Using the National Water Model as a Hypothesis-Testing Tool

Richard P. Hooper[1], Grey S. Nearing[2], Laura S. Condon[3]

[1]*Dept. of Civil and Environmental Engineering, Tufts University, Medford, MA 02155*
[2]*Dept. of Geological Sciences, University of Alabama, Tuscaloosa, AL 35487*
[3]*Dept. of Civil and Environmental Engineering, Syracuse University, Syracuse, NY 13244*

ABSTRACT

High performance computing has enabled the creation of high-resolution, continental-scale models such as the NOAA's National Water Model (NWM) which predicts hourly streamflow at 2.7 million locations in the conterminous US using the National Hydrographic Dataset Plus (NHDPlus) as the reference geofabric. This high resolution model provides a novel opportunity to bridge the gap between the scale of process research (at hillslope to headwater catchment scale) and the operational scale of river basins where predictions are needed for water resources management and hazard. We present a study design that uses the NWM in a hypothesis-testing (i.e., rejectionist) framework to assess process representations included in both "physically based" models, such as ParFlow-CLM, and conceptual models, such as the USGS Precipitation Runoff Modeling System. An information theoretic framework is proposed to assess the ability of either of these approaches to extract information from available data to make reliable predictions of water transport in the subsurface and surface.

Keywords:
model performance, entropy measures

1.0 Introduction

Vast increases in computing power has made high-resolution continental-scale modeling possible. In the words of Bierkins et al. (2015), we are on the cusp of creating "locally relevant hydrological models everywhere." One such model is NOAA's National Water Model (NWM) which uses the NHDPlus hydrography of the conterminous US, to define 2.7 million river reaches with an average reach catchment area of 3 km^2 (NOAA, 2016). This geofabric—the NHDPlus reach catchment framework—not only has the resolution of typical experimental watersheds, also it is using a landscape discretization of catchments rather than a computational grid. This choice of catchment rather than grid as a landscape unit (although grids underlie the computation in NWM) make an important conceptual bridge to field scientists typically working at the scale of a few square kilometers.

Long-term experimental watersheds, such as the NSF Critical Zone Observatories (CZOs), ARS watersheds, or USFS experimental watersheds, offer a unique potential to advance understanding of hydrologic processes through the accumulation of data characterizing the landscape thoroughly enough to permit more sophisticated hypotheses to be tested. Data from long-term observatories have challenged conventional hydrologic wisdom on various fronts including the finding of the dominance of old water in storm hydrographs (e.g., Hooper and Shoemaker, 1986 at Hubbard Brook; Genereux and Hooper, 1998 in a review paper of multiple sites), the general disconnection between riparian and hillslope groundwater below a threshold of rainfall amount (e.g., Tromp-van Meervald and McDonnell, 2006 at Panola Mountain), and the two-water worlds observation that trees transpire water that is isotopically distinct from the water that runs off in streams (Brooks et al., 2010 at HJ Andrews and Evaristo et al., 2015 at multiple sites). All these findings challenge assumptions of homogeneity that underlie many hydrologic models but have not yet been incorporated in them.

Paradoxically, as computing power and sensing capabilities have increased, the division between field hydrologists and modelers has only deepened, as modelers take on heroic computing challenges at larger spatial scales and field hydrologists deploy sensors to study hillslopes and small catchments in ever greater detail. A fundamental need in hydrology is to develop techniques and computational infrastructure to improve dialog and feedback between field scientists and modelers in ways that can (1) directly initiate new cycles of research, and (2) help to directly improve forecasting models used for policy and planning.

Modeling groundwater transport, including flowpaths and residence times needed for linking chemical models with hydrology models, remains a major challenge - largely because of difficulties related to observing structure and characteristics of porous media across a range of scales. Fully physical models like ParFlow-CLM (Kollet and Maxwell, 2008) that numerically solve flow PDEs require extensive input parameters that must be interpolated and/or extrapolated from sparse hydraulic measurements of porous media. Alternatively, conceptual models apply simpler sets of semi-empirical equations intended to capture the most critical aspects of flow behavior, but often have empirical parameters that must be calibrated and may not have direct physical meaning. The latter – conceptual models – can more directly capture scientists' conceptual understanding of the system (hence the name), and are likely better for testing hypotheses that arise from field observatories like the CZOs, however they are less directly tied to the underlying physics of flow.

Large-scale models like the NWM require groundwater representations that will run reliably everywhere at a continental scale. We expect that, at arbitrary locations, full physically-based groundwater models like ParFlow will generally have larger parameter uncertainties, while conceptual groundwater models will generally have larger model structural uncertainties. We suggest novel benchmarking, process-diagnostics, and uncertainty quantification techniques to quantify these uncertainty-related tradeoffs

2.0 Background

2.1 The National Water Model

The work is motivated by the opportunities and challenges presented by the National Water Model (NWM) being developed by the Office of Hydrologic Prediction of the National Weather Service (NOAA, 2016). The NWM is based on WRF-Hydro (Gochis et al., 2013). Version 1 of WRF-Hydro uses a Dupuit-Forcheimer formulation for subsurface routing in the upper 2 m of soil and a linear reservoir for producing baseflow (Gochis et al., 2013). Citable documentation for the NWM is not currently available and some modifications have been made to the code, but it remains a simplified groundwater model that (arguably) serves the current purpose of flood prediction but limits the use of the NWM for more comprehensive "water intelligence" that NOAA seeks to provide where a more accurate representation of groundwater flowpaths and residence time are needed.

2.2 ParFlow-CLM

ParFlow-CLM is an integrated hydrologic model that represents water and energy fluxes from the bedrock through the top of the canopy using physically-based equations (Kollet and Maxwell, 2006; 2008; Maxwell and Miller, 2005). ParFlow simulates three-dimensional variably saturated subsurface flow using Richard's equation, and has fully integrated overland flow that is simulated with the shallow water equations. It is coupled to the Community Land Model and the combined ParFlow-CLM model solves the full water energy balance at the land surface. The integrated approach used by ParFlow allows for dynamic interactions between groundwater levels, soil moisture and land energy fluxes that can evolve throughout simulations. These interactions are not possible in surface water models that ignore or rely on a parameterized approach to groundwater-surface water exchanges. ParFlow has been used in a number of studies ranging from the watershed to the continental scale to demonstrate the importance of groundwater surface water interactions and the potential for feedbacks between groundwater depth and water availability at the land surface (e.g., Condon et al., 2015; Ferguson and Maxwell, 2010;

Maxwell and Kollet, 2008b; Maxwell and Condon, 2016; Maxwell et al., 2007; Mikkelson et al., 2013). ParFlow has also been used to estimate the residence time of groundwater at continental scale (Maxwell et al., 2016).

The physics-based approach used by ParFlow, and other distributed integrated models of its kind, is very computationally and data intensive. Distributed physical models minimize the number of empirical parameters which must be calibrated within the model in order to create a system which can more realistically respond to nonstationary conditions where parameters calibrated for one period may no longer hold. However, the tradeoffs are (1) a large number of required hydrologic properties (e.g. conductivity, porosity for every grid cell), and (2) increased computational expense, which, among other things, makes uncertainty assessment difficult. While ParFlow is an established tool that has been well validated in many previous studies, there are still open questions regarding the scaling behavior of physically based models. For example, are larger scale emergent behaviors observed at field sites, such as threshold response for connecting hillslopes with riparian zones (Tromp-van Meervald, 2006) observed in model predictions? Applying ParFlow to these intensively studied sites can provide additional insights into model performance.

2.3 Conceptual Models

Unlike physically based models where the structure and governing equations are fixed, conceptual models seek to define these model elements based upon patterns observed in field data. Generally, the simplest possible model is proposed that is elaborated only when data indicates a lack of fit. Kirchner (2009) provides an excellent example of developing a conceptual model. He poses a simple hypothesis—discharge is a monotonic function of water storage in the catchment—and then determines the model parameters through a clever analysis of field data. Thus, although these are empirical parameters they are determined from field measurements—not simply tuned—that are determend at the same scale as the model. Nonetheless, a conceptual model may require unmeasurable parameters. It is well recognized that the information content in input-output signals can determine only a small number of tunable parameters (e.g., Hooper et al., 1986; Jakeman and Hornberger, 1993) so the number of such parameters must be minimized. The challenge is using all the data available—including data collected inside the catchment—to inform the model structure and parameters. Hooper and Christophersen (1992) provide an example of using conservative tracers to determine routing parameters for a conceptual model that could not be determined from the rainfall and runoff signals.

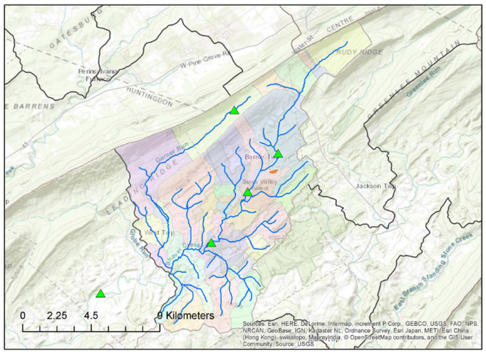
Contrast this approach to parameterization with physically based models where direct measurement of parameters is difficult due to the scale mismatch between the model grid size and the spatial support of the measurement, the number of parameters to be determined, and the underlying assumption that the hydraulic properties within a grid cell are homogeneous. Given the limitations of both physically based models and conceptual models, we believe it is worthwhile to conduct a systematic and rigorous comparison of the two modeling approaches.

2.4 Field Sites

Critical Zone Observatories provide a range of intensively sampled sites for executing this project design. Three are of particular interest because of previous work that has been done: Clear Creek, IA ((CC) and the Upper Sangamon River, IL (USR) in the Intensively Managed Landscapes CZO and Shavers Creek, PA in the Shale Hills/Susquehanna CZO. These sites are described extensively at their respective web sites (http://criticalzone.org/iml; http://criticalzone.org/shale-hills).

These basins range in area—SC, 90 km^2; CC, 270 km^2; USR, 3,690 km^2—and each provide different modeling challenges. SC is a largely undisturbed forested catchment while both CC and USR have extensive agriculture and tile drains. These sites are well instrumented and have had various geophysical surveys to develop hydrostratigraphic models. Other groundwater and integrated groundwater/surface water models have been developed at them, including a MODFLOW model of the USR (Herzog et al., 2003), a MIKE-SHE model of the USR (Stumpf, Illinois Geological Survey, pers. comm.) and RT-Flux-PIHM at SC (Bao et al., 2017). This provides a rich background of contrasting predictions in which to place the current work. Scientists at these sites will be engaged both at the development stage of the conceptual model and in evaluating the model predictions.

Of particular interest to this project is the relation between the NHDplus geofabric and these field sites. Figure 1 shows two of the sites (CC and SC) with the NHDplus stream network and the associated catchments (shaded areas) along with the current stream



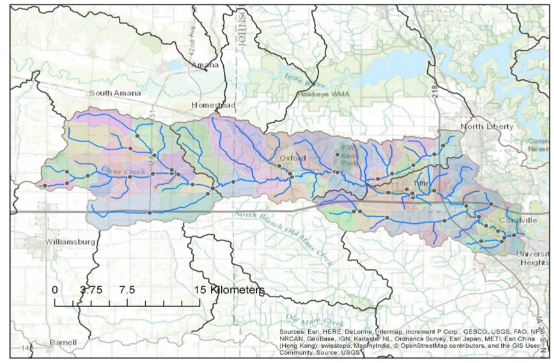


Figure 1. Maps showing the NHDplus geofabric for two of the study sites, CC (left) and SC (right).

gaging network operated by the CZOs and their partners. CC has a particularly extensive gaging net work throughout the basin. The NWM makes predictions of streamflow for each of these shaded basins.

3.0 Techniques for Model Diagnostics

Even when predictions conform to observations, we must have a way to determine whether a particular model configuration "gets the right answers for the right reasons" (Kirchner, 2006). While parameter identifiability is a problem for physically-based models like ParFlow, the primary danger with using a conceptual model is kludging (Clark, 1987). Kludging is when a model is tuned to match some set of available observations, without accounting for the possibility of compensating error structures in the internal dynamics of the model. A key aspect of model evaluation is an assessment of the structural adequacy of the model. We must measure two things: first, how much information and how much disinformation is provided by a particular modeling hypotheses, and second, how a given process-specific hypotheses changes information flow within the model.

Our recent work has developed a unified theoretical framework for measuring information and disinformation from individual process-specific hypotheses, and to do this in the presence of large and arbitrary input data uncertainties. Nearing & Gupta (2015) outlined the original motivation and examples, Nearing et al. (2016c) developed a philosophical basis for the approach, Nearing and Gupta (2017) turned that philosophical basis into a method of hypothesis testing, Nearing et al. (2016a) used the approach to separate uncertainty contributions due to missing information in (i) model structure, (ii) model parameters, and (iii) forcing data, and Nearing et al. (2016b) adapted the methodology to assess process-level deficiencies in model structures.

The challenge is to separate information gained by improving model structure from information gained or lost due to our (in)ability to effectively parameterize the models given available observation data. Markov Chain Monte Carlo (MCMC) is an apporach that can be used to integrate out parameter uncertainty in order to isolate information added or lost by the model structure alone. Dynamic process networks can be used to quantify information flow internal to the model. ParFlow can serve as an upper benchmark to measure information loss due to simulating groundwater flow with conceptual hypotheses, and a set of data-denial and data-corruption experiments can identify at what input data quality a particular conceptual model provide more and/or higher quality information than does a poorly-parameterized ParFlow.

The basic tool for evaluating internal process-diagnostics is the dynamic process networks (DPN) proposed by Ruddell & Kumar (2009). A DPN represents a dynamic system or dynamical systems model as a Bayesian network, whereby each node in the network represents a simulated variable (typically either a state or flux) and each edge in the network represents influence that one variable exerts on another at a certain spatiotemporal scale.

Influence between variables is quantified by measuring information transfers. To calculate the influence that one variable, , has on another variable, , at a particular timescale, , we integrate over the expected effect of probabilistically conditioning at time on the value of at time given all of the variables in the model other than at time . Note that even in deterministic models, each variable is probabilistic conditional on only a subset of other variables, and all variables are probabilistic in presence of input and parameter uncertainties (as identified by MCMC). Schreiber (2000) proposed a feasible Markov approximation of this integration that results in the following metric, called *transfer entropy* that quantifies the information transferred from variable to at timescale , and the variables in question (and) may represent state or flux variables integrated over any spatiotemporal extent. Thus, this metric can quantify couplings between observed or modeled variables at any spatiotemporal scale.

Two separate sets of DPNs (at specific spatiotemporal scales) are constructed – one from model-simulated data, and one from CZO-observed data. Each directed edge in each DPN is quantified at a particular scale using a metric like transfer entropy and differences between the edge-metrics in the modeled vs. observed DPNs can be evaluated. A positive difference along a particular network edge indicates that this pair of variables is coupled too strongly in the model, whereas a negative difference indicates that the model underestimates the role of one variable on determining the other.

4.0 Conclusions

A study design is described that uses the National Water Model structure to scale findings from experimental catchments to river basins. This approach will contrast the ability of physically based models with conceptual models to extract information from data available for model calibration with the assumption that more reliable predictions will result from models that more effectively extract information.

Acknowledgements

Funding for this work is provided as a subaward from CUAHSI from NSF grant EAR-1338606. Useful discussions about this study were held with Alva Couch (Tufts University), Tony Castronova (CUAHSI), Praveen Kumar (University of Illinois), and Dave Tarboton (Utah State University).

References

Bao, C., L. Li, Y. Shi, and C. Duffy (2017). Understanding watershed hydrogeochemistry: 1. Development of RT-Flux-PIHM, Water Resources Research, 53(3), 2328-2345.

Bierkens, M.F., V.A. Bell, P. Burek, N. Chaney, L.E. Condon, C.H. David, A. de Roo, P. Döll, N. Drost, J.S. Famiglietti, and M. Flörke, (2015), Hyper-resolution global hydrological modelling: What is next? Everywhere and locally relevant, Hydrol Process, 29(2), 310-320.

Brooks, J. R., H.R. Barnard, R. Coulombe, and J.J. McDonnell (2010), Ecohydrologic separation of water between trees and streams in a Mediterranean climate. Nature Geoscience, 3(2), 100.

Clark, A. (1987) The kludge in the machine. Mind & Language, 2(4), pp. 277-300.

Condon, L. E., A.S. Hering, and R.M. Maxwell (2015), Quantitative assessment of groundwater controls across major US river basins using a multi-model regression algorithm, Advances in Water Resources, 82, 106-123.

Evaristo, J., S. Jasechko, and J.J. McDonnell (2015), Global separation of plant transpiration from groundwater and streamflow. Nature, 525(7567), 91.

Ferguson, I. M. and R. M. Maxwell (2010), Role of groundwater in watershed response and land surface feedbacks under climate change. Water Resources Research, 46(10), W00F02, doi:10.1029/2009WR008616.

Genereux, D. P., and R.P. Hooper (1998), Oxygen and hydrogen isotopes in rainfall-runoff studies, 319-346, *in* Isotope tracers in catchment hydrology, C. Kendall and J. McDonnell, Elsevier, Amsterdam.

Gochis, D. J., W. Yu, and D.N. Yates (2013), The WRF-Hydro model technical description and user's guide, version 1.0. NCAR technical document, 120.

Herzog, B. L., D.R. Larson, C.C. Abert, S.D. Wilson, and G.S. Roadcap (2003), Hydrostratigraphic Modeling of a Complex, Glacial-Drift Aquifer System for Importation into MODFLOW, Groundwater, 41(1), 57-65.

Hooper, R. P. and C.A. Shoemaker (1986), A comparison of chemical and isotopic hydrograph separation, Water Resources Research, 22(10), 1444-1454.

Hooper, R. P. and N. Christophersen (1992), Predicting episodic stream acidification in the southeastern United States: Combining a long-term acidification model and the end-member mixing concept, Water Resources Research, 28(7), 1983-1990.

Hooper, R. P., A. Stone, N. Christophersen, E de Grosbois, and H.M. Seip (1988), Assessing the Birkenes model of stream acidification using a multisignal calibration methodology, Water Resources Research, 24(8), 1308-1316.

Jakeman, A. J., and G.M. Hornberger (1993), How much complexity is warranted in a rainfall-runoff model?. Water Resources Research, 29(8), 2637-2649.

Kirchner, J. W. (2006), Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, Water Resour. Res., 42, W03S04, doi:10.1029/2005WR004362.

Kirchner, J. W. (2009), Catchments as simple dynamical systems: Catchment characterization, rainfall-runoff modeling, and doing hydrology backward, Water Resour. Res., 45, W02429, doi:10.1029/2008WR006912

Kollet, S. J. and R.M. Maxwell (2006), Integrated surface–groundwater flow modeling: A free-surface overland flow boundary condition in a parallel groundwater flow model. Advances in Water Resources, 29(7), 945-958.

Kollet, S. J., and R. M. Maxwell (2008), Capturing the influence of groundwater dynamics on land surface processes using an integrated, distributed watershed model, Water Resources Research, 44(W02402), doi: 10.1029/2007WR006004.

Maxwell, R. M., and N. L. Miller (2005), Development of a coupled land surface and groundwater model, Journal of Hydrometerology, 6(3), 233-247, doi: doi:10.1175/JHM422.1.

Maxwell, R. M. and S. J. Kollet, (2008b), Interdependence of groundwater dynamics and land-energy feedbacks under climate change, Nature Geoscience, 1(10), 665.

Maxwell, R. M. and L.E. Condon (2016). Connections between groundwater flow and transpiration partitioning. Science, 353(6297), 377-380.

Maxwell, R. M., F.K. Chow, and S.J. Kollet (2007), The ground-water–land-surface–atmosphere connection: Soil moisture effects on the atmospheric boundary layer in fully-coupled simulations, Advances in Water Resources, 30(12), 2447-2466.

Maxwell, R. M., L.E. Condon, S.J. Kollet, K. Maher, R. Haggerty and M.M. Forrester (2016), The imprint of climate and geology on the residence times of groundwater, Geophysical Research Letters, 43(2), 701-708.

Mikkelson, K. M., R. M. Maxwell, I. Ferguson, J. D. Stednick, J. E. McCray, and J. O. Sharp (2013), Mountain pine beetle infestation impacts: modeling water and energy budgets at the hill-slope scale, Ecohydrology, 6(1), 64-72, doi: 10.1002/eco.278.

Nearing, G. S., Gupta, H. V., Crow, W. T., & Gong, W. (2013a). An approach to quantifying the efficiency of a Bayesian filter. Water Resources Research, 49(4), 2164-2173.

Nearing, G. S., Gupta, H. V., & Crow, W. T. (2013b). Information loss in approximately Bayesian estimation techniques: A comparison of generative and discriminative approaches to estimating agricultural productivity. Journal of hydrology, 507, 163-173.

Nearing, G. S. and H.V. Gupta, (2015), The quantity and quality of information in hydrologic models, Water Resources Research, 51(1), pp. 524-538.

Nearing, G. S., Mocko, D. M., Peters-Lidard, C. D., Kumar, S. V. and Xia, Y. (2016a) Benchmarking NLDAS-2 soil moisture and evapotranspiration to separate uncertainty contributions, Journal of Hydrometeorology, pp. 745-759.

Nearing, G. S., B. Ruddell, M.P. Clark, and B. Nijssen, (2016b), The Role of Data Assimilation in Model Diagnostics (invited). American Geophysical Union Fall Meeting, San Francisco, CA USA.

Nearing, G. S., Y. Tian, H.V. Gupta, M.P. Clark, K.W. Harrison, and S.V. Weijs, (2016c), A philosophical basis for hydrologic uncertainty, Hydrological Sciences Journal, 16(9), pp. 1666-1678.

Nearing, G. S. and H.V. Gupta (2017), Information vs. uncertainty as a foundation for a science of environmental modeling, https://arxiv.org/abs/1704.07512.

NOAA (2016), National Water Model: Improving NOAA's Water Prediction Services, http://water.noaa.gov/documents/wrn-national-water-model.pdf. 2 pp.

Ruddell, B. L., and P. Kumar (2009), Ecohydrologic process networks: 1. Identification, Water Resour. Res., 45, W03419, doi:10.1029/2008WR007279.

Schreiber, T. (2000) 'Measuring information transfer', Physical review letters, 85(2), pp. 461.

Tromp-van Meerveld, H. J. and J.J. McDonnell (2006), Threshold relations in subsurface stormflow: 1. A 147-storm analysis of the Panola hillslope. Water Resources Research, 42(2), W02410, doi:10.1029/2004WR003778 .