

Inference and Attribution in Watershed Hydrology: A Discussion of *Gupta et al. [2015]*

FRST 590: Statistical Methods in Hydrology

Submitted: 29 March 2019 Prepared by: Dan Kovacek (35402767)

1.0 Motivation

A commentary on the current state of research in the effect of changing land use and land cover (LULC) on streamflow and floods at the catchment scale is presented in *Rogger et al. [2017]*. In the process of delineating gaps in the existing research, the authors describe the need for new approaches to obtain more general statements on impacts, citing the regularity with which studies obtain contradictory results for the same *kind of change*, or intervention. *Rogger et al. [2017]* highlights two such studies:

“Some recent publications such as the paper of Gupta et al. [2015] on the relative impacts of climate and land use changes on streamflow or that by Alila et al. [2009] about the effects of forest practices on floods have triggered scientific debates with the results being criticized by many scientists.” (Rogger et al. 2017)

To gain more quantitative insights into the impacts of LULC on hydrological trends, perhaps new quantitative approaches are needed, as *Rogger et al. [2017]* argues. A clearer understanding of the distinguishing characteristics and appropriate use of existing approaches may be equally valuable. *Cox [2006]* argues the translation of a subject-matter problem into a formal statistical question is often the most critical part of the analysis. The aim of this paper is to determine whether the conclusions arrived at in *Gupta et al. [2015]* are justified by the approach. First, a general outline of statistical inference is presented to provide context for the subject-matter translation problem. A discussion of the model employed by *Gupta et al. [2015]* then follows to determine its capacity for inference. Finally the conclusions of *Gupta et al. [2015]* are compared to the model’s capacity for inference.

2.0 Background

2.1 Statistical Inference: Frequentist and Bayesian

Some of the difficulty in reviewing the literature is due to the prevalence of value statements invoking blame, guilt, and fear, none of which contribute to the understanding of science. (Lloyd and Oreskes 2018) Certainly no discipline or body of literature is perfect, however some of the lack of understanding of statistics often decried in the literature may instead be an indication of the

similarities between the established paradigms of statistical inference. The prominent statistician D.R. Cox broadly summarized inferential statistics by the following paradigms, by the briefest of definitions (Cox 2006):

- **Frequentist:** inference of system behaviour is measured from data alone, assuming fixed parameters, and
- **Bayesian:** inference of system behaviour is measured from data, but prior knowledge is incorporated by assuming probabilistic parameters.

Lindley [2000] states the concern of statistical analysis is evaluating uncertainty, and the fundamental problem of statistical inference is in using past data to predict future data. Significance level and confidence, which are descriptions of parameters and not data, do not obey the probability calculus. As such, statistical inference can be measured only by probability, but both Frequentist and Bayesian paradigms measure inference probabilistically. (Lindley 2000)

While the treatment of the statistical discipline in *Lindley [2000]* entirely avoids the language of causality and attribution, causal inference is a more recently established paradigm (despite independent origins in the 1920s from both Barbara Burks and Sewall Wright) putting causality central in the approach to statistical inference. (Pearl 2009) The capacity to evaluate nonexistent (what-if) scenarios, or counterfactuals, is the more advanced level of inference that the field of artificial intelligence strives for, and mere association (i.e. linear regression, machine learning) is the most primitive level. (Pearl and Mackenzie 2018) Statistical inference can thus reasonably include both associative and causal sub-categories. *Information* and *Likelihood* are additional established paradigms of statistical inference that are beyond the scope of this discussion.

The variety of ways of expressing like methods is a natural outcome of the application of statistics across the breadth of academic disciplines with little reason or opportunity to share ideas. Proof of the apparent interchangeability of methods can be seen in a random sample of titles by giving an academic journal database the key words “Frequentist” and “Bayesian”. Similarly, part of the challenge in reviewing the hydrological literature lies in the nuanced description and integrated application of statistical methods. While there is a current trend favouring process-based hydrological analysis over purely empirical approaches, there remains a lack of consensus in the definition and measurement of hydrological connectivity. (Bracken et al. 2013) In the field of hydrological research, there are numerous and varied approaches to the measurement and prediction of runoff, as well as to the attribution of physical causes to trends in observed data. (A. Viglione et al. 2016) *Bracken [2013]* invokes the Bayesian paradigm, with “*hydrological connectivity*” suggesting the incorporation of prior knowledge into a model, and *Viglione [2016]* invokes causality with “*the attribution of physical causes*”. *Rogger et al. [2017]* also invokes the language of causality in their criticism of the discipline:

“Studies that examine the impact of land use changes on streamflow and floods often obtain contradictory results for the same kind of

change.” (Rogger et al. 2017)

Analysis of hydrometric data is undertaken in order to base decisions upon expectations of future behaviour of some parameter. As such, hydrological analysis is inherently inferential, rather than descriptive. To gain any level of practical understanding of runoff at the watershed level, a model of some form must be employed. Input variables to hydrological models are discrete observations in time and space, representing samples of components and mechanisms of the hydrologic cycle. One of the central tasks in the study of watershed hydrology is the determination of an appropriate model for the characterization of timing and quantity of runoff at a spatio-temporal scale of interest. It is the model selection that determines the paradigm of inference of the study. The established inferential paradigms are not mutually exclusive, rather there are a variety of valid approaches to characterizations of the hydrologic cycle, and the validity of approach is dependent upon on the question being asked of the data.

2.2 Model Selection: Deterministic, Stochastic, and In-between

Hydrological processes occur on many different scales, and are both deterministic and stochastic in nature. *Sivakumar [2017]* uses the two terms as extremes in describing the complexity of systems, with a third term to describe the space between the two:

- **stochastic:** nonlinear interactions dominate the hydrologic cycle yielding random and irreproducible states of the real system,
- **deterministic:** order and dependence exist at some spatiotemporal scales, such as annual river flow and daily temperature, and
- **chaotic:** systems governed by three or more degrees of freedom (independent variables required to describe the state of a system) can be deterministic in the short term, but irreproducible unpredictable in the long term due to sensitivity to initial conditions.

Large scale atmospheric circulation transports water (mass) and energy (heat) in dynamic systems increasing in complexity with spatial scale. The degrees of freedom (independent variables) of large scale atmospheric circulation models employ in the order of 20-40 independent variables to describe the state of a system at a point in time. (Fraedrich, Ziehmman, and Sielmann 1995) The nonlinear interactions of the components and mechanisms in the hydrologic cycle represent the stochastic, or random, nature of the hydrologic system, requiring stochastic models. But some level of determinism does exist in the hydrologic cycle. In the short term, there is determinism and order in a low complexity system such as synoptic rainfall and resultant runoff in a small, highly developed catchment. A watershed made slightly more complex by increased size, or addition of any number of forms of storage may gain sufficient degrees of freedom, and as a result be unpredictable and irreproducible in the long-term due to sensitivity to initial conditions.

Something I’m not sure fits here

A final point about the criticism levelled at the hydrology discipline from *Rogger et al. [2017]*:

Although results from individual studies are legitimate, it is difficult to obtain general statements on the impacts since each study takes a rather narrow and study-specific perspective. (Rogger et al. 2017)

From an epistemological standpoint, if studies investigating attribution of the same intervention arrive at contradictory results, one or both must be false. Alternatively, if the methodologies from individual studies are in fact both valid, they must be asking different questions.

3.0 The Statistical Model of *Gupta et al. [2015]

“Rather than idealized angels of reason, scientific models are powerful clay robots without intent of their own, bumbling along according to the myopic instructions they embody.” (McElreath 2018)

In terms of the hypotheses (1), while Gupta et al. [2015] focus on the driving factors of a detected increasing trend in *mean annual* runoff, Alila et al. [2009] focus on the effect of forest harvesting on *peak annual* runoff, in terms of frequency and magnitude. The two hypotheses are distinct, but may not be entirely unrelated, as will be discussed. The framework to evaluate methodologies (2) is based on Mertz et al. [2012], who propose as ingredients of attribution: evidence of consistency, evidence of inconsistency, and provision of confidence interval.

different statistical approaches one truth, path is many -frequentist inference

path of action. (Neymann-Pearson) uses rules to govern behaviour such that in the long run, we will not be wrong too often. p values and alpha levels to reject or accept the null hypothesis. this is a rule to govern our behaviour in the long run. tells us nothing about the current test.

path of knowledge: focuses on likelihoods. based on observed data, what is likeliest of the true value? focuses on relative evidence between hypotheses. Bayes without priors.

path of belief (Bayes): incorporate a priori information. express evidence in terms of ‘degrees of belief’.

2.0 Study Summaries

In the subsequent sections, a condensed synopsis is presented, along with a brief summary of the commentary elicited by each study. Reproduction of results is not the purpose of this paper as the hypotheses of the studies differ sufficiently,

making direct comparison difficult. Instead, a discussion of the epistemological nature of the two studies is developed in Section 3 to arrive at a common territory to enable meaningful comparison.

2.1 Synopsis of Gupta et al. [2015]

Analysis of measured runoff in 29 watersheds in Iowa and Minnesota demonstrates an increasing trend of annual runoff, coincident with a positive trend in annual precipitation. Gupta et al. [2015] attempt to quantify the relative influence of precipitation and LULC to the observed increase in runoff. A secondary goal of the study is looking into the confounding effects of changing crop systems and loss of wetlands to explain the observation of approximately constant evapotranspiration (ET) over time.

Gupta et al. [2015] separates the measured record into two periods consistent with a Before-After-Control-Impact (BACI) analysis framework, citing extensive adoption of plastic drain tile in agricultural practices in the mid-1970s as the LULC intervention, as adopted in numerous previous studies (refer to Gupta et al. [2015]). Using annual precipitation and runoff volumes, Gupta et al. [2015] posits that a change in the linear relationship between precipitation and (the natural logarithm of) runoff should be indicative of a change in how the watershed converts precipitation to streamflow. (Foufoula-Georgiou et al. 2016)

Gupta et al. [2015] tests for shifts in streamflow versus precipitation relationships using a series of models of varying complexity, presented in more detail in the subsequent section. The study found the model coefficients to be statistically significant at the 5% level for 19 out of 29 watersheds using a multivariate linear regression model, and additionally all 29 watersheds exhibited a shift in 5-year moving averages of precipitation vs. runoff. Given these observed trends, the authors conclude that precipitation was the sole driver of increased streamflow, and LULC did not make a significant contribution.

In terms of the secondary question of ET, Gupta et al. [2015] conclude that despite substantially different potential ET in the predominant crop systems over the two time periods, no statistically significant trend in ET over time was found.

2.2 Gupta et al. Model Methodology: ANOVA

Runoff and precipitation data from 1966 to 2009 were used in the BACI study, with the breakpoint between the two periods, corresponding to widespread adoption of plastic tile drainage in agriculture, set at 1975. The system of models are described by the following equations (1), (2), and (3):

$$\ln(Q_{all}) = \beta_0 + \beta_1 \cdot P_{all} + \beta_2 \cdot I + \beta_3 \cdot P \cdot I$$

(1)

$$\ln(Q_{all}) = \beta_4 + \beta_5 \cdot P_{all} + \beta_6 \cdot I$$

(2)

$$\ln(Q_{all}) = \beta_7 + \beta_8 \cdot P_{all}$$

(3)

Statistical tests of the coefficients (β_0, \dots, β_8) are used to measure whether the relationship between streamflow and precipitation for the two periods of interest are better represented by

In each model, I has a value of 0 or 1 based upon the period, such that pre and post-change periods are assigned to separate coefficients. ANOVA tests for significant difference in the

Temporal trends in annual precipitation are demonstrated in two ways, one using the Mann-Kendall nonparametric test, and the other by calculating mean annual precipitation for three periods: 1920-1949, 1950-1979, and 1980-2009.

-where does the moisture come from?

-what about uncertainty in ET estimation? Methodology:

$$ET = PPT - Q - \Delta S - D$$

Pan evaporation is a measure of evaporative demand, and is driven by humidity gradients, temperature, wind speed, and solar insolation. (Roderick et al. 2007) Investigating a widely observed global trend in decreasing pan evaporation, *Roderick et al.* [2007] modeled the components of evaporative demand and attributed the decline in measured pan evaporation between 1975 and 2004 to a reduction in wind speed along with regional reduction in insolation. Note that pan evaporation data were used based on a single location to represent evapotranspiration across all of Minnesota and Iowa. Average wind speeds are spatially variable across Minnesota and Iowa (Harding and Snyder 2012)

-Gupta says seasonal runoff ratio changes are not appropriate due to dependence of runoff on antecedent soil moisture in previous season.

The study cites evidence of the spread of agricultural practices, including the use of drainage ditches and subsurface drain tile, beginning in the early 1900s. This assumption thus neglects the existing drainage and subsurface drain tile, in use for three quarters of a century prior to the set breakpoint in study periods (1975). Numerous related studies viewing widespread adoption of plastic drain tile in the mid 1970s as the major cause of increased runoff ((K. E. Schilling and Libra 2003), (Broussard et al. 2008), (Wang and Hejazi 2011), (Xu et al. 2013), (Schottler et al. 2014)). But without evidence of performance and/or soil moisture measures to defend the null hypothesis (no effect of drainage tile), the

intervention being investigated is then limited to the performance of modern plastic drain tile versus the older clay tile.

Alila: -dominant process theory (moderate correlation between April 1st SWE and peak flows) -frequency pairing -chronological pairing -what does Alila say are the causes of changes in variability? What are the ways he suggests this is demonstrated in the data? -what is the logical fallacy of composition? The inference that something is true of the whole from the fact that it is true of some part of the whole

-in pursuing the argument of the effect of forest storage on the frequency of floods, there is an implicit argument that forest harvesting, which tends to increase variability of runoff, changes the FFC. If the effect of forest harvesting translates the FFC in the positive vertical direction, the mean is necessarily affected. If the effect of forest harvesting has no effect on the lowest probability events, but has an effect on the higher probability events, the mean is necessarily affected. The logical

Gupta: -chronological pairing -ANOVA/ANCOVA

-possibility of delayed or transient effects of intervention (Murtaugh 2002)

-autocorrelations are not characteristic enough to distinguish random from deterministic chaotic signals (Sivakumalan 2017)

-Interaction between variables occurs in many different ways (often in feedback forms) and in varying degrees of nonlinearity, and so the variables are interdependent on each other, i.e. every variable is dependent on every other variable in a direct or indirect way. The hydrologic cycle is an excellent example of nonlinear interactions and interdependencies among variables, since every component in the hydrologic cycle is connected with every other component, either directly or indirectly (Sivakumar 2017;2016;2011;)

1. What are the interventions being investigated?
2. What are the specific hypotheses?
3. Are the hypotheses consistent with each other, or diametrically opposed?

Summary of *Gupta et al. [2015]*

Objective

Methodology

Supporting Evidence

Commentary

Summary of *Alila et al. [2009]*

Objective

Alila argues pairing of floods in ANCOVA and ANOVA fail to account for changes in flood frequency due to deforestation.

-investigate point about original intent of ANOVA/ANCOVA for detecting changes in means. Is it appropriate to use for extremes? ANOVA/ANCOVA don't account for changes in variance. How to show this?

Methodology

Supporting Evidence

Commentary

References

- Bracken, L. J., J. Wainwright, G. A. Ali, D. Tetzlaff, M. W. Smith, S. M. Reaney, and A. G. Roy. 2013. "Concepts of Hydrological Connectivity: Research Approaches, Pathways and Future Agendas." *Earth-Science Reviews* 119: 17–34.
- Broussard, Whitney, Peter A. Raymond, Neung-Hwan Oh, and R. E. Turner. 2008. "Anthropogenically Enhanced Fluxes of Water and Carbon from the Mississippi River." *Nature* 451 (7177): 449–52.
- Cox, D. R. 2006. *Principles of Statistical Inference*. Cambridge, UK;New York; Cambridge University Press.
- Foufoula-Georgiou, Efi, Patrick Belmont, Peter Wilcock, Karen Gran, Jacques C. Finlay, Praveen Kumar, Jonathan A. Czuba, Jon Schwenk, and Zeinab Takbiri. 2016. "Comment on 'Climate and Agricultural Land Use Change Impacts on Streamflow in the Upper Midwestern United States' by Satish c. Gupta et Al." *Water Resources Research* 52 (9): 7536.
- Fraedrich, Klaus, Christine Ziehmann, and Frank Sielmann. 1995. "Estimates of Spatial Degrees of Freedom." *Journal of Climate* 8 (2): 361–69.
- Harding, Keith J., and Peter K. Snyder. 2012. "Modeling the Atmospheric Response to Irrigation in the Great Plains: Part Ii: The Precipitation of Irrigated Water and Changes in Precipitation Recycling." *Journal of Hydrometeorology* 13 (6): 1687–1703.
- Lindley, Dennis V. 2000. "The Philosophy of Statistics." *Journal of the Royal*

Statistical Society. Series D (the Statistician) 49 (3): 293–337.

Lloyd, Elisabeth A., and Naomi Oreskes. 2018. “Climate Change Attribution: When Is It Appropriate to Accept New Methods?” *Earth’s Future* 6 (3): 311–25.

McElreath, Richard. 2018. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. 1st ed. Vol. 122. Boca Raton: CRC Press/Taylor & Francis Group.

Pearl, Judea. 2009. *Causality*. New York: Cambridge University Press.

Pearl, Judea, and Dana Mackenzie. 2018. *The Book of Why: The New Science of Cause and Effect*. First. New York, NY: Basic Books, Hachette Book Group.

Roderick, Michael L., Leon D. Rotstajn, Graham D. Farquhar, and Michael T. Hobbins. 2007. “On the Attribution of Changing Pan Evaporation.” *Geophysical Research Letters* 34 (17): L17403.

Rogger, M., M. Agnoletti, A. Alaoui, J. C. Bathurst, G. Bodner, M. Borga, V. Chaplot, et al. 2017. “Land Use Change Impacts on Floods at the Catchment Scale: Challenges and Opportunities for Future Research: LAND Use Change Impacts on Floods.” *Water Resources Research* 53 (7): 5209–19.

Schilling, Keith E., and Robert D. Libra. 2003. “INCREASED Baseflow in Iowa over the Second Half of the 20th Century.” *JAWRA Journal of the American Water Resources Association* 39 (4): 851–60.

Schottler, Shawn P., Jason Ulrich, Patrick Belmont, Richard Moore, J. W. Lauer, Daniel R. Engstrom, and James E. Almendinger. 2014. “Twentieth Century Agricultural Drainage Creates More Erosive Rivers.” *Hydrological Processes* 28 (4): 1951–61.

Sivakumar, Bellie. 2017;2016;2011; *Chaos in Hydrology. Bridging Determinism and Stochasticity*. Dordrecht: Springer Netherlands.

Viglione, Alberto, Bruno Merz, Nguyen Viet Dung, Juraj Parajka, Thomas Nester, and Günter Blöschl. 2016. “Attribution of Regional Flood Changes Based on Scaling Fingerprints.” *Water Resources Research* 52 (7): 5322–40.

Wang, Dingbao, and Mohamad Hejazi. 2011. “Quantifying the Relative Contribution of the Climate and Direct Human Impacts on Mean Annual Streamflow in the Contiguous United States.” *Water Resources Research* 47.

Xu, Xianli, Bridget R. Scanlon, Keith Schilling, and Alex Sun. 2013. “Relative Importance of Climate and Land Surface Changes on Hydrologic Changes in the US Midwest Since the 1930s: Implications for Biofuel Production.” *Journal of Hydrology* 497: 110–20.