Hydrol. Earth Syst. Sci. Discuss., 9, C3378–C3382, 2012
www.hydrol-earth-syst-sci-discuss.net/9/C3378/2012/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Data-based discharge extrapolation: estimating annual discharge for a partially gauged large river basin from its small sub-basins" *by* L. Gong

Anonymous Referee #2

Received and published: 27 July 2012

General Comments

This manuscript attempts to offer a potentially useful and interesting approach to data-driven hydrology, but the sometimes confusing presentation obscures what seems to be a straightforward underlying concept. The approach appears to be elegantly elementary statistics: choose a representative sample (sub-basins) that mimics the behavior of the broader population (the encompassing basin). The sample criteria are precipitation P and potential evapotranspiration PET; the sought behavior is annual discharge. The mean discharge of the representative sub-basin sample is shown to compare favorably with discharge estimates of the entire gauged portion of the basin.

Unfortunately, this main concept, if I have it correctly, is buried within an overly lengthy introduction, an unfocused theory section, a methods section with elements that would be more appropriate for the theory section, insufficient description of the primary procedure (scale-extrapolation), and insufficient space given to results and discussion. These shortcomings are compounded by grammar errors, confusing or ambiguous terminology, as well as figures and writing that are not essential to the core concepts. In short, this manuscript needs more focus on the main idea, less peripheral information, and more clarity overall. The author should develop a more coherent narrative to guide the reader to the main conclusions and rely less on section numbering that jumps suddenly between distinct topics.

With that said, and assuming that I correctly understand the main ideas, methods, and conclusions, the substance of this manuscript has potential to offer a meaningful contribution to data-driven hydrology. I recommend that it should be considered for publication after major revisions, and I would be glad to review the revised manuscript.

âĂČ Specific Comments

1. The term “scale-extrapolation” is not defined precisely even though it is central to the manuscript. You certainly extrapolate discharge measurements from small regions to a larger region, but I’m not clear on what the combined term is supposed to mean. It suggests to me that you’re extrapolating some scale. I assume you have a very specific definition in mind, but left to my own imagination the terms seems redundant. 2. How is “annual” defined? Jan – Dec, or as a water year such as Oct – Sep? 3. Starting in the abstract and throughout the paper you write about the potential for this method to bracket uncertainty of discharge estimates, but you don’t actually do this in the manuscript. I can imagine how this might be done, but the idea doesn’t deserve so much attention unless you show it in practice. 4. The introduction section is quite dense. The information presented is mostly relevant, but a more concise version with

paragraphs to organize the various sub-topics would be a favor to the reader. 5. In the dataset section, paragraph 1, your validation of the CRU precipitation data is somewhat confusing. You argue that comparison with the SHMI data (presumably only in Sweden vs the entire Baltic Basin?) shows that the CRU data can be considered reliable even after 1990 because the remaining CRU stations still captured spatial variability of annual precipitation, at least at the scale of a 30-min grid. But certainly you don't mean that the reduction in CRU gauges could have had an effect on the precipitation itself, as suggested by your phrasing in the sentence on lines 11-13. This should be clarified. 6. In the last paragraph of the dataset section, my confusion over basin categorization and terminology began. This section tells me that 100 gauged sub-basins are in the study. I suggest that you define the "gauged basin area" here rather than in 4.1 so that the meaning of the "active sub-basins" can be treated separately. The difference is simple enough, but it took a pen and paper for me to keep track of which set of basins you were referring to at different points in the manuscript. 7. The theory section gives several good justifications for your data-driven approach, but it lacks a unifying theme, such as the simple statistical concept that a representative sample set can mimic the behavior of the population. Also, the acknowledgement of Budyko is necessary, but the way it was written I was initially concerned that you were going to use that sort of long-term partitioning analysis to drive your yearly estimates. I was glad that you didn't. 8. My feeling is that a revised, more concise and focused version of sections 4.2 and 4.3 should be in the Theories section. Sections 4.2 and 5.1 present a proof of principle that a small sample set can be used to represent the climate of the broader region. This is important, but it has too much space in the paper and it repeatedly interrupts the main story about estimating discharges. Section 4.3 undoubtedly belongs in the Theories section. Over halfway into the section you explain that Budyko's theory and equations aren't used in your method. I reach the opposite conclusion as you from Figure 1b – that Budyko's equation isn't very good at predicting interannual partitioning (and isn't meant to be). The purpose, method, and outcome of the theoretical test with equation (3) are not obvious. 9. The assumption that each cell is an independent nonlinear unit is too important to be buried deep in section 4.3. 10. Section 4.4 needs clarification. An illustrated example of the procedure for one sub-basin would be very useful. I tried to make my own illustration of the method you describe, but I couldn't quite make sense of step 4. If discharge is normalized by the sub-basin area so that it is a runoff depth averaged over the sub-basin, then why is it necessary to take a weighted mean of the discharge for the area overlapping the selected cells and the sub-basin? Is the broader basin discharge (runoff) taken to be the average of the runoff for all of the sub-basins selected during steps 1 through 3? Or am I misunderstanding the method? Either way, elaboration and clarification are needed. 11. The relationship between nonlinearity and bias observations in Section 5.2 is not clear to me. Couldn't we also see bias in a linear system if the cells were not representative of the basin climate? 12. In the first paragraph of section 5.3, I don't know where this new list of 77 sub-basins came from. I had 100 sub-basins in mind from the gauged area, then 51 sub-basins in the active area, and now 77 sub-basins are being discussed. 13. The discharge estimates are impressive and unexpected. I wouldn't have guessed that P and PET would be sufficient to provide such good results. In particular, I would expect interannual storage in snowpack to be a larger factor, but this seems to be at least partially accounted for by the two climate variables. 14. As previously mentioned, I would like to see more discussion of uncertainty, as well as a further exploration of whether it's obvious that this method should work or if there is something surprising about the results. 15. Does the method work at smaller scales, e.g. to predict discharge of a mid-size basin using embedded sub-basins? 16. What are the results if only basin-wide similarity is used as cell selection criteria, rather than imposing variation similarity within each sub-basin as well? Would this result support or diminish the importance of the non-linearity of individual basins? 17. A more robust discussion of the statistics would be helpful, for instance how lack of independence between climate cells might impact the results, and how the observed residuals compare with theoretical expectations for a given sample size.

Technical Corrections

There are many grammatical and idiomatic errors, the assistance of a good editor would be useful for the next revision.

Page 1 Line 2: you probably mean “so far”, but you certainly don’t want to use the very subjective “by far” to start your paper.

P1 L3: “accessing” – do you mean “assessing”?

P1 L22 (and elsewhere): “recourse” should be “resource”. I didn’t note anywhere that “recourse” would be the appropriate word.

P2 L8: it is surprising to me that as much as 50% of the world’s land area might be gauged, this seems very high.

P4 L9: Should be “(figure not shown)”?

P4 L17: WATCH is not defined.

Equation (1) should have variables defined.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 9, 6829, 2012.