**Hydrology and
Earth System
Sciences
Discussions**

Interactive comment on "An operational hydrological ensemble prediction system for the city of Zurich (Switzerland): skill, case studies and scenarios" *by* N. Addor et al.

Dr. Werner (Referee)

micha.werner@deltares.nl

Received and published: 21 March 2011

Review HESSD-8-715-2011

An operational HEPS for the city of Zurich: skill, case studies and scenarios

Ador et al.

General comments:

This is an interesting paper – while it is presented as a case study, the work does touch in many of the basic research questions in forecast verification. I think the use of the two scenarios is very nice – while the results of these is not surprising, it does show

how crucial the operation of the lake is to the whole system, despite its command over only some 50% of the catchment. This then leads to two cost-loss functions which need to be balanced- will be interesting to hear how this work progresses.

Overall the use of English is good – but there are some occasional slips – perhaps these were added after the review by Dingwall – a quick re-read may weed out the most striking ones.

Specific comments:

Page 719: It may be good to elaborate a little more on how the Sihl passes underneath the station. I would mention explicitly that there are five culverts, two of which are closed during construction. To be honest I am not sure if "tide" gates does not confuse – perhaps better just mention that the culverts are closed by gates.

Page 720: It may also help to be more explicit on how water is normally diverted from the lake to Lake Zurich to pass through the hydropower station. Also – if a drawdown is required – is this then done through the penstock thus generating additional power due to increased flow, or is this always released into the Sihl towards Zurich? I can imagine this would influence the cost/loss ratio of the decision.

Page 722: In the discussion on the use of the COSMO-LEPS model, the first 12 hours of the model run are disregarded. This it is mentioned is close to what happens operationally. For the COSMO-7 model, however, the model run initiated at 00:00 is used directly. Under normal operational conditions this would not have been the case, as that run would not yet be available, and the 12:00 run would be used. Would it not have been more logical to apply the same reasoning to the selection of which of the two COMSO-7 models to use. This may warrant some comment as to how much difference this could make.

Page 724: The significance of large driftwood reaching Zurich is a little unclear to me. Perhaps this could be reformulated in some way.

Page 724: An approach that is often used to resolve the issue of undersampling due to the limited number of events is to use thresholds derived as exceedence quantiles in the hindcast period. Whilst these may be significantly lower than real warning thresholds and raise the question on how well skill at these lower levels extrapolates to higher levels, the issue of too few events is reduced in this way. Again, it may be good for the authors to comment on approaches taken by others (applied in work referred to), and why they chose to take the approach they have (estimating quantiles from a much longer period of record).

Page 731-732: The discussion shows that flows are generally overpredicted by the model (as shown in the rank historgrams). Perhaps it would be interesting to understand if the authors considered any methods to correct for this bias. As I understand it, this may be achieved by recalibrating the model, but this may also be detrimental (as noted). Have any statistical postprocessing methods for reducing bias been considered, such as quantile regression?

Page 734: In the discussion of the ROC diagrams, it is interesting to note that for the high thresholds Q=0.99 the skill is quite low. I think that besides the issue of FAR and Misses, that this is quite relevant for the users – in any case that depends at which threshold the decision is taken to evacuate the two closed channels and give the underpass under the station the full capacity. If this is at the lower threshold, then maybe the poor skill of the higher threshold is not that relevant (for that particular user). I guess that this exercise also is quite cheap when compared to the possible loss – this leading to the conclusion that falser alarms are more acceptable to misses.

Technical Comments (selected)

Page 725: Line 17: hence these scores Page 725: Line 19: most interest to the Page 729: Line 16: Suggest to use "small" instead of "weak" – or low discharge Page 731: Line 13-14: It could indicate a greater robustness Page 733: Line 1-2: only forecasts for days with a maximum discharge

C502

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 8, 715, 2011.