



2013

# The Importance of Open Data and Software for Large Scale Hydrological Modelling

L. Strömbäck

*Swedish Meteorological and Hydrological Institute, lena.stromback@smhi.se*

B. Arheimer

*Swedish Meteorological and Hydrological Institute*

G. Lindström

*Swedish Meteorological and Hydrological Institute*

C. Donnelly

*Swedish Meteorological and Hydrological Institute*

D. Gustafsson

*Swedish Meteorological and Hydrological Institute*

*See next page for additional authors*

Follow this and additional works at: <https://scholarsarchive.byu.edu/openwater>

## BYU ScholarsArchive Citation

Strömbäck, L.; Arheimer, B.; Lindström, G.; Donnelly, C.; Gustafsson, D.; Isberg, K.; Pechlivanidis, I.; Andersson, J.; and Pers, C. (2013) "The Importance of Open Data and Software for Large Scale Hydrological Modelling," *Open Water Journal*: Vol. 2 : Iss. 1 , Article 32.

Available at: <https://scholarsarchive.byu.edu/openwater/vol2/iss1/32>

This Article is brought to you for free and open access by the All Journals at BYU ScholarsArchive. It has been accepted for inclusion in Open Water Journal by an authorized editor of BYU ScholarsArchive. For more information, please contact [scholarsarchive@byu.edu](mailto:scholarsarchive@byu.edu), [ellen\\_amatangelo@byu.edu](mailto:ellen_amatangelo@byu.edu).

---

# The Importance of Open Data and Software for Large Scale Hydrological Modelling

## **Authors**

L. Strömbäck, B. Arheimer, G. Lindström, C. Donnelly, D. Gustafsson, K. Isberg, I. Pechlivanidis, J. Andersson, and C. Pers

## **The Importance of Open Data and Software for Large Scale Hydrological Modelling**

L. Strömbäck, B. Arheimer, G Lindström, C. Donnelly,  
D. Gustafsson, K. Isberg, I. Pechlivanidis, J. Andersson, C. Pers  
Swedish Meteorological and Hydrological Institute,  
Folkborgsvägen 17, 601 76 Norrköping, Sweden  
Email: lena.stromback@smhi.se

An important goal for hydrological research at SMHI in Sweden is large scale hydrological modelling. The goal of this work is to simulate hydrological conditions over large geographical areas, such as Europe and the Arctic region. Our most important tool is the HYPE model, which is designed to enable simulation over large areas. However, another important prerequisite for these large scale models is the availability of open data, which enables us to collect the information required for the model setup. In this paper we will give an overview over the work and focus on the aspects that make large scale modelling possible. In particular we will describe the HYPE model, the use of open data sources for modelling, how data is adapted to the models and how we feed results back to the community. In the end of the paper we give an overview of future perspectives and make an overview of the new EU project SWITCH-ON.

### **HYPE and large scale modelling**

The Hydrological Predictions for the Environment (HYPE) model is the core of our work at Hydrological Research at SMHI. The model is a dynamic, semi-distributed, process-based, integrated catchment model (Lindström et al., 2010). It uses well-known hydrological and nutrient transport concepts and can be applied for both small and large scale assessments of water resources and status. In the model, the landscape is divided into classes according to soil type, vegetation and altitude. The soil representation is stratified and can be divided in up to three layers. Water and substances are routed through the same flow paths and storages (snow, soil, groundwater, streams, rivers, lakes) considering turn-over and transformation on the way towards the sea.

The HYPE model is developed for being easy to set up for large scale applications and to make use of available data about geography. The process of setting up the model for a new geographic area is therefore highly dependent on using and adjusting external data sources. Based on the general description of the geography and other related information the model can simulate river runoff and other variables important for hydrological applications. The geographic area to be modelled consists of one or several catchment areas. Each catchment is divided into subbasins which is the basic unit for our geographical setup.

For each model setup in HYPE a number of soil and land use classes are defined. The hydrological properties are connected to these classes. For each subbasin in the model its hydrological properties are given as percentages of soil and land use classes. To describe nutrient transformation the soil is modelled as up to three layers which can be different for each land use. As an example, agricultural land classes commonly use three soil layers. Other properties can be connected to these classes, for instance, for agricultural land we describe which crops are most common and how these are fertilized. The HYPE model is calibrated based on properties of each soil and land use class for the whole geographic area. This allows us to calculate factors related to water resources and water quality with very high spatial detail for large geographic areas. Towards the end of the calibration process, further fine tuning of the model can be done based on defined regions.

An important aspect for the success in large scale models is the HYPE code itself. The HYPE code is flexible, efficient and constantly evolving. This is a great advantage when setting up hydrological models in different parts of the world with very different geographic conditions; varied access to information and with different purposes. The current version of the code and other tools enables us to

relatively fast set up a model for a new region in the world. Analysing this first model leads to new hydrological insights which further develop not only the particular model, but also the HYPE model and code.

## Large scale modelling with HYPE

With the HYPE model and other software and tools developed and used at the Hydrological research unit it has been possible to apply the HYPE model in many different regions of the world. The map in figure 1 shows the geographic extent of models available and under development.



Figure 1: Geographic extent of developed HYPE models.

For the European continent, we currently have three different models; The **S-HYPE** model for Sweden with a median subbasin resolution of 7 km<sup>2</sup>; The **BALT-HYPE** model for the Baltic Sea with a median sub-basin resolution of 315 km<sup>2</sup>; and the **E-HYPE** model for the whole Europe with a median sub-basin resolution of 200 km<sup>2</sup>. All the above models simulate hydrological and nutrients for their respective areas. They are running operationally at SMHI.

For other parts of the world we have a number of models at different stage of development. For South America we have developed the **LPB-HYPE**, a model of the La Plata Basin, simulating hydrological variables (median subbasin resolution = 500 km<sup>2</sup>). The model with the largest geographic extent is the **ARCTIC-HYPE** for the entire Arctic region (defined here as all land areas with runoff to the Arctic Ocean). This model simulates hydrological variables and first results from a hindcast simulation have been delivered for 1961-2009. We also have on-going development of models for the Niger River in Africa, **NIGER-HYPE**, the Indian region, **In-HYPE**, and the Arab (Middle East and Northern Africa) region, **MENA-HYPE**. For **In-HYPE** and **NIGER-HYPE** the plan is to deliver first results during 2013.

## The use of open data for setting up a model

All our large scale models are dependent on reliable data for their setup. For Sweden, we can make benefit of databases for Sweden hosted by our own and other Swedish institutes. For larger regions we are dependent on data from open European or global databases. Table 1 gives an overview of the data used for the setup a selection of our model setups.

While setting up a new model, a lot of effort is put into finding and selecting appropriate datasets, quality check and integrate them into a model setup for HYPE. To speed up this process we have developed our own tool WHIST (World Hydrological Information Setup Tool). This tool uses radar data on the topography to define the basic subbasin structure of the model. The tool can then combine the information with other data, e.g. soil and land use into a first version of the model setup. WHIST can be applied in several different ways to develop information for hydrological models out of

different types of databases. From the information of topographic databases WHIST can aid in delineating the subbasins, and the linking (routing) between them. The tool can also add information from other databases to determine the properties of each subbasin for example, soil characteristics and soil types. WHIST is developed at SMHI and a flexible tool that can be customized and tailored for different uses.

	Sweden	Europe	Arctic	Niger River	India	La Plata Basin
<b>Acronym in this paper</b>	S-HYPE <sup>k</sup>	E-HYPE	ARCTIC-HYPE	NIGER-HYPE	In-HYPE	LPB-HYPE
<b>Total area (km<sup>2</sup>)</b>	525 000	9.6 million	23.3 million	2.1 million	3.3 million	3.1 million
<b>No. of sub-basins</b>	37000	35000	30770	803	6010	6182
<b>No. of discharge stations</b>	303	1007	1349	71	43	68
<b>Topography</b>	SVAR (SMHI)	HydroSHEDS <sup>a</sup> , Hydro 1K <sup>a</sup>	Hydro 1K <sup>a</sup>	HydroSHEDS <sup>a</sup>	HydroSHEDS <sup>a</sup>	HydroSHEDS <sup>a</sup> , Hydro 1K <sup>a</sup>
<b>Land Use</b>	CORINE <sup>a</sup>	CORINE <sup>a</sup> , GLC2000 <sup>a</sup>	GLC2000 <sup>a</sup>	GLC2000 <sup>a</sup>	GLC2000 <sup>a</sup>	GLC2000 <sup>a</sup>
<b>Soil</b>	SLU	ESD <sup>a</sup> , DSMW <sup>a</sup>	HWSD <sup>b</sup>	HWSD <sup>b</sup> , WISE <sup>c</sup>	HWSD <sup>b</sup>	ISRIC World Soil Information
<b>Discharge observations</b>	SMHI	GRDC <sup>a</sup> , EWA <sup>a</sup> , BHDC <sup>a</sup>	GRDC <sup>a</sup> , USGS <sup>d</sup> , R- ArcticNET <sup>e</sup>	GRDC <sup>a</sup> , ABN <sup>f</sup>	GRDC <sup>a</sup>	GRDC <sup>a</sup> , and local observations
<b>Precipitation</b>	PTHBV (SMHI)	ERA- INTERIM <sup>a</sup>	WFDEI <sup>g</sup>	WFD <sup>h</sup>	APHRODITE	ERA- INTERIM <sup>a</sup>
<b>Temperature</b>	PTHBV (SMHI)	ERA- INTERIM <sup>a</sup>	WFDEI <sup>g</sup>	WFD <sup>h</sup>	AphroTemp	ERA- INTERIM <sup>a</sup>
<b>Lakes and reservoirs</b>	SVAR (SMHI)	GLWD <sup>a</sup>	GLWD <sup>a</sup>	GLWD <sup>a</sup> , Grand <sup>i</sup>	GLWD <sup>a</sup> , Grand <sup>i</sup>	-
<b>Irrigation</b>	-	EIM <sup>a</sup> , GMIA <sup>a</sup>	-	-	GMIA	-
<b>Crop types</b>	SCB	MIRCA2000 <sup>j</sup> , CAPRI <sup>a</sup>	-	-	-	-

Table 1: An overview of underlying data for a selection of our HYPE models. For more information about the databases a:(see Donnelly *et al.*, 2013), b:(FAO *et al.*, 2009), c: (Batjes, 2012), d:(USGS, 2011), e:(R-ArcticNET, 2011), f:(ABN, 2008), g:(Weedon *et al.*, 2012), h:(Weedon *et al.*, 2011), i:(Lehner *et al.*, 2011), j: (Portmann *et al.*, 2010). k: ([www.vattenweb.se](http://www.vattenweb.se))

## Sharing results

In addition to understanding hydrological modelling and processes, an important goal with our work is to make our data and tools available as open data and services. We currently do this in three different ways; result from hindcast simulations are available on the web; web-based tools for further exploration of our results and finally the HYPE OSC that makes the HYPE source code available for public use. In this section we will give brief information of each of these resources.

**HYPE simulations on the web:** The S-HYPE, BALT\_HYPE and E\_HYPE models are currently available via the web site [hypeweb.smhi.se](http://hypeweb.smhi.se) for export and exploration of HYPE simulations. The basis for all these websites is a map where the user can select a sub basin and download hindcast simulation data for water and nutrients. In addition the E-Hype and S-HYPE sites contain results from climate impact studies.

**Web-based tools:** The S-Hype version of this website (*vattenweb.smhi.se*) is used by the water authorities to fulfil the Water Framework Directive and the Marine Strategy Framework Directive. It is used for characterization, forecasts, and scenario analyses. Therefore, this website contains more information and tools compared to our other applications. In particular, the web site contains tools that can be used for evaluation of the performance of S-HYPE compared to observations. In addition there is a scenario tool making it possible to experiment with changes in nutrient load from different sources. Both these tools are interactive and enable further evaluation and exploration of the model results.

**The HYPE open source community:** This community makes the source code of HYPE available for anyone interested in further development of HYPE. The HYPE OSC (*hype.sourceforge.net*) is an open source initiative under the Lesser GNU Public License taken by SMHI to strengthen international collaboration in hydrological modelling and hydrological data production. The hypothesis is that more brains and more testing will result in better models and better code. The code is transparent and can be changed and learnt from. New versions of the main code are delivered frequently. The main objective of the HYPE OSC is to provide public access to a state-of-the-art operational hydrological model and to encourage hydrologic expertise from different parts of the world to contribute to model improvement. HYPE OSC is open to everyone interested in hydrology, hydrological modelling and code development – e.g. scientists, authorities, and consultancies.

The HYPE source code is designed to efficiently handle large scale modelling for forecast, hindcast and climate applications. The code is under constant development to improve the hydrological processes, efficiency and readability. In the beginning of 2013 we released version with new and better modularization, corresponding to hydrological processes which will make the code easier to understand and further develop for a new user. An important challenge in this process is to produce code that is easy for anyone to understand and work with, but still maintain the properties that make the code efficient enough for large scale applications.

### Future perspectives: SWITCH-ON

The work with large scale modelling in our group has shown the importance of available open datasets. However, our experience also shows that quality checking and adaption of available resources to fit hydrological requirements is a time consuming work. Therefore it would be a great benefit if more resources adapted to hydrological needs were available to the community. In the fall 2013 we will start a new European Cooperation, the SWITCH-ON project with 16 partners. In this project we will address these questions and work on improving the processes and data sources.

One important goal for the project is to create Spatial Information Platform for easy access and integration of water-related data. The platform will be designed for “tailoring” environmental data to water applications. Existing and evolving open data will be used, such as the GEOSS data-CORE and its 284 water products, and the GMES programme. These datasets provides essential data for water modelling, for instance land-cover and land-use data, seasonal and annual change monitoring and climate forcing. The SWITCH-ON platform will also connect with on-going progress of open public-sector data portals, which is prescribed in the INSPIRE directive by 2013.

The focus of the project is to foster new forms of collaborative research and new innovative products and services through openness and collaboration by boosting the use of, and to add value to, existing and emerging open datasets. The platform will provide a shared research environment as a Virtual Water-Science Laboratory. Or ambition is that this platform will stimulate the scientific debate and new achievements through rigorous comparison of research results from different methods, assumptions or models. In addition an aim in the project is to provide the basis for and develop new products and services for operational use, leading to new businesses across environmental knowledge domains.

## Acknowledgements

This work is an overview of several projects performed at the hydrological research unit at SMHI. Many persons have contributed to various parts of this work. The development of the models and tools have been funded by the European Commission, the Swedish Environmental Protection Agency, the Swedish Research Council Formas, the Swedish Research Council (Vetenskapsrådet) and the Swedish International Development Cooperation Agency. The development of the S-Hype model, has in addition, been funded by the Swedish water authorities.

## References

- ABN. (2008) Niger-HYCOS Hydrological Information System. Retrieved March 1, 2012, from <http://nigerhycos.abn.ne/>
- Andersson, J.C.M.; Pechlivanidis, I.G.; Gustafsson, D.; Donnelly, C. ; and B. Arheimer (2013) Key Factors for Improving Large-scale Hydrological Model Performance. *Proceedings of the 13<sup>th</sup> International Conference on Environmental Science and Technology, Athens, Greece. Paper no. CEST13\_0753.*
- Batjes, N. H. (2012) ISRIC-WISE derived soil properties on a 5 by 5 arc-minutes global grid (ver. 1.2) ( No. 2012/01). Wageningen, The Netherlands: ISRIC - World Soil Information.
- Donnelly, C., Arheimer, B., Capell, R., Dahné, J. & Strömqvist, J. (2013) Regional overview of nutrient load in Europe – challenges when using a large-scale model approach, E-HYPE.
- FAO, IIASA, ISRIC, ISS-CAS & JRC. (2009) Harmonized World Soil Database (version 1.1), 38. Rome, Italy: FAO and IIASA.
- Lehner, B., Liermann, C. R., Revenga, C., Vörösmarty, C., Fekete, B., Couzet, P., Döll, P., *et al.* (2011) High-resolution mapping of the world's reservoirs and dams for sustainable river-flow management. *Frontiers in Ecology and the Environment* **9**(9), 494–502. doi:10.1890/100125
- Lindström, G., Pers, C., Rosberg, J., Strömqvist, J. & Arheimer, B. (2010) Development and testing of the HYPE (Hydrological Predictions for the Environment) water quality model for different spatial scales. *Hydrology Research* **41**(3-4), 295–319. doi:10.2166/nh.2010.007
- Portmann, F. T., Siebert, S. & Döll, P. (2010) MIRCA2000—Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling. *Global Biogeochem. Cycles* **24**(1), GB1011. doi:10.1029/2008GB003435
- R-ArcticNET. (2011) R-ArcticNET Pan-Arctic River Discharge Data. Retrieved December 7, 2011, from <http://www.r-arcticnet.sr.unh.edu/v4.0/AllData/index.html>
- USGS. (2011) USGS National Water Information System. Retrieved November 2, 2011, from <http://waterdata.usgs.gov/nwis>
- Weedon, G. P., Gomes, S., Balsamo, G., Best, M. J., Bellouin, N. & Viterbo, P. (2012) WATCH forcing data based on ERA-INTERIM. Retrieved December 15, 2012, from <ftp://rfddata:forceDATA@ftp.iiasa.ac.at>
- Weedon, G. P., Gomes, S., Viterbo, P., Shuttleworth, W. J., Blyth, E., Österle, H., Adam, J. C., *et al.* (2011) Creation of the WATCH Forcing Data and Its Use to Assess Global and Regional