

Interactive comment on “Predictability, stationarity, and classification of hydraulic responses to recharge in two karst aquifers” by A. J. Long and B. J. Mahler

A. J. Long and B. J. Mahler

ajlong@usgs.gov

Received and published: 2 November 2012

Authors' responses to review comments by Anonymous Referee #2, RC C4922

Responses to general comments (overall message): (1) Predictability: Yes, this is a good point. Quantitative information on goodness of fit will be added to the abstract and conclusions. (2) Stationarity: This is an important point. A summary of stationarity of the sites will be added to the abstract and conclusions. (3) Classification: Yes, we can be more specific on classification in the abstract, conclusions, and in the body. Also, see response to comment on section 3.5.

C5082

Specific comments and responses:

-p9588, line 1-2: how did you determine recharge areas for the different sites?

Authors' response: The general recharge areas correspond to the formation outcrops and are shown on figure 2 (study areas). The specific recharge area for each site was assumed to be the outcrop area directly upgradient from the site. The length of the recharge area (measured parallel to the formation outcrop) is not precisely defined, but we assume that the primary influence on the response at each site does not extend a long distance along the length of the outcrop. A description of this will be added in a revision.

-p9588, line 9; remove "temperature"?

Authors' response: Yes, thanks for catching this error.

-section 3.2: several sites included recharge to the aquifer from local streams in addition to direct precipitation. Description of how this was modeled should be added to the recharge modeling section (2.1) which now only deals with direct precipitation.

Authors' response: Good point. We'll add that here.

-p9590, line 16: note that Nash values can also be negative.

Authors' response: Agree. This will be added.

-p9590, line 22-25: this does not make sense to me. By using the entire observation period (wet plus dry) in the denominator but only dry or wet period in the numerator you artificially inflate the Nash values.

Authors' response: Actually, we calculated the Nash values for calibration and validation periods separately (not wet and dry periods). Either way, the reviewer is correct. However, what should have been explained in the text is that the denominator was scaled down by multiplying it by the ratio of the time length of the calibration or validation period to the length of the full period. As a check, when we calculated C-nash

C5083

for the full period, it was between the values for the calibration and validation periods. This will be clarified in a revised manuscript.

-p9591, line 25: the Nash value for the validation period is likely a too optimistic estimate of future Nash/error values because model selection was performed on the validation period (which then basically acts as a "calibration" period for picking the right model structure). A better estimate of future errors can be made by testing the final model structure on a third independent data period.

Authors' response: Yes, this is true. This will be mentioned in a revision

-Instead of Nash, why not use RMSE (root mean square error), which provides a more direct measure of modeling errors?

Authors' response: We chose to use the Nash criterion because it provides a comparison of the magnitude of residuals (errors) to the overall amplitude of fluctuation. We think this is more meaningful for time-series, especially when we are comparing sites with different output amplitudes. For example, if you had one spring that fluctuated by 2000 L/s and another that fluctuated by only 100 L/s, and each spring had an RMSE of 10 L/s, we think that the first spring is a better fit because the error is a smaller fraction of the variability. But RMSE says the two models are equally good. We think that C-nash is a better gage of how well the model fits the fluctuations in the measured time series. Also, C-nash is useful for comparing a model of head to a model of spring flow; RMSE would not be comparable.

-section 3.5: discussion of the classification results would benefit from making a link with physical/hydrological characteristics of the sites; in other words, do the results confirm physical knowledge of these sites (e.g. flow vs head site; insights from previous studies etc).

Authors' response: This is an excellent point and one that deserves a major revision of section 3.5. Upon further analysis, we can make several points to address this.

C5084

Because the differences between the wet- and dry-period IRFs seemed to be very informative, we have added an additional metric to better describe these differences. Several inferences can now be made to relate the revised PCA results to the physical systems.

The general curve type, either lognormal or exponential, that primarily composed each of the IRFs was the primary factor in relating the sites' IRFs to hydrogeologic factors. The two major categories, exponential and lognormal, have different plotting characteristics in PCA and also are separated spatially within the two aquifers. For both aquifers, hydrogeologic differences exist between these two areas. The Edwards aquifer is wide in the area occupied by the lognormal category, where groundwater flow is circuitous; the aquifer is narrow in the area occupied by the exponential category, where groundwater flow is generally straight. The circuitous flow path might be related to the delayed response defined by lognormal IRFs. For the Madison aquifer, the area occupied by the lognormal category is near the axis of the main anticline that defines the Black Hills uplift, where fracturing and subsequent development might be related to IRF characteristics.

Differences in springs and wells also were considered. Springs commonly discharge water from a large surrounding area, whereas none of the wells used in this study are pumped. Despite these differences, there is no apparent distinction between springs and wells that is evident from PCA (figs A & B), which indicates that the difference between springs and wells is small compared with regional differences in the aquifers.

On the basis of revised PCA results, the Madison aquifer seems to have larger heterogeneity than the Edwards aquifer. First, the separation in principal component space of Madison aquifer sites, which is a measure of IRF differences, is significantly larger than that of the Edwards sites based on a rank-sum statistical test. Second, the Madison aquifer has larger differences between wet- and dry-period IRFs than does the Edwards aquifer. Third, the Madison aquifer has all four of the double-peaked IRFs, which suggests greater differences between conduit and fracture porosity for the Madi-

C5085

son.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 9, 9577, 2012.

C5086