Review of:

hess-2011-416 (Submitted on 13 Dec 2011)

Water balance estimation in high Alpine terrain by combining distributed modeling and a neural network approach (Berchtesgaden Alps, Germany)
G. Kraller, M. Warscher, H. Kunstmann, S. Vogl, T. Marke, and U. Strasser
Handling Editor: Andrea Butturini, abutturini@ub.edu

Principal criteria:
Scientific significance (substantial contribution to scientific progress): Good (2)
Scientific quality (scientific approach and applied methods valid): Good (2)
Presentation Quality: Good-Fair (2-3)
(see technical details at the end of the review)

My overall impression (Summary):

The authors describe the application of an artificial neural network (ANN) to improve the quality of the results (i.e. modelled vs. observed water balances and discharge curves) of a distributed hydrologic model in a high Alpine catchment characterized by a karstic environment. The problems arising from unknown flow paths and flow patterns in karstic systems are comprehensively described (and they are apparent in many other model applications, too). The proposed approach to consider unknown subsurface flows by defining boundary conditions to the groundwater flow as results of an ANN instead of “simply” correcting the total runoff is new and very promising, since it allows in principle to correct for external impacts at the places of origin of these impacts and thus allowing the model to run under more realistic (subsurface) conditions – which is essential for physically based models.

On the other hand, there are two major points of criticism:
1) It was, in my opinion, not sufficiently tried to exclude effects of interpolation and precipitation correction. At least for the Königsee-subbasin, the missing runoff could have been easily created by increasing precipitation by either a larger precipitation correction (especially for snow) and/or using other interpolation methods (like a combination of Thiessen polygons with a fix “lapse rate” or gradient, as is available in the used model WaSiM-ETH since May 2010 (version 8.07). In essence, this means, that the proposed correction of subsurface storages by ANN results could simply be a correction of missing precipitation – only more sophisticated and transferred to a later stage in the model chain. To be sure about the reason for the mismatch in observed vs. modelled storage balance and runoff, the water balances of the surrounding catchments to the south and south-east must be taken into account in order to estimate the potential inflow from these sources. Is it realistic that the required amounts of water can originate from the relative small areas of sufficient elevation (to ensure a sufficient gradient) outside the Königsee catchment? This becomes even more critical when looking at the Wimbach catchment: The two neighboring catchments, the Klausbachtal and the Königsee catchment, are contained in the model domain already, so subsurface water exchanges between these subcatchments wouldn't show up in the balance of the entire basin. The Königsee catchment may loose some water to the Wimbachtal, but this effect should partly cancel out the mismatch between modelled and observed runoff in that catchment (i.e. Königsee) as well. The required additional water for the Wimbachtal catchment equals approx. 63% of the modelled runoff (table 3, last row), which is around 800mm/a. So where does this water come from? I would strongly suggest to do some model runs with larger precipitation correction in order to show if and under which circumstances the runoff balance could be modelled without considering external inflows. If this correction would be unrealistic huge, then this would support the theory of the authors.

2) The effect of the ANN application seems to be quite limited. Although there is an improvement in the modelled storages for the months May to September 2007 (fig. 12), there are no or only minor improvements or even worse results in other years (April 2007, May/June 2008, April and July 2009, May, July, August 2010) not to speak about the large deviations in November 2007.

Consequently, the discussion and the results shown in figs. 14 and 16 are focussing on 2007, the only year for which the approach works really nice.

I'm missing a comparison of the observed hydrograph against the modelled hydrographs with and without ANN-corrections (similar to fig. 5, but for several years, not only one year). This could demonstrate the effectiveness of the approach much better than the total balance, since it will show the changes in the dynamic of the hydrograph.

I think the reason for the large underestimation of modelled runoff is rather a combination of precipitation underestimation and karst effects (unknown groundwater inflows/outflows). So, the ANN correction is a promising approach but must be justified by better evidence of the real value of the mean areal precipitation.

One thought about the method: Applying constant fluxes to the boundary cells (by the way: which cells where selected by which criteria?) will work for underestimated storages only (so the storage change applied by the ANN is positive), because then the applied flux is positive and the additional groundwater will flow through the subsurface system to the rivers. But what if the storage change is negative? A negative boundary condition (negative constant flux) at the higher elevation rim of the catchment will have almost no effect (except on the cells the boundary condition was applied to). Wouldn't it be better to apply negative boundary fluxes to all cells of a catchment or at least to cells at the lower end of the aquifer? Since the ANN results may well give positive and negative corrections, this must be implemented somehow in the model in order to apply the model+ANN corrections in karst environments where larger fractions of the groundwater flow out of the basin.

My recommendation is to publish the paper after the suggested additional checks and analysis are done (try to model the runoff with higher precipitation corrections or larger precip.-gradients and estimate the possible inflow from the surrounding areas as well as showing the results for several years (not only 2007), also as compared hydrographs).

Different Aspects Details:

1. Does the paper address relevant scientific questions within the scope of HESS?
Yes, the hydrologic modelling of karst systems is still a challenge

2. Does the paper present novel concepts, ideas, tools, or data?
Yes, using an ANN to derive boundary conditions for groundwater modelling is a new, promising concept

3. Are substantial conclusions reached?
Yes, partly – as mentioned above, the results should be analysed for additional years and be compared to model results with higher precipitation corrections.

4. Are the scientific methods and assumptions valid and clearly outlined?
Yes, the ANN application is described sufficiently (general descriptions on ANN applications are referenced)

5. Are the results sufficient to support the interpretations and conclusions?
For the specific year 2007 – yes. As mentioned earlier, this should be supported by analysing the potential for groundwater inflow from the surrounding areas.

6. Is the description of experiments and calculations sufficiently complete and precise to allow their reproduction by fellow scientists (traceability of results)?
yes

7. Do the authors give proper credit to related work and clearly indicate their own new/original contribution?
yes

8. Does the title clearly reflect the contents of the paper?
Almost. Suggestion: since the water balance even in high Alpine areas usually can be estimated quite good with distributed models if there is no karst system involved, the title should contain a hint to the karstic conditions.

9. Does the abstract provide a concise and complete summary?
yes

10.Is the overall presentation well structured and clear?
Yes (technical errors are commented below)

11. Is the language fluent and precise?
yes

12. Are mathematical formulae, symbols, abbreviations, and units correctly defined and used?
yes

13. Should any parts of the paper (text, formulae, figures, tables) be clarified, reduced, combined, or eliminated?
Yes, see comments above and below.

14. Are the number and quality of references appropriate?
yes

15. Is the amount and quality of supplementary material appropriate?

n.a.

Detailed comments/technical issues:

general:

- The schema in figure 2 and the description of the method do not clarify, if the ANN is applied to each grid cell or to the sub catchments only (I assume the latter).
- If monthly mean values for T, QS and RH are used as input for the ANN, and the result (Storage) is valid for the same month: how were the monthly mean values of interpolated inputs calculated? In another model run in advance to the “real” run? Or are the results valid for the month following the month, the mean input values are valid for?
- How are the lakes modelled? Using the fully integrated lake model (with connection to unsaturated and saturated zone etc.) or with the conceptual lake model included in channel routing? There is no mention of this module in the model description
- The Wimbachtal catchment with its large groundwater storage (4 years mean retention time) indicates that the model initialization could require evenly long or even more time to make sure the internal groundwater dynamic is in an equilibrium state. There is no mention on how long the initialization period was.
- Also, the average aquifer depth (and soil depth) is not specified. I have no idea if it is tens of meters or hundreds of meters (except the legend of figure 15, suggesting a maximum aquifer thickness of 30 m). An aquifer of a few hundred meters thickness would be appropriate, according the description of the study area (p. 221, lines 1-10)
- since the ANN corrects the groundwater fluxes, the affected runoff components are rather the slower ones. Consequently, the model efficiency should be shown not only by using the (linear) NSE or r^2 (which emphasizes the peak flows) but also by using the logarithmic results (r^2 and NSE of the logarithms of the modelled runoff).
- What is the reason for using monthly time steps for the ANN? Since karst systems often react much faster, I would think of time steps of 10 days or even shorter...

Page 220, line 11: are glaciers really part of the dominant biotopes?
Page 221, line 1 and 2: something is missing here (or the “which” is unnecessary)
Page 222, line 3: what version of the model was used?
Page 222, line 8-10:

- IDW weight is 0.25 despite the strong altitude gradients of 47mm/100m?
- Why are all the other input data interpolated using IDW and not using elevation dependent regression? At least temperature and humidity in figure 15 strongly look like results of regression rather than IDW – so is it simply wrongly described?

Page 222, line 14ff: what are the main soil types?
Page 223, line 15: total unsaturated zone thickness (sum of all layers thickness’s)?
Page 224:

- was precipitation correction applied? As mentioned in the general comments, this is often crucial for good model results in high Alpine regions.
- Line 13: a good fit of a single station is not really a good criterion for an interpolation method using linear regression (since it is the essence of a regression to fit a line between observation points to minimize the overall error – usually the line touches not a single point). On the other hand, IDW will always fit for each of the stations, so a combination of REG + IDW will of course fit the single stations much better than the regression alone.
- There are other interpolation methods available in WaSiM, e.g. a method combining Thiessen polygons with fix elevation gradients (lapse rates). Try these methods (as mentioned earlier in combination with P-corrections)

Page 225, line 11/12: grammar: “we assume **the...”** (instead of that)?...
Page 226

- line 2: fig. 7 shows years 2002 to 2010, not 2001 to 2011 as stated here
- line 3: storage is “always” positive...: no, it's not (e.g. Oct. 2005). And it may also be negative in the not shown winter months
- line 4: “The monthly sums of modeled runoff show positive values”. What else should one expect?
- Line 6: negative peaks are not recognizable for Klausbachtal. Shouldn't Wimbachtal be meant here?
- Line 7/8: “there is more runoff than precipitation coming in, indicating groundwater inflow and explaining the amount of measured annual runoff.” → difference to annual runoff could also be explained by underestimated precipitation.
- line 18: figure 8 shows years 2007 to 2009, not 2007 to 2011

page 227, line 5: figure 2 is not the correct context here
page 228, line 22/23 (equations 7 and 8): explain x and y more specific
page 229,

- line 4: m=444? Should rather be 44. Also fig.12 indicates the year to range from 2007 to 2010, not to 2009 as stated in the text here
- line 7: should be October 2010 instead of October 2009
- in 2008, the ANN results are not really better than the distributed model results, and in 2010, the ANN results are even worse then the distributed model ones. This should not be kept back.
- Line 24: measure of boundary flux is m/s, not m/s per hour. (the boundary flux is applied perpendicular to the grid cell, so for a single grid cell, the boundary flux is in m^3/s with regard to the cell size)

Page 230, line 18-20: this is not really convincing. The constraint “by the given station data and interpolation method...” should be overcome by trying other interpolation methods and/or precipitation correction, as mentioned before
Page 231,

- line 3: “Evaporation is calculated from zero to over 1400...”: does this include interception? A value of zero is very uncommon.
- line 4: humidity ranges rather from 70% to 82% than from 0.7% to 0.82%
- line 17: “ ...boundary flux in m/s **per time step of one hour**”: boundary flux is only in m/s, not in m/s per hour (see comment on page 229, line24)
- Line 22: please add efficiencies for logarithms of runoff

page 240, table 3:

- is it modelled evaporation (and precipitation) or observed, e.g. by eddy covariance towers?
- Balances for gauge St. Leonhard would be nice to see the downstream effects of runoff underestimation

page 243, table 6: please add values for logarithms (R^2_{log} and NSE_{log})
page 244, fig. 1: the river network would be a nice overlay
page 245, fig. 2: typo in calculated (last but one row)
page 246:

- soil classification and geology: please use colours here. The B/W pictures are unusable
- land use classification: no recognisable difference between ice/firn and water...

page 247, fig4: please use different colours for different types of stations (daily, hourly etc.)
page 248, fig.5: such a figure would be nice for model results with applied ANN-corrections
page 249, fig.6: please specify if the model results are with or without ANN-corrections
page 250, fig.7:

- specify the used model: with or without ANN-corrections.
- The picture shows monthly sums from June to October, the legend describes is as from July to October.
- Years range from 2002 to 2010 in the figure, but from 2001 to 2011 in the description
- the pictures itself contain a legend for the three basins Hintersee, Wimbachtal and Königseetal. Hintersee is not referenced in the text legend. Is Klausbachtal meant?

Page 251, fig.8: specify: modelled results with or without ANN-corrections (same for all other figures)
page 252, fig.9: it looks quite confusing (10 grey lines...). A bold coloured line indicating the monthly mean values would be nice
page 253, fig 10:

- the style of this figure is out of line with the other figures: fat, grey lines, too fat arrows
- please explain the variables QS, Tm, RH and Sobs in the text legend

page 254, fig 11: what is the value of the correlation coefficient?
Page 257, fig 14: mismatch between Qobs in the text legend and Qmod in the picture (and Sobs,real and Sobs)
page 258, fig 15: exchange groundwater level and relative soil moisture in the text legend to match the order of the figures.