Kourzeneva, E., Kheyrollah Pour, H., and Duguay, C.: Impact of partly ice-free Lake Ladoga on temperature and cloudiness in an anticyclonic winter situation—a case study using a limited area model, *Tellus A*, 66, 23929, <https://doi.org/10.3402/tellusa.v66.23929>, 2014.
- GLDB: dataset and technical documentation, available at: <http://www.flake.igb-berlin.de/ep-data.shtml>, last access: May 2019.
- Kourzeneva, E., Bouttier, F., and Fischer, C.: Global dataset for the parameterization of lakes in numerical weather prediction and climate modelling, *ALADIN Newsletter*, July–December, Meteo-France, Toulouse, France, 46–53, 2009.
- Kourzeneva, E., Asensio, H., Martin, E., and Faroux, S.: Global gridded dataset of lake coverage and lake depth for use in numerical weather prediction and climate modelling, *Tellus A*, 64, 15640, <https://doi.org/10.3402/tellusa.v64i0.15640>, 2012.
- Raymond, P. A., Hartmann, J., Lauerwald, R., Sobek, S., McDonald, C., Hoover, M., Butman, D., Striegl, R., Mayorga, E., Humborg, C., Kortelainen, P., Durr, H., Meybeck, M., Ciais, P., and Guth, P.: Global carbon dioxide emissions from inland waters, *Nature*, 503, 355–359, <https://doi.org/10.1038/nature12760>, 2013.
- Samuelsson, P., Kourzeneva, E., and Mironov, D.: The impact of lakes on the European climate as simulated by a regional climate model, *Boreal Environ. Res.*, 15, 113–129, 2010.
- Stepanenko, V. M., Machulskaya, E. E., Glagolev, M. V., and Lykossov, V. N.: Numerical modeling of methane emissions from lakes in the permafrost zone, *Izvestiya Atmos. Ocean. Phys.*, 47, 252–264, 2011.
- Verpoorter, C., Kutser, T., Seekell, D. A., and Tranvik, L. J.: A global inventory of lakes based on high-resolution satellite imagery, *Geophys. Res. Lett.*, 41, 6396–6402, <https://doi.org/10.1002/2014GL060641>, 2014.