Lake water levels for calibration of the S-HYPE model

Göran Lindström

ABSTRACT

Lake water levels are easy to measure, and can be seen as a transformation of discharge. Water level measurements from 42 lakes in Sweden were analysed, and eight of these lakes were modelled using the Swedish S-HYPE model. The objective was to test if a hydrological model can be calibrated using water level data instead of discharge. A semi-analytical lake routing was developed to resolve fast variations in water levels in small lakes. Seven tests were made in which data were sampled from real observations and the model was calibrated to each data set. It was found that water levels are useful for calibration of hydrological models, even without measuring discharge and establishing a traditional rating curve. The exponent in the rating curve equation could be set to 2 as a standard value. Approximate rating curves can be identified by model calibration. The results were improved considerably already when only four observed water levels were used (mean $NSE = 0.88$ for both water levels and discharge), compared to the general model (-1.44 and 0.76, respectively). Calibration of both the rating curve and the inflow gave nearly as good discharge simulations (mean $NSE = 0.92$) as traditional calibration using recorded discharge (0.93).

Key words | calibration, lake, level, rating curve, S-HYPE

Göran Lindström
Swedish Meteorological and Hydrological Institute,
Norrköping SE-601 76,
Sweden
E-mail: goran.lindstrom@smhi.se

INTRODUCTION

Sweden has about 100,000 lakes with a surface area >0.01 km^2. Water levels in the six largest lakes are measured and reported daily on the internet by the Swedish Meteorological and Hydrological Institute (SMHI). In addition, water levels in natural lakes are measured with the main purpose of estimating discharge through established rating curves. The raw water level data are usually not used for other purposes. Water level data are, for instance, not quality controlled and corrected, for example for ice jams, in the same way as discharge data. Lakes have a major impact on the hydrological and hydro-chemical conditions, and lake routing has been modelled explicitly in the Swedish HBV model since 1985 (Bergström *et al.* 1985).

S-HYPE is a hydrological model that covers all of Sweden in high spatial resolution (Strömqvist *et al.* 2012). It is based on the HYPE model code (Lindström *et al.* 2010). Both HYPE and S-HYPE are developed continuously and current versions are available at http://hype.sourceforge.net/ and http://vattenwebb.smhi.se/. The present S-HYPE version

(2012 2.0.0) has 36,692 sub-basins with a median area of 7 km^2. Lake water levels are explicitly modelled for 8,581 lakes. There are about 400 discharge stations in the model, which means that about 99% of the basins are ungauged. It is practically impossible to measure discharge at all points of interest, which is why models are used as a complement. A large number of discharge measurements are required, at different flow situations, before a reliable rating curve can be established. The cost of establishing a rating curve based on ten measurements was estimated at roughly 17,000 euro (Lena Brahm-Eriksson, personal communication). High flows, in particular, are difficult to capture and measure. Rating curves are thus often extrapolated for high flow situations. Rating curves at the SMHI are divided into 1, 2 or 3 intervals, corresponding to 3, 7 or 11 parameters. The large number of parameters makes it difficult to generalize this information to the many ungauged lakes in S-HYPE.

Since 2008, the SMHI has operated a network of temporary water level measurements that are rotated to new

locations after 2–3 years. The idea is that even short time series can provide valuable information, especially on the dynamics of a catchment (see also e.g., Seibert & Beven 2009). The objectives of this study were to:

1. improve the lake routing scheme of the HYPE model;
2. test if lake water level data are useful for calibration of hydrological models even without measuring discharge;
3. test if a hydrological model combined with lake water level data could be used in the establishment of rating curves as a cost-effective complement to existing techniques.

METHODS AND DATA

Calibration by super parameters

The S-HYPE model was originally calibrated with general parameters, i.e., parameters that were constant in the whole application or coupled to soil type or vegetation type (Strömqvist et al. 2012). These parameters were mainly calibrated for small, lake-free representative basins, dominated by specific land-uses and soil types. General lake routing parameters were thereafter established for a set of small basins. It proved difficult to find parameter values that were applicable to all of Sweden. For instance, the same characteristics for clay were assumed in the southern part of the country as in the north. To allow adaption to local conditions and improve model performance, the concept of super parameters has been introduced in the HYPE model. Super parameters have been formulated for both water quantity parameters and parameters which only affect water quality. They act within specified parameter regions and govern how key parameters deviate from the general values. Super parameters either affect only one parameter, or a set of related parameters. The following super parameters were used in this study:

1. Precipitation factor (%)
2. Evaporation factor (%)
3. Temperature correction (°C)
4. Recession coefficient factor for all soil layers and surface runoff (%)
5. Lake outflow factor (%).

The super parameters are a convenient compromise between a completely general model and a model where all parameters are calibrated for each gauging station. For instance, the relation between parameters and recession rates from individual soil layers can be maintained. This reduces the risk of unrealistic parameter combinations. The use of super parameters has improved the S-HYPE results considerably, and made the model more useful for forecasts and warnings, for example. The long-term vision is to continuously improve the model structure and input data and reduce the use of local adjustments.

Lake routing

The rating curve for a lake outlet is typically written as:

$$q = k(w - w_0)^p$$

where q is the outflow (m³/s), w is the water level (m), w_0 is a threshold (m) and p is an exponent. The outflow parameter k determines the outflow at a water level 1 m above the threshold. Water levels can thus also be interpreted as a transformation of discharge:

$$(w - w_0) = (q/k)^{1/p}.$$

In fact, water levels contain additional information, since the outflow may cease ($q = 0$) during dry spells ($w < w_0$). It is more convenient to use the height h (m) above the threshold ($h = w - w_0$). Assuming that variations in the lake surface area A (m²) are negligible, the water balance for the lake will be:

$$q = i - \frac{Adh}{dt}.$$

The net inflow i (m³/s) accounts for river inflow, evaporation and precipitation on the lake surface. Even if the water level does not uniquely determine the outflow from a specific lake, it still measures the discharge as the accumulated difference between inflow and outflow since an initial level h_0:

$$h(t) = h_0 + \frac{1}{A}\int_0^t (i - q)d\tau.$$

In reality, the inflow varies in time, and the water balance equation has to be solved dynamically. In the first

version of the HYPE model (Lindström *et al.* 2010) lake routing was performed with a daily time step, in the simplest numerical way, using an explicit calculation. This works well as long as the flows are small compared to the volume of water above the threshold ($A \cdot h$). In small lakes with high inflow this numerical scheme was too simple, and lake water stages collapsed. The 8,581 lakes in the present S-HYPE version range in area from 0.002 to 5,435 km^2. Shorter time steps were therefore tested. The problem is to find a time step which is short enough for resolving fast variations in small lakes, but not so short that it leads to a waste of computing resources for large lakes. Water levels can only be used in the calibration if the model itself can resolve the variations with sufficient accuracy. One alternative is to use a more advanced numerical scheme, but for the HYPE model a semi-analytical approach was developed. The rating curve is first linearized at the level (h_0) at the start ($t = 0$) of each time step. If the inflow is assumed to be constant during a time step, the linearized equation can be solved by standard methods. For instance, integration over the time step Δt from 0 to t gives:

$$h_t = h_0 + (i/c/A - h_0/p)(1 - e^{-c\Delta t}).$$

Here h_t is the water level at the end of the time step and c (unit s^{-1}) is the linearized recession rate:

$$c = \frac{pkh_0^{p-1}}{A}.$$

The inverted value ($1/c$) gives a linearized time constant (T), which governs the time with which an inflow hydrograph will be delayed while passing through a lake. It is usually more convenient to convert T from seconds to days or hours. The time constant will vary with water level due to the non-linearity of the rating equation.

The mean outflow $\bar{q}$ during the time step can thereafter be obtained from the water balance:

$$\bar{q} = i - \frac{A(h_t - h_0)}{\Delta t}.$$

An advantage of this semi-analytical approach is that there is no need for conversion of k when changing from

for instance, daily to hourly time steps. For comparison with recorded water levels (daily means), the calculated levels were also converted into daily means: $(h_0 + h_t)/2$. This simplification was used here, but an exact linearized average can be computed as: $h_0 + \bar{q}/c/A - h_0/p$. This formulation gives an even higher agreement with daily mean observed water levels.

Computational details

Rating curves in S-HYPE

In the current S-HYPE version there are 42 lakes for which simplified rating curves have been established based on measurements of water level and the corresponding discharge from the official rating curve. The official rating curves (with 3–11 parameters) were simplified into curves using only three parameters. This was done to enable the use of the semi-analytical solution described above and to simplify generalization to lakes where there are no measurements (about 99% of the lakes in S-HYPE). The rating curve can be motivated from the conversion from potential to kinetic energy at a hydraulic control section, which gives $v = \sqrt{2gh}$. For a parabolic cross-section the width is proportional to $\sqrt{h}$. The outflow is then obtained by integration over the cross-section area A:

$$q = \int_A v dA = \int_A \sqrt{2gh} dA \propto \int_0^h z dz \propto h^2.$$

This thus suggests that $p \approx 2$ is appropriate for parabolic sections. This shape was assumed to be more typical for natural lake outlets than either rectangular or triangular (with 90° angle) cross-sections, corresponding to $p = 1.5$ and $p = 2.5$, respectively (see further e.g., Maidment 1992).

Calibration criteria

The commonly used Nash–Sutcliffe efficiency (*NSE*) criterion (Nash & Sutcliffe 1970) is computed as:

$$NSE = 1 - \frac{mse}{V_r}.$$

Table 1 | Selected lakes and discharge stations, from north to south. The lake area as a fraction of the whole upstream area is given

Lake	Discharge station	Main river no.	Station no.	Basin area km²	Lake area km²	Fraction %
Torneträsk	Abisko	1	2357	3,346	330	9.9
Kukkasjärvi	Kukkasjärvi	3	1160	494	3.7	0.8
Räktjärv	Räktjärv	4	17	17,387	13	0.1
Karatj	Karats	9	1403	1,174	60	5.1
Överstjuktan	Skirknäs	28	2275	418	23	5.5
Ånnsjön	Ånnsjön	40	2151	1,563	58	3.7
Hasselasjön	Hassela	44	2273	651	8.4	1.3
Allgunnen	Rörvik	98	200	159	14	8.6

NSE compares the mean square error *mse* to the variance V (r for recorded data). For evaluation of water levels a slight re-formulation was made. The mean square error was re-written (see e.g., Gupta *et al.* 2009):

$$mse = E[(c - r)^2] = E[(c - r + \bar{r} - \bar{r})^2]$$
$$= V_c + V_r - 2C_{cr} + bias^2.$$

C here denotes the covariance and c computed data. This means that the contribution to *mse* from the systematic error (*bias*) can be estimated directly. Simultaneous calibration of the two rating parameters k and w_0 is difficult since they are so inter-dependent. In the case of water level data, the bias can always be removed by changing w_0. Calibration is thus made much more efficient by optimizing the modified criterion:

$$NSE^* = NSE + \frac{bias^2}{V_r}.$$

The bias can thereafter be eliminated by adjustment of w_0. This procedure is essentially the same as first maximizing the correlation and thereafter adjusting the threshold. *NSE* is often calculated on-line during a model simulation by reformulating the variance as:

$$Vr = E[(r - \bar{r})^2] = E[r^2] - E[r]^2.$$

This formulation can yield inaccurate results for water levels, due to numerical problems. The threshold w_0 may be several hundred metres above sea level, but the variations in lake water levels are only a few metres. It is, therefore, in practice, better to evaluate water levels above the threshold (i.e., $h = w - w_0$ instead of w).

Test calibrations using recorded water levels

Eight relatively large lakes with existing rating curves were selected from the full S-HYPE model, representing different sizes and geographical locations (Table 1). Hypothetical experiments were made by sampling from existing data series in the SMHI archive. All simulations were made for 1999-01-01 to 2008-12-31, with 1989–1998 as a warm-up period. All calibrations were made manually, by maximizing *NSE** for water levels and *NSE* for discharge. For all estimations of the lake rating parameters, a general value of $p = 2$ was assumed. Note that there actually is no daily recorded discharge, but only estimates of discharge using different rating curves (Figure 1).

Seven types of simulations were made (Table 2). The first (method Ref) is the reference to which the other tests are compared. It is the general S-HYPE model described above. Some of the eight test basins have been used in the general model calibration, but with very little influence from each one. In the reference situation, there is no
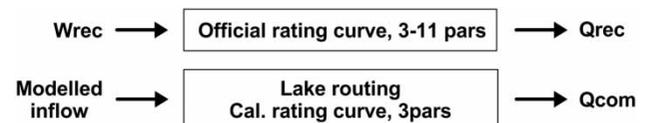

Figure 1 | Principles for estimation of recorded (rec) and computed (com) lake outflow.

Table 2 │ Set-up of test calibrations

	Method	Calibration of k using W data	Calibration using Q data	Calibration of inflow (super parameters)	Notes
1	Ref	–	–	–	General model, reference
2	4obs	4 observations	–	–	Four obs. points during 2 years
3	Max–min	Max–min	–	–	Only the range used, 2 years
4	kW	10 years, daily	–	–	
5	CalW	10 years, daily	–	Yes	
6	CalQ	–	10 years, daily	Yes	The official rating curve was used
7	Q(W)	–	–	–	Q by rating curve from method CalW and recorded W

calibration to local data using super parameters. The general lake rating parameters are used. In the second (4obs) and third simulations (max–min), the value of only a few observations was studied, both based on 2 years of data. In the second simulation only four water level observations were used, and in the third simulation only the range (max–min) was used. The four measurements in method 4obs were chosen as one measurement, the 15th of the months with highest and lowest recorded water levels for each lake, during the 2 years. These months are strictly speaking not known, but the hydrological regime in different parts of the country is well known. In the fourth method (kW), 10 years of daily water levels were used for calibration of the k parameter. In methods 2–4, only the rating parameter k was calibrated and the inflow was not modified.

In the fifth method (CalW), both the lake routing (k) and super parameters that affect the lake inflow were calibrated. This corresponds to a normal calibration but using only water levels instead of discharge. The official discharge was not used in the calibrations 2–5, but only for subsequent evaluations. The sixth method (CalQ) is the traditional situation in which the official rating curve (although simplified, see above) is used for the lake routing, and the inflow to the lake is calibrated using 10 years of discharge data (by the official rating curve). The last method (Q(W)) basically measures how well the rating curve established using only water level data corresponds to the official rating curve. This is done by comparing the discharges according to the two curves, both driven by the observed water levels (Wrec in Figure 1).

RESULTS AND DISCUSSION

Out of 42 lakes for which simplified rating curves were developed, most had exponent p values of around 2. The median value was 2.0 and the mean 2.3, due to some lakes with high p values. For all lakes, a reasonable agreement could furthermore be established for $p = 2$, by adjusting the k value and w_0 appropriately. This is in agreement with the assumption of parabolic lake outlets made above, as well as values reported by Maidment (1992). The assumption of $p = 2$ as a general value thus seems appropriate.

The results from the calibration tests are summarized in Tables 3 and 4 and Figure 2. The general model simulated discharge rather well for these sites (mean $NSE = 0.76$), considering that this is partly a PUB case (prediction in ungauged basins) (Blöschl et al. 2013) since the general model was not adapted to local data. The test basins are fairly large and have an outlet lake, two factors which usually contribute to better results. On the other hand, most water levels were simulated poorly using general lake parameter values.

Figure 3 shows selected results for Lake Torneträsk (Abisko). This was one of the lakes where the use of water level data gave the largest improvements compared to the general model. It was chosen for illustration since the methods and results could be seen clearly. A dramatic improvement was found already from the adaption of the rating parameter k with the use of four water level observations or the range during 2 years in this lake. In general, the simulated water levels improved considerably already with very few water level observations, but there were a few lakes where discharge

Table 3 │ NSE for water levels

Lake/Method	Ref	4obs	Max–min	kW	CalW	CalQ
Torneträsk	0.144	0.964	0.968	0.968	0.976	0.978
Kukkasjärvi	0.528	0.883	0.883	0.884	0.890	0.889
Räktjärv	−6.745	0.871	0.882	0.877	0.926	0.926
Karatj	−3.873	0.930	0.931	0.933	0.949	0.935
Överstjuktan	−1.436	0.919	0.917	0.919	0.922	0.919
Ånnsjön	−1.668	0.635	0.813	0.898	0.926	0.919
Hasselasjön	0.616	0.902	0.902	0.912	0.920	0.914
Allgunnen	0.890	0.921	0.923	0.925	0.948	0.946
Mean	−1.443	0.878	0.895	0.915	0.932	0.928

Table 4 │ NSE and volume errors (V.E.) for discharge

Lake/Method	Ref	4obs	Max–min	kW	CalW	CalQ	Q(W)
Torneträsk	0.519	0.951	0.941	0.943	0.966	0.969	0.975
Kukkasjärvi	0.872	0.887	0.887	0.888	0.886	0.897	0.992
Räktjärv	0.877	0.858	0.864	0.862	0.912	0.924	0.958
Karatj	0.744	0.910	0.910	0.908	0.926	0.935	0.946
Överstjuktan	0.719	0.890	0.888	0.890	0.904	0.918	0.981
Ånnsjön	0.600	0.770	0.802	0.838	0.878	0.907	0.959
Hasselasjön	0.925	0.924	0.924	0.920	0.917	0.932	0.965
Allgunnen	0.819	0.879	0.886	0.900	0.943	0.951	0.986
Mean NSE	0.759	0.884	0.888	0.894	0.917	0.929	0.970
Mean V.E.	−3%	−3%	−3%	−3%	−2%	−2%	−2%
Mean ABS (V.E.)	6%	6%	6%	6%	6%	2%	6%
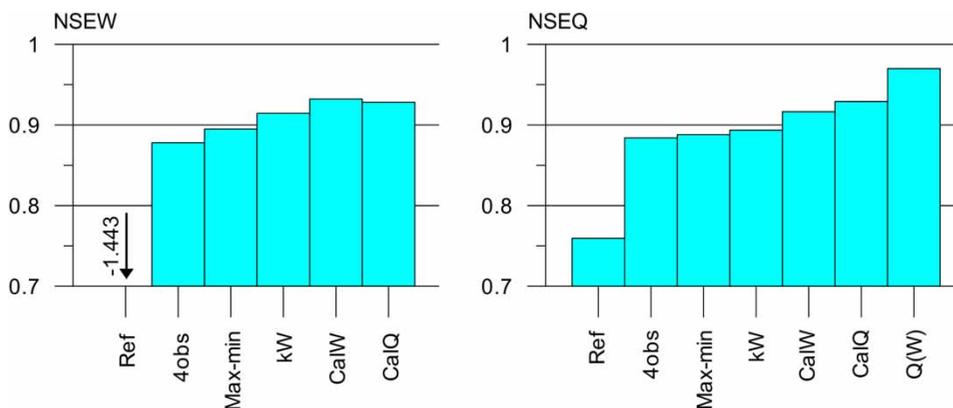
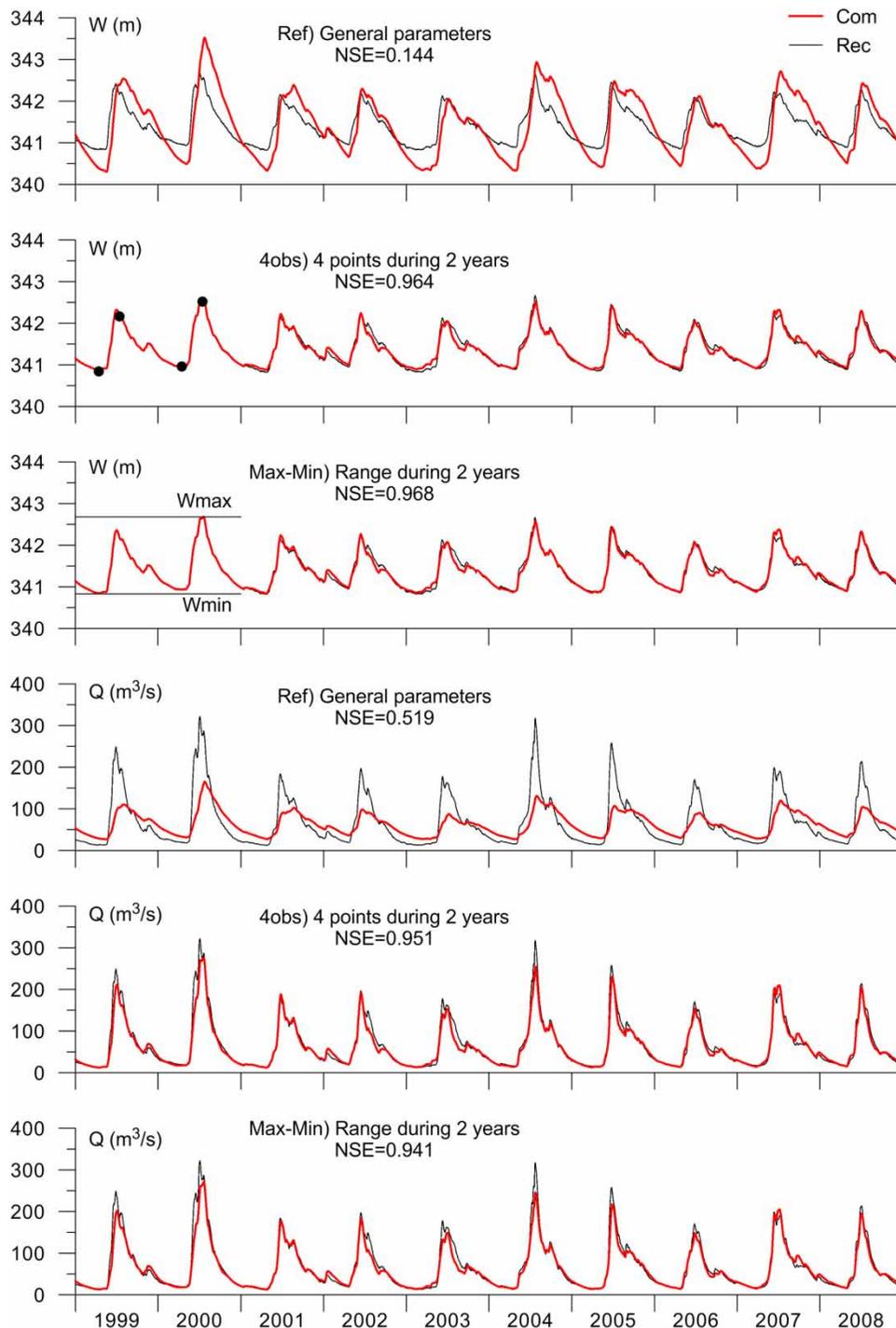**Figure 2** │ Summary of the results, mean NSE for the eight basins.

Figure 3 | Example of results for Lake Torneträsk for the general, partly PUB-type calibration, and the two calibration methods that use the least observations.

was not always improved. On average, however, both water levels and discharge were improved considerably (Figure 2). It is interesting to note that the improvements were so large with very little data: for instance, four water level measurements or the range between minimum and maximum levels during 2 years. Before the 1970s, water levels in large lakes in the SMHI network were often read manually a few times per week. The results obtained here show that this

was sufficient for most purposes due to the slow variation in water levels in large lakes.

A rough estimate of the range between minimum and maximum water levels can be made without actually performing measurements, for instance, from local knowledge or other soft data (see e.g., Seibert & McDonell 2002). It would, for instance, be very difficult to make a reliable estimation of the range in outflow from Lake Torneträsk during 2 years, 1999–2000 ($=310 \text{ m}^3/\text{s}$), without measurements (see Figure 3). It would, on the other hand, be much easier to make a reasonable estimate of the range in water levels ($=1.85 \text{ m}$). Even a rough estimate improves the results considerably in many cases.

Full use of the water level data during 10 years, including re-calibration of the inflow, gave nearly as good discharge simulations (on average 0.917, Table 4) as a traditional calibration using recorded discharge (0.929). The water level calibration furthermore resulted in a slightly better simulation of water levels (0.932 versus 0.928). The largest part of the improvement is, however, already obtained by adjusting only the rating parameter k. This is easily done in practice. The adjustment of k improves the modelled water levels in the particular lake. If the lake is small compared to the upstream area (as for instance Kukkasjärvi, Räktjärv and Hassela), the downstream flow is not improved so much by this adjustment. However, for large lakes, such as Torneträsk in Figure 3, the downstream discharge is improved dramatically by only adjusting the rating parameter k. The use of water levels can improve the timing of discharge, by adjusting parameters affecting storage of water in a basin and the timing of snow melt.

In Table 3 it can also be seen that water level variations were modelled successfully after calibration. This is also the case for Lake Räktjärv, which is small compared to the upstream area (0.1%), and one of the lakes where the previous lake routing method failed completely. The reason for failure is that the time constant (T) at high water levels in this lake is only 6–7 hours, which is why a daily time step is not sufficient. The water levels were modelled adequately by the new lake routing scheme, both in the lakes in this paper and in other lakes in S-HYPE. This shows that the new lake routine works well in practice.

Volume errors were rather low to start with. The average volume error for the general model was −3%, and the

average absolute volume error was 6% (Table 4). Only one of the calibrations using water level data (CalW) involved adjustment of the inflow volume. This calibration only slightly improved the volume errors compared to the general model. Full calibration using discharge reduced the average absolute volume error to 2%. *NSE* is much more sensitive to timing errors than to volume errors. In practice, even after a traditional optimization of *NSE*, there is often a significant remaining volume error (see e.g., Lindström 1997). The volume is, furthermore, quite uncertain in recorded discharge series. In practice, the mean discharge can change by ±5–10%, or even more, when a rating curve is revised. For flood warnings, the timing can be more important than the actual discharge level, since the modelled flow is usually compared to statistics based on model results from a reference period (see e.g., Bergstrand *et al.* 2014).

The rating curves that are established from the model calibration can also be used for estimation of discharge using the recorded water levels. Figure 4 shows a comparison between the modelled rating curve ($p = 2$) and the
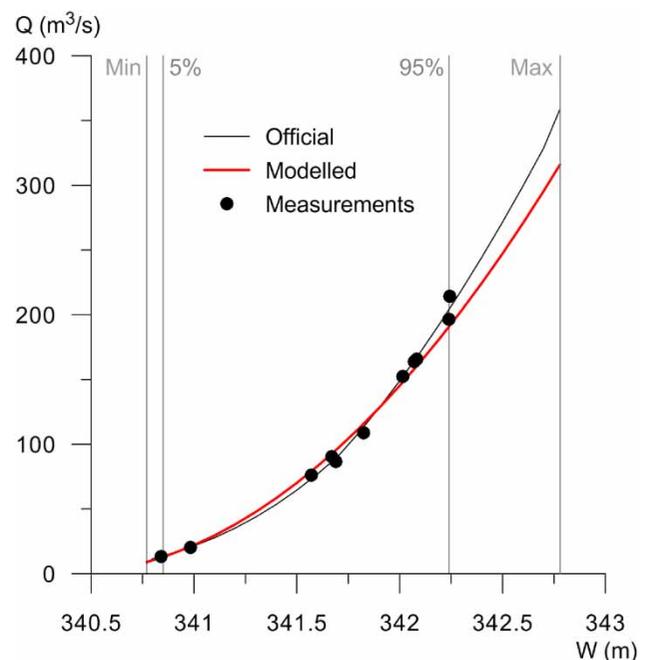

Figure 4 | Example of rating curves for Lake Torneträsk. Official rating curve, consisting of two intervals with seven parameters, the actual discharge measurements on which the curve is based, and the rating curve estimated from the model calibration to water levels, with three parameters (method kW, k estimated using 10 years of daily water levels). Percentiles of measured water levels in the full data series 1984–2014 are shown for comparison.
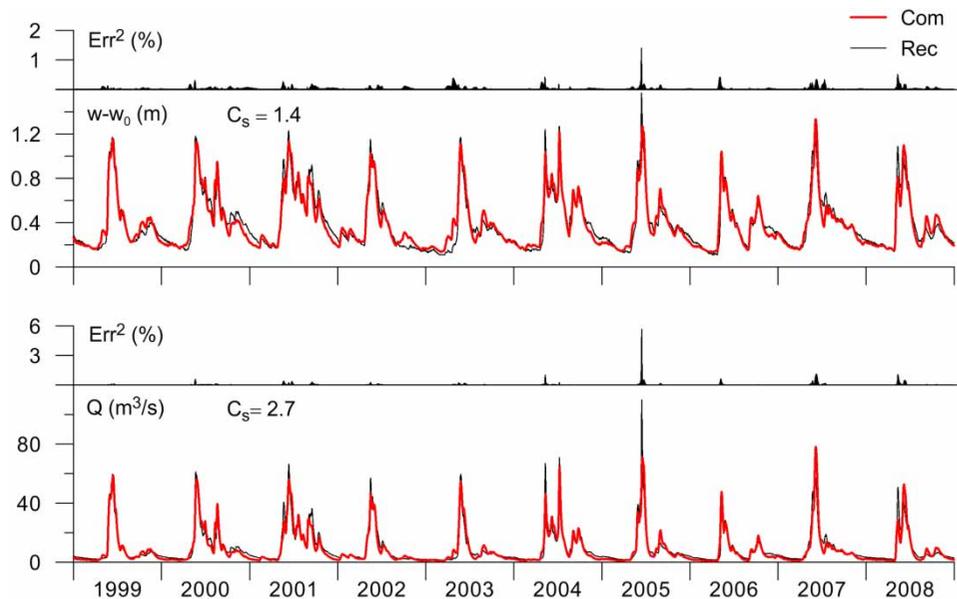
Figure 5 | Example of results for Lake Överstjuktan, method CalW (calibration using 10 years of daily water levels). Cs = the skewness of the recorded values. The error Err² is the squared error for each day as a percentage of the total squared error.

official rating curve ($p \approx 2.8$). The official curve is divided into two intervals, and thus has seven parameters. The two curves in the figure were derived with two completely different methods. They mostly agree well but differ at high levels. Neither curve is exactly correct. Together they illustrate that conversion from water levels to discharge is not made without uncertainties (see further e.g., Westerberg *et al.* 2011). What is usually called recorded discharge is, in fact, modelled discharge using recorded water levels as input data (cf. Figure 1). The discharges estimated by the official rating curves and the estimated curves, however, agreed well (average $NSE = 0.97$, Table 4), which should be high enough for most practical applications.

Calibration of hydrological models using water level data directly may have additional advantages. First, it involves no extrapolation of the rating curve. Second, it makes use of the real measurements of water levels and not discharge modelled via an estimated rating curve. Third, calibration using discharge is difficult due to the high variability and skewness of this variable (see e.g., Sorooshian & Gupta 1995). Figure 5 shows the dynamics of both water levels above the estimated threshold w_0 and discharge for Lake Överstjuktan (Skirknäs). Both the coefficient of variation and the skewness are about twice as high for discharge as for water levels. Furthermore,

when the differences between modelled and recorded values are squared as in the *NSE* criterion, large errors dominate at the expense of smaller errors during low flow. This is here clearly illustrated during the spring flood of 2005. Flow data are therefore sometimes transformed by taking logarithms for evaluation of low flows. The direct use of water level data is equivalent to using a power transformation of the discharge, usually the square root of discharge ($p = 2$). Individual events do not stand out as dramatically for water levels as they do for discharge (Figure 5). This should reduce the risk of sacrificing low flows during a calibration, although this was not investigated further here.

CONCLUSIONS AND CONCLUDING REMARKS

Water levels are easy and cheap to measure. They contain much of the information in a discharge record. Measured water levels were found to be useful for calibration of the S-HYPE model, even without measuring discharge and establishing a rating curve. Modelled water levels in the lake itself improve substantially, and on many occasions the water levels themselves are of great interest. Furthermore, the downstream flow is usually improved considerably if the lake is large in relation to the upstream

area. It is mainly the timing of discharge, rather than the runoff volume, that is improved. What is usually referred to as recorded discharge in the calibration of hydrological models is actually modelled using a rating curve and recorded water levels as input. An alternative is to use the actual measurements, i.e., water levels, in the calibration. Extrapolation of the rating curve is thereby avoided. Water levels are furthermore less skewed than discharge. Water level data deserve more attention than they have so far received and more efforts should also be spent on quality control of these measurements. The main conclusions from the present study are as follows:

- A semi-analytical lake routing method was developed for the HYPE model in order to resolve fast variations in water levels in small lakes with large inflow. It was found that the exponent p in the rating curve equation could be set $= 2$ as a standard value.
- Hydrological models can be improved considerably, compared to a model with general parameters, by calibration to water levels, even without using measurements of discharge and establishing a rating curve. In the relatively large lakes in this study, the results were improved already by using only four observed water levels (mean $NSE = 0.88$ for both water levels and discharge), compared to the general model (-1.44 and 0.76, respectively).
- Calibration of both the rating curve of a lake, and the inflow from upstream areas, gave nearly as good discharge simulations (mean $NSE = 0.92$) as a more traditional calibration where recorded discharge data were used (0.93). Approximate rating curves can be estimated by model calibration and thereafter used with recorded water levels as input for calculation of a discharge. This approximation of the recorded discharge agrees well with the official recorded discharge.

The methods and results should be valid for other regions with lakes, for instance, the Nordic countries. In Finland, water levels are, in fact, already used for model calibration (Vehviläinen *et al.* 2005). Further work should be done on how hydrological models can be combined with water level data to establish both timing and volume of discharge, as a complement to current techniques. The results presented here, however, show that even very little information on water levels is useful for improved calibration of hydrological models, even without discharge measurements and traditional establishment of a rating curve.

ACKNOWLEDGEMENTS

REFERENCES

Bergstrand, M., Asp, S. & Lindström, G. 2014 Nationwide hydrological statistics for Sweden with high resolution using the hydrological model S-HYPE. *Hydrol. Res.* **45** (3), 349–356.

Bergström, S., Brandt, M. & Carlsson, B. 1985 Hydrologisk och hydrokemisk modellberäkning i sjörika skogsområden [Hydrological and hydrochemical simulations in basins dominated by forests and lakes]. *Vatten* **41** (3), 164–171.

Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A. & Savenije, H. (eds) 2013 *Runoff Predictions in Ungauged Basins – Synthesis across Processes, Places and Scales*. Cambridge University Press, Cambridge, UK, p. 465.

Gupta, H. V., Kling, H., Yilmaz, K. K. & Martinez, G. F. 2009 Decomposition of the mean squared error and NSE performance criteria: implications for improving hydrological modeling. *J. Hydrol.* **377**, 80–91.

Lindström, G. 1997 A simple automatic calibration routine for the HBV model. *Nordic Hydrol.* **28** (3), 153–168.

Lindström, G., Pers, C. P., Rosberg, R., Strömqvist, J. & Arheimer, B. 2010 Development and test of the HYPE (Hydrological Predictions for the Environment) model – a water quality model for different spatial scales. *Hydrol. Res.* **41** (3–4), 295–319.

Maidment, D. R. 1992 *Handbook of Hydrology*. McGraw-Hill, New York, USA.

Nash, J. E. & Sutcliffe, J. V. 1970 River flow forecasting through conceptual models. Part I – a discussion of principles. *J. Hydrol.* **10** (3), 282–290.

Seibert, J. & Beven, K. J. 2009 Gauging the ungauged basin: how many discharge measurements are needed? *Hydrol. Earth Syst. Sci.* **13**, 883–892.

Seibert, J. & McDonell, J. J. 2002 On the dialog between experimentalist and modeler in catchment hydrology: use of

soft data for multicriteria model calibration. *Water Resour. Res.* **38** (11), 1241.

Sorooshian, S. & Gupta, V. K. 1995 Model calibration. In: V. P. Singh (ed.), *Computer Models of Watershed Hydrology.* Water Resources Publications, Highland Ranch, CO, USA, pp. 23–68.

Strömqvist, J., Arheimer, B., Dahné, J., Donnelly, C. & Lindström, G. 2012 Water and nutrient predictions in ungauged basins: set-up and evaluation of a model at the national scale. *Hydrol. Sci. J.* **57** (2), 229–247.

Vehviläinen, B., Huttunen, M. & Huttunen, I. 2005 Hydrological forecasting and real time monitoring in Finland: the watershed simulation and forecasting system (WSFS). In: *Innovation, Advances and Implementation of Flood Forecasting*. Technology, Conference Papers, Tromsø, Norway, 17–19 October 2005.

Westerberg, I., Guerrero, J.-L., Seibert, J., Beven, K. J. & Halldin, S. 2011 Stage-discharge uncertainty derived with a non-stationary rating curve in the Choluteca River, Honduras. *Hydrol. Process.* **25**, 603–613. doi:10.1002/hyp.7848.