# Comment on "Climate and agricultural land use change impacts on streamflow in the upper midwestern United States" by Satish C. Gupta et al.

**Patrick Belmont[1], John R. Stevens[2], Jonathan A. Czuba[3], Karthik Kumarasamy[1], and Sara A. Kelly[1]**

[1]Department of Watershed Sciences, Utah State University, Logan, Utah, USA, [2]Department of Mathematics and Statistics, Utah State University, Logan, Utah, USA, [3]Department of Civil, Environmental and Geo-engineering, St. Anthony Falls Laboratory and National Center for Earth Surface Dynamics, University of Minnesota, Minneapolis, Minnesota, USA

**Abstract** The paper "Climate and agricultural land use change impacts on streamflow in the upper midwestern United States" by Satish C. Gupta, Andrew C. Kessler, Melinda K. Brown, and Francis Zvomuya (hereafter referred to as Gupta et al.) purports to evaluate "the relative importance of changes in precipitation and LULC (land use, land cover) on streamflow in 29 Hydrologic Unit Code 008 watersheds in the Upper Midwestern United States." However, as we report here, the approach used by Gupta et al. is wholly inadequate for making such an evaluation. Gupta et al. use strong language to criticize other studies and imply a level of certainty that goes well beyond, and in some cases is entirely unsupported by, the results they have presented. We take this opportunity to point out several critical flaws in their study.

## 1. The Statistical Tests Gupta et al. Use to Examine Changes in Precipitation-Streamflow Relationships are Wholly Inadequate for Identifying Effects of Land Use and Artificial Drainage

### 1.1. Due to High Variability and Small Sample Size, the Gupta et al. Analysis has Very Little Power to Detect What Appears to be a Large Change in Hydrology, Rendering it Highly Susceptible to a Type II Error

Precipitation-streamflow relationships naturally exhibit considerable variability. This variability, combined with relatively short records (typically ∼ 80 years), provides very little statistical power to identify changes in slope or intercept of the relationships, and thus increases the risk of type II error (probability of not detecting a change that is real and substantial). In cases where statistical inference directly affects environmental decisions, public welfare, or the economic interests of advocacy groups (such as agriculture) the implications of committing a type II error (in this case, failing to identify the hydrologic change caused by land use and agricultural drainage) may be egregious [*Hoenig and Heisey*, 2001; *Rosner et al.*, 2014]. Statistical power is defined as the probability that an effect (e.g., change in slope of the precipitation-streamflow relationship) could be detected, if it exists within a given data set. Generally statistical power of 80% is considered acceptable.

As a representative example, we fit Gupta et al.'s Model 1 using data from the Blue Earth watershed (including the Watonwan watershed, from USGS and PRISM databases) and verified general assumptions of the model (see Supporting Information S1 for methods). Figure 1a shows the data and replicated analysis for the Blue Earth watershed (with difference between slopes of 0.0005), where the statistical power to detect such a difference is only about 7%. In other words, if the difference in slopes identified by Gupta et al. is in fact real, the very limited data set evaluated by Gupta et al. would only have a 7% probability of detecting it. Additionally, we fit Gupta et al.'s Model 2 (that allows different intercepts but the same slopes for the two periods) but this time focusing on various sizes of $\beta_6$ (the shift between the two periods). The statistical power to detect the actual observed difference as significant was about 35%. This may explain why 10 of 29 watersheds are found to have significant period shifts (based on $\beta_6$ in Gupta et al.'s Model 2).

Low statistical power should be expected for relationships found to be nonsignificant [*Hoenig and Heisey*, 2001], and would not necessarily be a matter of concern if the observed change in the precipitation-streamflow relationship was inconsequential. However, the apparent change in slope represents a nearly
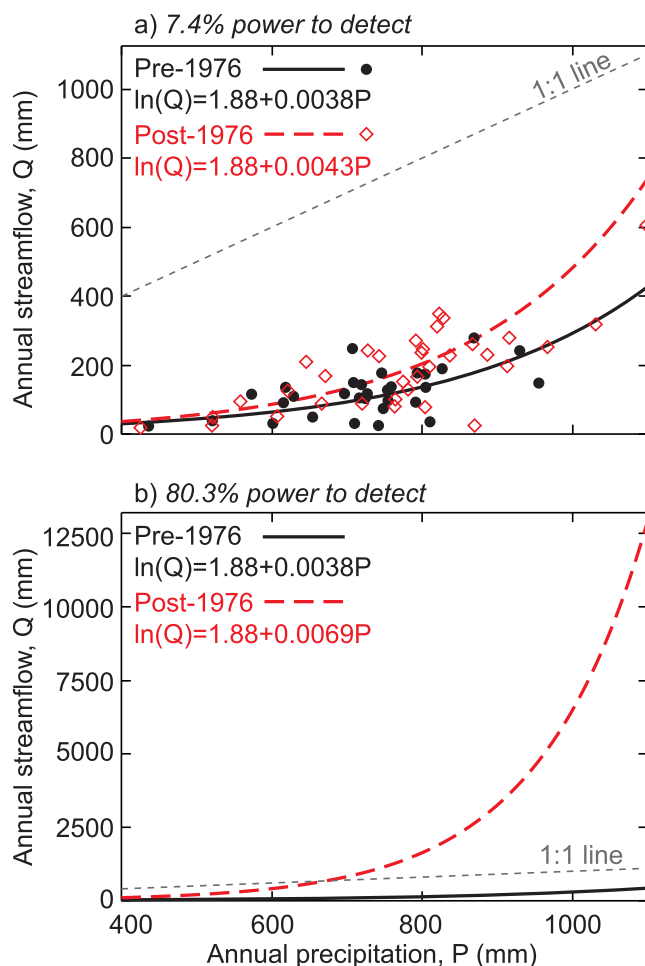
**Figure 1.** Power analysis for slope differences in the Blue Earth River data. (a) The post-1976 line is the fit from the actual data, and (b) the relationship necessary to achieve approximate 80% power to detect any slope difference from the pre-1976 period.

50% increase in annual streamflow for a year with 800 mm of precipitation, a value that is exceeded frequently in both time periods evaluated by Gupta et al. and even this 50% increase in annual streamflow understates the potential implications. For example, 10% exceedance flows have increased by 65% between the two time periods. This large increase in high flows translates into even greater increases in erosion and sediment transport, especially in the highly dynamic geomorphic settings of southern Minnesota [*Belmont*, 2011; *Gran et al.*, 2013; *Stout et al.*, 2014; *Schaffrath et al.*, 2015]. The large hydrologic changes observed throughout the upper Midwest have far reaching implications and demand rigorous analyses that link cause and effect at the appropriate time and space scales in order to develop effective strategies to improve water quality.

Gupta et al. are quick to interpret the failure to reject the null hypothesis as proof of no change in the precipitation-streamflow relationship. However, in a world increasingly characterized by nonstationary hydrology it is essential that we move beyond such rudimentary null hypothesis testing and critically evaluate the probability and potential costs associated with committing both type I (over-design) and type II (under-prepare) statistical errors [*Rosner et al.*, 2014].

## 1.2. Given the Small Sample Size and Large Natural Variability in the Data Set, to Obtain an Acceptable Level of Statistical Power Would Require an Impossibly Large Shift in the Relationship

To determine how big of an increase in the precipitation-streamflow relationship would be required to obtain an acceptable level of statistical power, we used the probability distribution method [*Gbur et al.*, 2012] and artificially modified the slope of the relationship in increments of 0.0005 (see Supporting Information S1 for explanation of methods). Figure 1b shows that it would require a very large effect size (a difference in slope of $\beta_3 = 0.0031$) before the power comes close to generally acceptable levels (such as 80%). More specifically, in order to have 80% statistical power to detect any difference between the two periods' slopes (i.e., discern land use effects on the precipitation-streamflow relation) in the Blue Earth watershed using Gupta et al.'s method, the difference would need to be so large that streamflow would need to exceed precipitation in the post-1976 period. In other words, given the small sample size and high variability in the data only a physically impossible difference in period slopes would be detectable with an acceptable level of statistical power.

We note that the slope difference ($\beta_3$) between periods is not statistically significant for any watershed in Gupta et al.'s Model 1. However, the lack of a significant result (at $\alpha = 0.05$) in the observed data cannot be considered proof of no slope difference between the two periods, as the nature of variability in the precipitation-streamflow relationship is so high that even an otherwise appropriate statistical model is simply underpowered to detect any difference within the realm of physical possibility (Figure 1b).

**Table 1.** Sensitivity of Probability Values of Regression Coefficients Relating ln(Q) Versus P in Two Hierarchical Models Used by *Gupta et al.* [2015]

| USGS Streamflow Gage (Site Number) | p Values | | | | | | |
| | Model 1[a] | | Model 2[b] | | | | |
| | $\beta_2$ | $\beta_3$ | $\beta_6$ | Record Years | Breakpoint[c] | B\|A Years[d] | Comments |
|---|---|---|---|---|---|---|---|
| Le Sueur River near Rapidan, MN (05320500) | 0.98 | 0.78 | 0.08 | 1940–2009 | 1975 | 36\|34 | same as in Table 2, |
| | 0.90 | 0.91 | 0.18 | 1942–2009 | 1975 | 34\|34 | *Gupta et al.* [2015] |
| | 0.88 | 0.92 | 0.14 | 1941–2010 | 1975 | 35\|35 | |
| | 0.78 | 0.95 | 0.04[e] | 1940–2011 | 1975 | 36\|36 | |
| | 0.82 | 0.90 | 0.04[e] | 1940–2014 | 1975 | 36\|39 | |
| Blue Earth River near Rapidan, MN (05320000) | *Exclude Watonwan precipitation* | | | | | | |
| | 0.83 | 0.64 | 0.14 | 1940–2009 | 1975 | 36\|34 | same as in Table 2, |
| | 0.85 | 0.70 | 0.25 | 1942–2009 | 1975 | 34\|34 | *Gupta et al.* [2015] |
| | 0.82 | 0.67 | 0.22 | 1941–2010 | 1975 | 35\|35 | |
| | 0.98 | 0.75 | 0.08 | 1940–2011 | 1975 | 36\|36 | |
| | 0.96 | 0.73 | 0.07 | 1940–2014 | 1975 | 36\|39 | |
| | *Include Watonwan precipitation* | | | | | | |
| | 0.83 | 0.63 | 0.11 | 1940–2009 | 1975 | 36\|34 | |
| | 0.90 | 0.74 | 0.21 | 1942–2009 | 1975 | 34\|34 | |
| | 0.88 | 0.70 | 0.19 | 1941–2010 | 1975 | 35\|35 | |
| | 0.98 | 0.74 | 0.06 | 1940–2011 | 1975 | 36\|36 | sensitivity discussed in text |
| | 0.94 | 0.70 | 0.06 | 1940–2014 | 1975 | 36\|39 | |
| Whetstone River near Big Stone City, SD (05291000) | 0.34 | 0.53 | 0.10 | 1932–2009 | 1975 | 44\|34 | as in Table 2, *Gupta et al.* [2015]; |
| | 0.67 | 0.91 | 0.11 | 1932–2009 | 1975 | 44\|34 | data not filtered |
| | 0.51 | 0.42 | 0.45 | 1942–2009 | 1975 | 34\|34 | filtered data[f] |
| | 0.48 | 0.37 | 0.33 | 1941–2010 | 1975 | 35\|35 | |
| | 0.52 | 0.37 | 0.22 | 1940–2011 | 1975 | 36\|36 | |
| | 0.60 | 0.37 | 0.03[e] | 1940–2011 | 1990 | 51\|21 | |
| | 0.69 | 0.55 | 0.33 | 1972–2009 | 1990 | 19\|19 | |
| | 0.58 | 0.37 | 0.09 | 1967–2014 | 1990 | 24\|24 | |
| | 0.91 | 0.59 | 0.01[e] | 1932–2014 | 1990 | 59\|24 | filtered data |
| | 0.57 | 0.33 | 0.02[e] | 1940–2014 | 1990 | 51\|24 | |

[a]$\ln(Q_{all}) = \beta_0 + \beta_1 P_{all} + \beta_2 I_B + \beta_3 P_{all} I_B$.
[b]$\ln(Q_{all}) = \beta_4 + \beta_5 P_{all} + \beta_6 I_B$.
[c]Breakpoint referred to here is the last year included in the before-change period.
[d]B\|A years refers to the number of years before (B) and after (A) the breakpoint, respectively.
[e]Intercept significantly different from zero at the 5% level.
[f]Years between 1932–1939 with >10% missing daily streamflow values were not included in the analysis.

### 1.3. The Statistical Methodology is Not Robust to Slight Changes (a Few Years) in the Time Period Used in the Analysis

The start date of each record used in Gupta et al.'s analyses was arbitrarily selected as when the streamflow gage began operating, which varied from 1902 to 1946. For the Le Sueur watershed, the period used in Gupta et al.'s analysis was 1940–2009, B\|A=36\|34 years (shorthand for 36 years in the before change period and 34 years in the after change period), for which they reported a p value for the coefficient $\beta_6$ of model 2 (referred to as the $\beta_6$ p value for shorthand) of 0.08 and interpreted this as no significant change at the 5% level from LULC change in annual precipitation versus streamflow relationship. However, when we fix the before and after change period lengths to the same number of years, depending on the number of years used, we obtain substantially different values that change the conclusions. For instance, when using the period 1940–2011 (B\|A=36\|36 years), the p value on $\beta_6$ changes to 0.04 (Table 1), which means the coefficient $\beta_6$ is now statistically significant at the 5% level suggesting that LULC change has affected the annual streamflow-precipitation relationship according to Gupta et al.'s methodology. We observed this level of sensitivity (roughly +/- a factor of two) of the $\beta_6$ p value in the other two watersheds we analyzed (Table 1).

### 1.4. The Statistical Methodology is Not Robust to the Systematic Removal of a Single Point

We further assessed the sensitivity of the $\beta_6$ p value by systematically removing a single data point and repeating the analysis. This analysis was performed for the Blue Earth watershed for the period 1940–2011

(B|A=36|36 years; we also included the Watonwan HUC 8 watershed precipitation data as discussed in section 1e). The $\beta_6$ p value for this period without removing a single point was 0.06, but varied from 0.01 to 0.12 (again roughly +/- a factor of two) when removing only a single data point from the analysis. In seven instances, removal of a single point brought the p value below the 5% level. Gupta et al.'s highly sensitive statistical methodology does not inspire confidence in $\beta_6$ p values around 0.10 (within a factor of two of 0.05) as being statistically insignificant at the 5% level.

### 1.5. It Appears That Gupta et al. Excluded a Large Portion of the Blue Earth Watershed Precipitation Data, While Using Streamflow Data for the Entire Watershed

The precipitation data used for comparison with the streamflow gage on the Blue Earth River at Rapidan, MN (USGS 05320000) must account for the Blue Earth HUC8 watershed as well as the Watonwan HUC8 watershed because both are upstream of the gage and are included in the USGS streamflow measurements. We were only able to exactly reproduce Gupta et al.'s results for the Blue Earth watershed when excluding the Watonwan HUC8 watershed precipitation data but we had no problem exactly reproducing Gupta et al.'s results for the Le Sueur watershed. After weighted averaging (by drainage area) the precipitation data from the Blue Earth (64%) and Watonwan (36%) HUC8 watersheds, we observed that the $\beta_6$ p values decreased by 0.03 (Table 1).

### 1.6. The Relevant "Breakpoint" for the Whetstone Watershed is 1990, Not 1975

*Foufoula-Georgiou et al.* [2015] showed that a more meaningful hydrologic breakpoint (i.e., year before/after which a hydrologic change due to land use/land cover change should be tested) for this watershed is 1990, based on analysis of land cover change (transition from corn and small grain to corn and soybean) as a surrogate for the intensification of agricultural drainage. We understand that this paper was published after Gupta et al.'s and that they were not aware of this finding, but we mention this here to show that by changing the breakpoint, the $\beta_6$ p value can range between 0.01 and 0.33, depending on the number of years considered (Table 1). Note that annual streamflow data that were missing >10% of their daily values were removed before performing the analysis. Missing streamflow data (not estimated) are common before 1940 and for the Whetstone River between 1932 and 1940 up to 40% (in 1934) of the streamflow record was missing in a given year. We observed a much larger sensitivity of this watershed compared to the others considered, due to the fewer number of data points in the after-change period from shifting the breakpoint. The key point here is that the methodology is not robust as you can obtain a large range of p values just by arbitrarily considering different lengths of record even when the breakpoint is based on relevant land use/ cover information.

## 2. The Annual Timescale Evaluated by Gupta et al. Fails to Resolve the Myriad Hydrologic Changes Caused by Human Alterations of Agricultural Systems in the Upper Midwest

Most of Gupta et al.'s conclusions rely on analysis of annual streamflow and precipitation data. The annual timescale obscures the impacts of land cover, crop type and artificial drainage, which are known to vary over event to seasonal timescales [*Lytle and Poff*, 2004; *Schilling and Helmers*, 2008; *Schottler et al.*, 2014; *Foufoula-Georgiou et al.*, 2015]. Analysis of how, specifically, humans have altered the hydrologic system of agricultural watersheds requires greater specificity regarding the processes or streamflows of interest and the time periods for which those processes or streamflows occur. For example, change in mean annual streamflow cannot readily be translated into changes in erosive power, sediment transport, nutrients and riparian dynamics, all of which have been identified as critically altered in this system [*Wilcock*, 2009; *Belmont et al.*, 2011; *Gran et al.*, 2011; *Lenhart et al.*, 2013].

The fact that the annual scale obscures the more critical, higher frequency impacts of artificial drainage, such as changing peak flows and hydrograph shape, was made clearly by *Foufoula-Georgiou et al.* [2015] and was also discussed by *Schottler et al.* [2014]. To further illustrate this point, we utilize results from a well calibrated SWAT (Soil and Water Assessment Tool) model we developed for the 2880 km$^2$ Le Sueur watershed. For the sake of brevity, we only briefly describe in Supporting Information the development, calibration, validation and application of the 175 subbasin model.
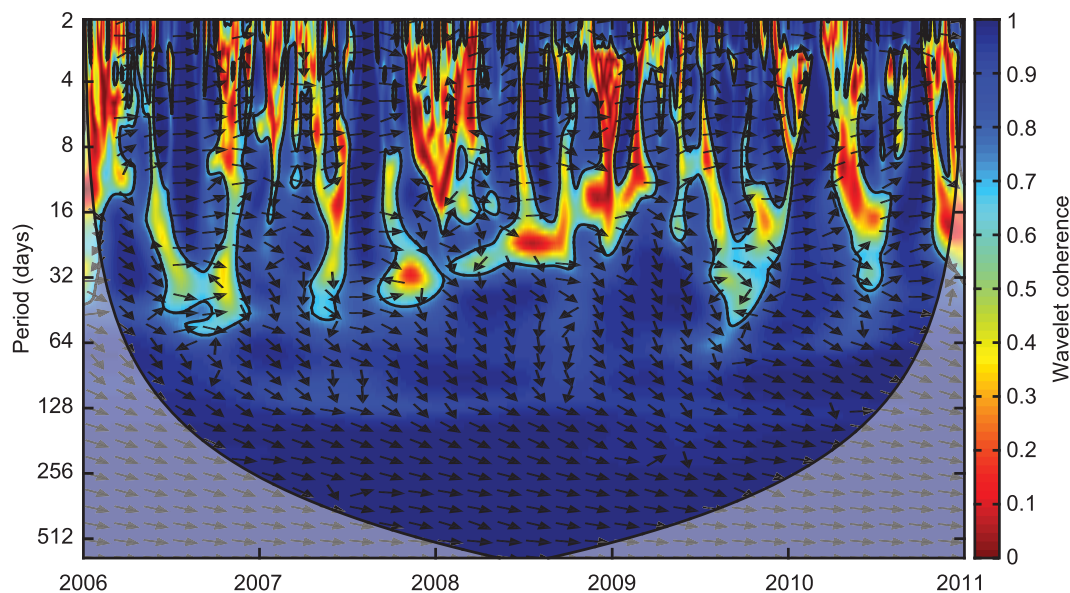
**Figure 2.** Wavelet coherence (a parameter similar to traditional correlation coefficient) between the tile versus no-tile scenarios. High coherence areas are denoted with cool colors, low coherence is denoted with hot colors. Contour lines shown as thick black lines represent the 95% confidence level for the red noise (provides confidence in coherence estimates and differentiates from signal over Poisson processes) and the region where edge effects introduced by finiteness of the signal affects interpretation is shown in a lighter shade. The vectors denoted by the arrows indicate the phase difference with in-phase relationship indicated by a right-pointing arrow and anti-phase indicated with a left-pointing arrow. A downward pointing vector means the first signal leads the second signal by one quarter of the periodicity (half the difference between in-phase and anti-phase, i.e., 0.25 days; and so forth) while an upward pointing vector means the second signal leads the first signal by one quarter periodicity. The horizontal axis shows time in days and the vertical axis is the wavelet period in days.

Using a time-frequency analysis via wavelets, we have examined the difference between the two scenarios (tile versus no tile) for different timescales (Figure 2, with 1 period ≈ 1 day). We applied the wavelet coherence technique described in *Grinsted et al.* [2004] and used the Mortlet wavelet for feature extraction [*Foufoula-Georgiou and Kumar*, 1995] to quantify where and how the hydrograph changed with the simulated removal of tiles. Highly continuous coherence areas (cool colors) for long duration (periods> 64 days) indicates strong correlation between the two signals. In contrast, the dominance of low coherent features (hot colors) for shorter periods (<30 days) indicates differences between the tile versus no tile scenarios, as rising and falling limbs of hydrographs are considerably steeper in the tiled scenario. Thus, these results indicate that artificial tile drainage most strongly affects submonthly hydrograph timescales (steepening rising and falling limbs of hydrographs) and is imperceptible at annual timescales. This modeling result is entirely consistent with the joint quantile-quantile analysis of *Foufoula-Georgiou et al.* [2015] based on Copulas, so any concerns regarding the ability of SWAT to simulate the relevant hydrologic processes would also need to argue against the analysis of physical data presented therein.

## References

Belmont, P. (2011), Floodplain width adjustments in response to rapid base level fall and knickpoint migration, *Geomorphology*, *128*(1-2), 92–102.

Belmont, P., et al. (2011), Large shift in source of fine sediment in the Upper Mississippi River, *Environ. Sci. Technol.*, *45*, 8804–8810, doi:10.1021/es2019109.

Daggupati, P., H. Yen, M. J. White, R. Srinivasan, J. G. Arnold, C. S. Keitzer, and S. P. Sowa (2015), Impact of model development, calibration and validation decisions on hydrological simulations in West Lake Erie Basin, *Hydrol. Processes*, *29*(26), 5307–5320, doi:10.1002/hyp.10536.

Foufoula-Georgiou, E., and P. Kumar (1995), *Wavelets in Geophysics*, vol. 4, Academic, San Diego, Calif.

Foufoula-Georgiou, E., Z. Takbiri, J. A. Czuba, and J. Schwenk (2015), The change of nature and the nature of change in agricultural land-scapes: Hydrologic regime shifts modulate ecological transitions, *Water Resour. Res.*, *51*, 6649–6671, doi:10.1002/2015WR017637.

Gbur, E. E., W. W. Stroup, K. S. McCarter, S. Durham, L. J. Young, M. Christman, M. West, and M. Kramer (2012), *Analysis of Generalized Linear Mixed Models in the Agricultural and Natural Resources Sciences*, Am. Soc. of Agron., Soil Sci. Soc. of Am., and Crop Sci. Soc. of Am., Madison.

Gran, K. B., P. Belmont, S. S. Day, N. Finnegan, C. Jennings, J. W. Lauer, and P. R. Wilcock (2011), Landscape evolution in south-central Minnesota and the role of geomorphic history on modern erosional processes, *GSA Today*, *21*(9), 7–9.

Gran, K. B., N. Finnegan, A. L. Johnson, P. Belmont, C. Wittkop, and T. Rittenour (2013), Landscape evolution, valley excavation, and terrace development following abrupt postglacial base level fall, *GSA Bull.*, *125*(11-12), 1851–1864.

Grinsted, A., J. C. Moore, and S. Jevrejeva (2004), Application of the cross wavelet transform and wavelet coherence to geophysical time series, *Nonlinear Processes Geophys.*, *11*(5/6), 561–566, doi:10.5194/npg-11-561-2004.

Gupta, S. C., A. C. Kessler, M. K. Brown, and F. Zvomuya (2015), Climate and agricultural land use change impacts on streamflow in the upper midwestern United States, *Water Resour. Res.*, *51*, 5301–5317, doi:10.1002/2015WR017323.

Hoenig, J. M., and D. M. Heisey (2001), The abuse of power: The pervasive fallacy of power calculations for data analysis, *Am. Stat.*, *55*(1), 19–24.

Lenhart, C. F., M. L. Titov, J. S. Ulrich, J. L. Nieber, and B. J. Suppes (2013), The role of hydrologic alteration and riparian vegetation dynamics in channel evolution along the lower Minnesota River, *Trans. Am. Soc. Agric. Biol. Eng.*, *56*(2), 549–561, doi:10.13031/2013.42686.

Lytle, D. A., and N. L. Poff (2004), Adaptation to natural flow regimes, *Trends Ecol. Evol.*, *19*(2), 94–100.

Rosner, A., R. M. Vogel, and P. H. Kirshen (2014), A risk-based approach to flood management decisions in a nonstationary world, *Water Resour. Res.*, *50*, 1928–1942, doi:10.1002/2013WR014561.

Schaffrath, K. R., P. Belmont, and J. M. Wheaton (2015), Landscape-scale geomorphic change detection: Quantifying spatially variable uncertainty and circumventing legacy data issues, *Geomorphology*, *250*, 334–348.

Schilling, K. E., and M. Helmers (2008), Effects of subsurface drainage tiles on streamflow in Iowa agricultural watersheds: Exploratory hydrograph analysis, *Hydrol. Processes*, *22*, 4497–4506.

Schottler, S. P., J. Ulrich, P. Belmont, R. Moor, J. W. Lauer, D. R. Engstrom, and J. E. Almendinger (2014), Twentieth century agricultural drainage creates more erosive rivers, *Hydrol. Processes*, *28*(4), 1951–1961.

Stout, J. C., P. Belmont, S. P. Schottler, and J. K. Willenbring (2014), Identifying sediment sources and sinks in the Root River, Southeastern Minnesota, *Ann. Assoc. Am. Geogr.*, *104*(1), 20–39.

Wilcock, P. R. (2009), Identifying sediment sources in the Minnesota River Basin, in *Minnesota River Sediment Colloquium Report*, Minn. Pollut. Control Agency, pp. 1–18, St. Paul, Minn. [Available at https://www.pca.state.mn.us/sites/default/files/wq-b3-43.pdf, accessed online September 18, 2016.]