

Bridging groundwater models and decision support with a Bayesian network

Michael N. Fienen,¹ John P. Masterson,² Nathaniel G. Plant,³ Benjamin T. Gutierrez,⁴ and E. Robert Thieler⁴

Received 21 December 2012; revised 31 July 2013; accepted 18 August 2013; published 9 October 2013.

[1] Resource managers need to make decisions to plan for future environmental conditions, particularly sea level rise, in the face of substantial uncertainty. Many interacting processes factor in to the decisions they face. Advances in process models and the quantification of uncertainty have made models a valuable tool for this purpose. Long-simulation runtimes and, often, numerical instability make linking process models impractical in many cases. A method for emulating the important connections between model input and forecasts, while propagating uncertainty, has the potential to provide a bridge between complicated numerical process models and the efficiency and stability needed for decision making. We explore this using a Bayesian network (BN) to emulate a groundwater flow model. We expand on previous approaches to validating a BN by calculating forecasting skill using cross validation of a groundwater model of Assateague Island in Virginia and Maryland, USA. This BN emulation was shown to capture the important groundwater-flow characteristics and uncertainty of the groundwater system because of its connection to island morphology and sea level. Forecast power metrics associated with the validation of multiple alternative BN designs guided the selection of an optimal level of BN complexity. Assateague island is an ideal test case for exploring a forecasting tool based on current conditions because the unique hydrogeomorphological variability of the island includes a range of settings indicative of past, current, and future conditions. The resulting BN is a valuable tool for exploring the response of groundwater conditions to sea level rise in decision support.

Citation: Fienen, M. N., J. P. Masterson, N. G. Plant, B. T. Gutierrez, and E. R. Thieler (2013), Bridging groundwater models and decision support with a Bayesian network, *Water Resour. Res.*, 49, 6459–6473, doi:10.1002/wrcr.20496.

1. Introduction

[2] Decision makers rely on forecasts of system response to future conditions to develop adaptive management strategies. Future conditions—particularly those involving climate change—must be understood in the context of their uncertainty [National Research Council, 2009]. Molina *et al.* [2010] suggest that for water resources management, two main approaches for forecasts with uncertainty are Decision Support Systems (DSS) and process models. Broadly, a DSS is a framework in which decision makers are able to evaluate the potential consequences of various choices they are faced with in managing resources [Poch *et al.*, 2004]. A DSS is most useful if it can explore a uni-

verse of choices involving interconnected processes, each with associated uncertainty, without relying on the laborious process of changing conditions in individual models and rerunning them or contacting a new set of experts for information about each choice. Process models, on the other hand, have the ability to incorporate exhaustive system detail but at great cost of construction, simulation runtime, and often, stability. This dichotomy between process models and DSSs is not a sharp one—indeed, process models are routinely used for decision making. However, there is a gap between a process model including the necessary detail to adequately represent complex processes and the need for expedience in the decision-making process. The goal of this work is to illustrate one way to emulate a process model, preserving much of the insight obtained from the detailed process model, in a computationally efficient way, while still propagating uncertainty.

[3] A fully coupled process model that represents all aspects of a natural system would be ideal, but is often difficult to develop. Progress has been made hydrologic modeling creating such models by integrating surface water with groundwater processes [Markstrom *et al.*, 2008; Therrien *et al.*, 2012] and solute transport with groundwater flow [Zheng, 1990; Langevin *et al.*, 2007]. However, even detailed process models that accurately simulate the underlying processes suffer from uncertainty and struggle to represent complex environmental conditions [Oreskes *et al.*,

¹U.S. Geological Survey, Wisconsin Water Science Center, Middleton, Wisconsin, USA.

²U.S. Geological Survey, New England Water Science Center, Northborough, Massachusetts, USA.

³U.S. Geological Survey, St. Petersburg Coastal and Marine Science Center, St. Petersburg, Florida, USA.

⁴U.S. Geological Survey, Woods Hole Science Center, Woods Hole, Massachusetts, USA.

Corresponding author: M. N. Fienen, U.S. Geological Survey, Wisconsin Water Science Center, 8505 Research Way, Middleton, WI 53562, USA. (mnfienen@usgs.gov)

1994]. Coupled process models also often require prohibitively long runtimes and often suffer from numerical instability that leads to nonconvergence problems [Doherty and Christensen, 2011]. A transparent and efficient assessment of salient model results with estimates of their uncertainty is critical for transparently supporting the needs of decision makers.

[4] In process models, the uncertainty in parameters can be propagated to forecasts through linear methods [e.g., Fienen et al., 2010], conditional realizations as in calibration-constrained Monte Carlo [e.g., Kitanidis, 1995; Tonkin and Doherty, 2009], or nonlinear methods such as Markov Chain Monte Carlo [e.g., Michalak and Kitanidis, 2003; Michalak, 2008; Steinschneider et al., 2012]. In process models, all changes in stresses or conditions must be explicitly represented in a model and the corresponding outcome must be evaluated. To quantify forecast uncertainty, propagation methods such as those discussed above must be employed. Even in the linear case, a separate evaluation of uncertainty for each candidate new stress must be performed. The computational expense incurred and numerical instability experienced by these individual evaluations often precludes practical inclusion of the analysis into integrated water resources management in which the natural process response is only one component of a broader decision-making framework.

[5] A DSS is intended to be fast and, while many processes may be included (including socioeconomic, ecological, and physical) [Sadoddin et al., 2005; Barthel et al., 2008; Guillaume et al., 2012], major simplifications are often made to mitigate the numerical instability and computational expense incurred when evaluating process model responses to stresses. Examples include a simplified mass-balance analysis [Martin de Santa Olalla et al., 2007] or elicitation of expert knowledge without direct modeling of the process [Stiber et al., 2004; Zickfeld et al., 2007]. In the mass balance approach, an analytical solution of mass balance is made to apply over an entire region lumping the processes that are explicitly simulated with a numerical model. Elicitation of expert knowledge provides ongoing challenges in quantifying levels of belief and posing questions of experts in a systematic and accurate way.

[6] There is a need for a bridge between simple models—or the reduction of the process to “expert opinion” only [e.g., Morgan et al., 1990]—and exhaustive use of process models in evaluating the uncertainty of system response to a predicted stress. Doherty and Christensen [2011] discuss pairing simple models with more complex models pulling useful attributes from each for making forecasts. Rather than relying on fluctuations about individual parameter values, however, perhaps the natural system has already provided the variability we need? The geometric arrangement of the system may sample enough variability in underlying parameter values (in addition to proximity to various boundary conditions and other variability that is harder to characterize using Monte Carlo methods) to characterize the uncertainty in forecasts through a systematic analysis of correlation. This is particularly so when geometry dominates as in the present case or when evaluating the interaction between pumping wells and surface water [e.g., Jenkins, 1968].

[7] The concept of using a surrogate, simpler model is similar to fitting a transfer function to a model rather than

to data alone and is a form of emulation [Kennedy and O'Hagan, 2001; Castelletti et al., 2012]. Other emulation approaches have been explored, seeking to learn associations between input parameters and associated output. For example, artificial neural networks [Nolan et al., 2012; Gusyev et al., 2012] can be trained on numerical model results, or high-dimensional model reduction [Li et al., 2001] can reduce model inputs and outputs to a simpler high-order polynomial representation.

[8] We explore the use of a Bayesian decision network (BN) [Jensen and Nielsen, 2001] to emulate a model of the Assateague island groundwater flow system, propagating uncertainty from inputs through to outputs. Long-term changes in the land surface and water-table position on barrier islands like Assateague can have a profound effect on species diversity and ecosystem function across the barrier-island landscape [Ehrenfeld, 1990; Najjar et al., 2000; Scavia et al., 2002]. Resource managers, tasked with making decisions to preserve natural habitat and recreational use need guidance to plan for changes in both island elevation and water-table position in the context of sea level rise, where both the drivers and responses are uncertain [National Research Council, 2009]. The forecast considered in this work is mean depth to water; however, only changes in water-table position are calculated by the groundwater model. We recognize that changes in land surface also affect depth to water, but the processes that affect that surface were not considered in this analysis.

[9] The approach of using a BN for model emulation has been applied to create DSSs from numerical models in coastal engineering applications [e.g., Plant and Holland, 2011a, 2011b]. In this manuscript, we present an example of emulating a numerical groundwater model using a BN. We show that a BN generally can capture the behavior of a numerical groundwater model using an emulation approach and can propagate uncertainty from inputs through forecasts. This method provides a means for forecasting without the long runtimes of the process model. If other process models can be emulated similarly, causal links among them can be made by combining multiple BNs into a single object-oriented BN [Molina et al., 2010; Carmona et al., 2011] or into a formal DSS.

[10] We also present metrics to evaluate the trade off between fit to calibration data and forecasting power, including sensitivity and cross validation. The resulting BN is suitable for inclusion into a DSS evaluating the mean depth to water response to changes in island shape due to sea level rise and can be readily linked to BNs created using similar techniques from other process models simulating related processes.

2. Assateague Island and Barrier Island Systems

[11] Assateague Island is a well-studied barrier island that has attracted the interests of coastal geologists and engineers for several decades [Dean and Perlin, 1977; Halsey, 1978; Leatherman, 1979; Krantz et al., 2009]. Assateague Island—like all barrier islands—is shaped by storms, washover events, inlet formation, and the recovery afforded by longshore transport [Hayes, 1979; Leatherman, 1979; Nummedal, 1983; Davis, 1994; Oertel and Kraft, 1994].

[12] The majority of Assateague island is under the jurisdiction of the National Park Service (NPS) or US Fish and Wildlife Service (FWS) so there are ongoing management efforts to maintain the island in its natural state to maintain suitable habitat for a range of plants and animals. Of particular importance is maintaining protected and threatened shorebird populations that use the island for breeding and stop-overs during migration [e.g., *Schupp et al.*, 2013]. A wild-horse population—made famous in a popular children’s book [*Henry and Dennis*, 1947]—also depends on shallow freshwater ponds connected to the groundwater system.

[13] Key to these island ecosystems is the unique vegetation assemblage and distribution [*Ehrenfeld*, 1990]. Vadose zone thickness in shallow, coastal aquifer systems is a major determinant of the establishment, distribution, and succession of vegetation on these migrating, low-profile barrier islands [*Hayden et al.*, 1995; *Shao et al.*, 1995; *Rheinhardt and Faser*, 2001; *Kirwan et al.*, 2007; *O’Connell et al.*, 2012]. Therefore, long-term changes in the land surface or water-table position can have a profound effect on species diversity and ecosystem function across the barrier-island landscape [*Ehrenfeld*, 1990; *Najjar et al.*, 2000; *Scavia et al.*, 2002].

[14] Given the inherently dynamic nature of barrier islands, and their unique ecosystems that are highly dependent upon specific hydrologic conditions, these systems are particularly vulnerable to the effects of sea level rise [*Ehrenfeld*, 1990]. Sea level rise is a major climate change impact that will have a profound effect on the ecohydrology of coastal systems. Recent projections [*Rahmstorf*, 2007; *Jevrejeva et al.*, 2010] suggest that sea level may be 0.6–1.5 m higher than present by 2100.

[15] The hydrologic response of coastal aquifer systems to sea level rise is well documented [*Sherif and Singh*, 1999; *Barlow*, 2003; *Masterson and Garabedian*, 2007; *Werner and Simmons*, 2009; *Essink et al.*, 2010]. Increases in sea level position result in increases in groundwater levels in coastal aquifers. This response in barrier-island aquifer systems with relatively thin Vadose zones (typically <1 m) increases the potential for the water table to intersect land surface, increased groundwater discharge to surface water, and reduces the volume of the underlying freshwater lens.

[16] Developing forecasting tools that would allow for the physical effects of sea level rise to be factored into coastal planning and adaptive management efforts is a primary focus of ongoing research efforts at Assateague island. Uncertainties in future system drivers and responses, and interdependencies among the dynamic island morphology, groundwater, and implications on habitat motivate an integrated decision support framework. Understanding the groundwater system and emulating it with a BN is an important step along the path to integrated decision support.

[17] Assateague island has varied physical dimensions, topography, and ecology along its 60 km length. Specific and distinct conditions can be considered archetypes of other midlatitude barrier islands worldwide. Furthermore, as sea level rise progresses, conditions in one barrier island setting transition to conditions observed at other barrier island settings. On Assateague island, the northern and

southern ends of the island are representative of the stressed conditions that are likely to result from rising sea level, whereas the central portions of the island represent more stable conditions both on Assateague and other islands. This motivates the generalization of behavior on Assateague, distilled into salient characteristics described below, to potentially be used to forecast response to sea level rise in other similar settings on different islands. In this work, such forecasts can be made by using a BN trained to a groundwater flow model of Assateague island.

3. Groundwater Flow Model

[18] A groundwater flow model was developed for Assateague island, as shown in Figure 1, to evaluate the groundwater system response to sea level rise. The model was simulated by using SEAWAT [*Langevin et al.*, 2007] to simulate both groundwater flow and salt transport. *Masterson et al.* [2013] document the details of the modeling effort summarized here. The model was constructed by using geologic and spatial information to physically represent the island geometry and properties. A uniform lateral grid spacing of 50 m was applied throughout the model. Ten layers, ranging in thickness from 0.5 to 12 m, comprise the vertical discretization to a maximum depth of 30 m.

[19] The model was calibrated to observations of groundwater elevation and contrasts in salinity detected in boreholes using data collected by *Banks et al.* [2012]. Borehole data were limited to several transects across the northern half of the island. As a result, the calibration was performed by using regularization to penalize deviation from initial property values to stabilize the inversion and to supplement the limited data with subjective a priori information regarding system properties. The paucity of data precluded use of a highly parameterized hydraulic conductivity field so hydraulic conductivity was represented by zones of piecewise constancy generally aligned with the long axis of the island. Four categories of recharge values, based on land use characteristics, were assigned as shown in Figure 1. The calibration was performed by using a Gauss-Levenberg-Marquardt parameter estimation approach implemented in PEST [*Doherty*, 2010] on a distributed parallel computing cluster at the USGS Wisconsin Water Science Center coordinated with the HTCCondor run management software [*Condor Team*, 2012].

[20] The first stage of calibration was to run SEAWAT with fully coupled groundwater flow and salt transport to “flush” the system over a long period (1000 simulated years) such that the salinity profiles in the wells were reasonably reproduced. This outlined the extent of the freshwater lens atop the seawater at depth in the groundwater system beneath the aquifer. Following this initial calibration, the freshwater-saltwater interface was pinned in place by running SEAWAT only considering flow and not saline transport. In this mode, variable fluid properties reflect the concentration of salt in the water but the position of the freshwater-saltwater interface is not able to move. At this stage, calibration is finally achieved by systematically adjusting hydrogeologic properties including horizontal and vertical hydraulic conductivity in piecewise constant zones and recharge based on land cover zonation. Prior

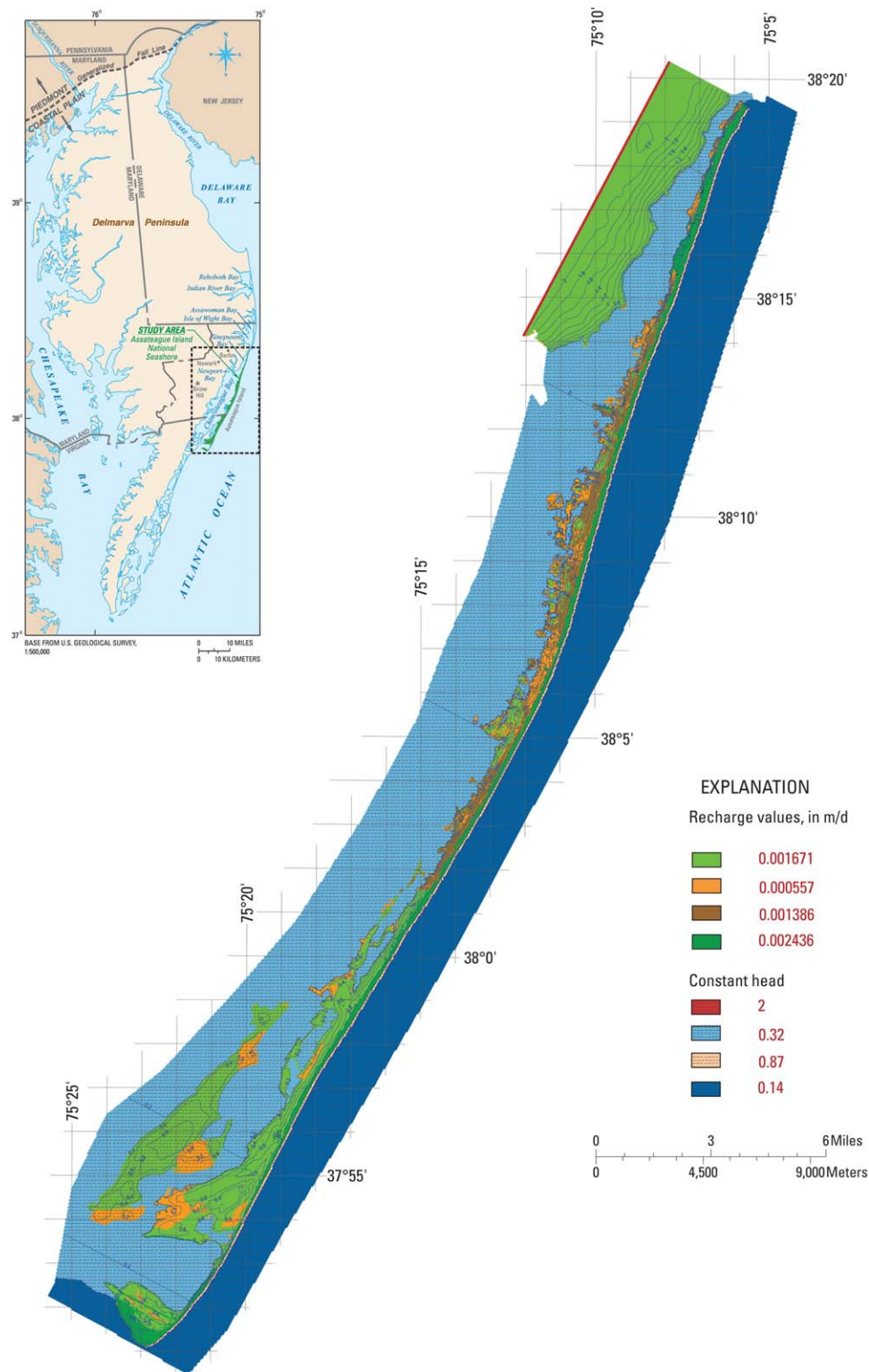


Figure 1. Site location and model overview, depicting recharge zones in the SEAWAT model.

values of the hydraulic parameters and head measurements were used as the calibration constraints.

[21] The head measurements are a snapshot in time despite the short-term transient nature of the system. As a result, obtaining extremely close correspondence between modeled values and head observations would constitute

overfitting and decrease the forecasting value of the model. In a Bayesian sense, the parameters estimated for, and forecasts made with SEAWAT model are conditional upon limited data set available for calibration. Rather than striving to reproduce the individual measurements too closely, the goal in this work was to result in a model that reasonably

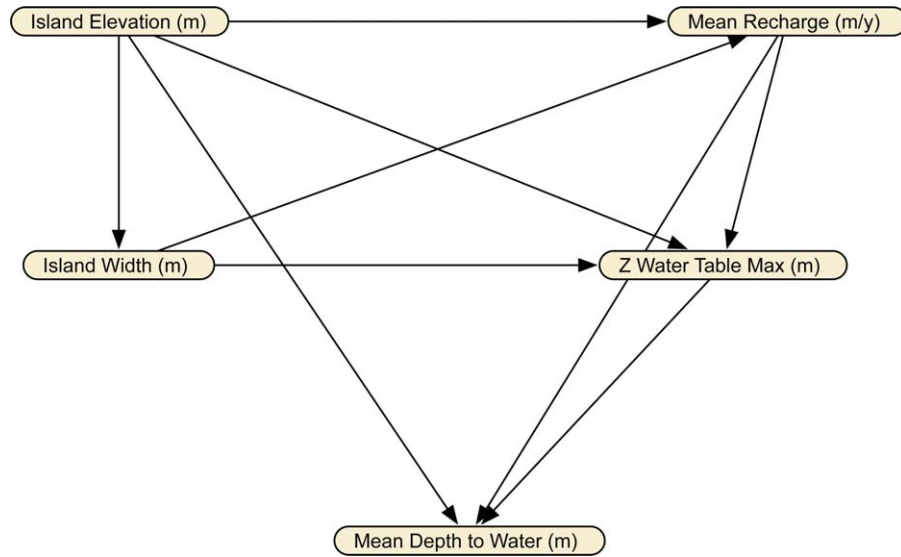


Figure 2. Layout of the Assateague island Bayesian network. Rounded boxes indicate nodes, arrows indicate edges, and arrow direction indicates general causal dependency direction.

captures long-term steady-state behavior of the site. The ability of the model to fulfill this goal is necessarily a somewhat subjective result [Fienen, 2013], but the overall goal is a simulator of sufficient quality to allow extension to similar areas beyond Assateague island.

[22] With the calibrated model, the characteristics of the island and groundwater system were extracted along model rows (generally normal to the long axis of the island model) to use for training the BN. The characteristics at each cross section were island width, maximum island elevation, mean depth to water, maximum water table elevation, and mean recharge. The values corresponding to the maximum or mean of a quantity are calculated by using all model cells along the cross section that are considered active by the groundwater model in the uppermost layer. As discussed in the following section, each of these characteristics was expected to play a role in determining the mean depth to water—the principal model forecast connected to the habitat questions driving this analysis.

4. Methods

[23] A Bayesian network (BN) is a directed acyclic graph [Korb and Nicholson, 2004], composed of nodes and edges. Nodes represent states of parameters or outcomes and can be Boolean or discrete bins. Continuous values of parameters must be discretized into bins. Edges form the connections between nodes and represent a correlated connection between the properties represented by the nodes. To create the BN, parameters that have the potential to drive the forecast of interest are identified and sampled at multiple locations throughout the numerical model domain. The underlying process must be ergodic so that the statistical and correlative characteristics are constant throughout the sample. In other words, the conditions over which the underlying model is built must be representative of conditions over which forecasts will be made with the BN. Minor violations of the ergodic assumption are likely to be encountered, but major violations can only be addressed by

segregating the data into locally (in time or space) ergodic packages. Molina *et al.* [2013] presents an example in which ergodicity was maintained by making BNs at successive time steps over a transient period.

4.1. Bayesian Network Structure

[24] In the case of the Assateague island SEAWAT model, the forecast of interest is mean depth to water along each cross section that can be determined by slicing the model along rows of the computational grid. Assembling both parameters and outcomes from the model in this way resulted in 1152 samples on which to build conditional probability tables (CPTs). The associated parameters are maximum island elevation, island width, mean recharge, and maximum water table elevation. Maximum water table elevation is also a model output but was evaluated as an intermediate forecast in some of the analyses. Figure 2 shows the layout of the dependencies among the parameters and the forecast of interest. The BNs in this work were constructed by using Netica [Norsys Software Corp., 1990–2012].

[25] Calculations are made using the BN based on conditional probabilities using Bayes' Theorem (adopting the symbology of Plant and Holland [2011a])

$$p(F_i|O_j) = \frac{p(O_j|F_i)p(F_i)}{p(O_j)} \quad (1)$$

where $p(F_i|O_j)$ is the posterior probability of a forecast (F_i) given (conditional on) a set of observations, (O_j), $p(O_j|F_i)$ is the likelihood function, $p(F_i)$ is the prior probability of the forecast, and (F_i) and $p(O_j)$ is a normalizing constant. In the remainder of this work, “observations” refer to outcomes of the model rather than the initial calibration data measured in the field. As discussed above, the calibration to field measurements was necessarily loose due to the paucity of data in both space and time. For training a BN, it would be ideal to use direct observations of the underlying system, but that level of information is rarely

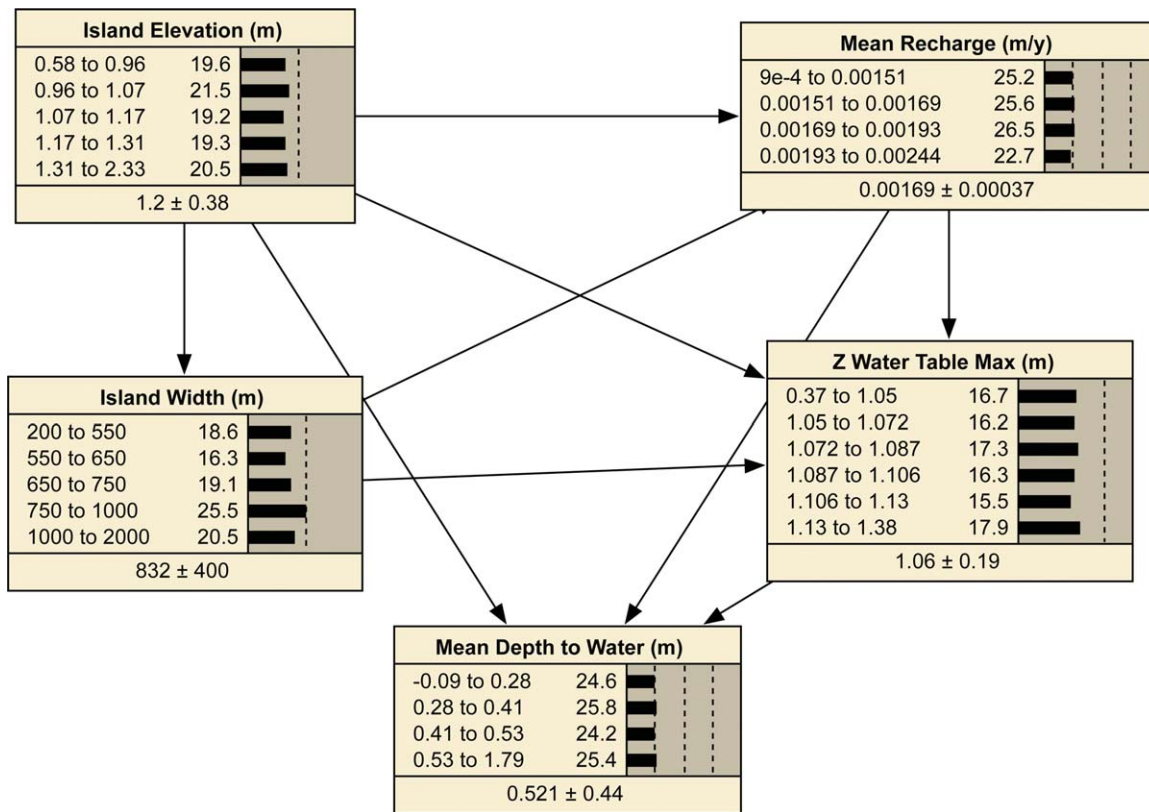


Figure 3. Prior probabilities—expressed as belief bars—for the Assateague island Bayesian network. Mean depth to water is the response variable and the other bins correspond to system states for input. Numbers on the left delineate bin boundaries, numbers on the right show probabilities, and the horizontal black bars graphically show the prior probability values. The numbers at the bottom of each node, including the \pm symbol, indicate the mean value and the associated standard deviation. This arrangement of bins within nodes is considered “optimal” using a combination of quantitative and qualitative metrics as discussed below.

possible and the model must serve as a proxy for the real system. In cases with sufficient data, a direct correspondence between the BN ability to reproduce process model “observations” and the process model ability to reproduce field observations could be assessed. The underlying process model is required to link the observations that can be made (such as island width and island elevation) to those that must be inferred by using the model (such as ground-water position and recharge). The posterior probability reflects an updating that is achieved by considering the entire chain of conditional probabilities of all bins connected to the node representing F_i . The likelihood function represents the probability that the observations (O_j) would be observed given that the forecast was perfectly known. This is a metric of the ability of the BN to function as a forecasting device and imperfections in such forecasts are a function of epistemic uncertainty. Epistemic uncertainty includes uncertainty due to model imperfection, data errors, and other sources. The prior probability of the forecast, $p(F_i)$, is the probability of a forecast without the benefit of the observations and the BN (or a process model or other experiment). $p(F_i)$ may be calculated by using expert knowledge, or may be assumed relatively uninformative to make the entire process as objective as practical (similar to an ignorance prior as in Jaynes and Bretthorst [2003]). A

common prior often used in BNs is the division of a node into bins of equal probability. This is the approach generally followed in this work, resulting in bins of equal probability or “belief” although it is not exactly an ignorance prior because the probability mass in each bin may differ due to variable bin widths. Figure 3 shows the layout of the Assateague island Bayesian network with prior probabilities expressed like histograms as “belief bars.” It is possible to evaluate the contribution to all uncertainty values calculated by the BN by expressing the uncertainty in the prior probabilities. In this work, the model is assumed (for the sake of proving the concept) to be perfect and the only prior variability is a function of sampling each value and assigning it to bins.

[26] Once a system is cast in a BN, new observations of system state are applied and propagated through the BN using Bayes’ theorem such that all forecasts made in the model are contingent upon the specific observations of system state. In other words, each forecast is associated with a specific configuration of observations of system state. Observations are indicated by selecting a bin and forcing the probability of a value in the node to be 100% (Figure 4). When this operation is performed, the Bayesian update propagates in each direction among nodes that are d-connected [Jensen and Nielsen, 2001], updating the

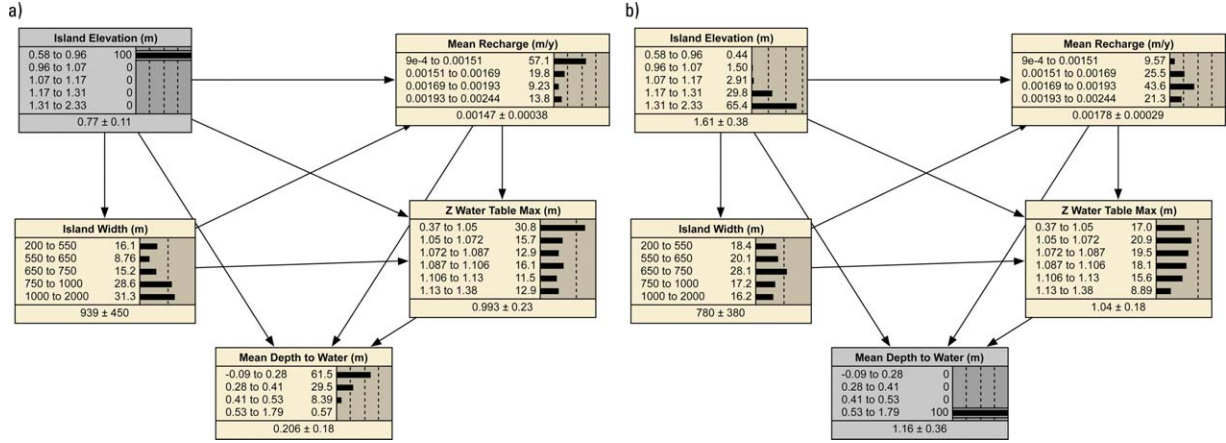


Figure 4. Updated probabilities for the Assateague island Bayesian network. (a) The lowest bin for island elevation has been selected with probability of 100%. All probabilities in the BN are updated by using Bayes' theorem and the updates cascade in a forecasting manner to all bins—including the response variable (Mean Depth to Water). (b) The response variable in the highest bin is selected with probability of 100%. After the updates, the probabilities in the BN reflect the combinations of parameters that are most closely associated with the selected outcome. In this way, the BN acts as a descriptive tool.

probabilities regardless of causal direction. In this way, correlations are expressed as well as causal responses. By selecting a suite of observations of state, the BN functions like a transfer function by providing an estimate of the forecast of interest and associated uncertainty.

[27] A key piece of a priori information is the establishment of edges connecting the nodes. Edges should reflect a cascade of causality grounded in an understanding of the underlying process being modeled. If multiple processes from different models are to be linked, the selection of edge relationships defines the linkage. Similarly, a BN emulating a numerical model could be linked to a simplified BN based on expert knowledge or empirical observations representing a different process. While machine learning can be used to teach a BN which parameters are connected to each other and to outputs, we adopt the more common method in which expert system understanding is used to specify these connections through the identification of nodes and edges. In this way, the BN honors the physical conditions known by the modeler, incorporated as soft knowledge.

[28] In Figure 3, arrows on the edges indicate the direction of causal dependence. When all nodes are d-connected, the direction of the edge arrows serve no purpose. However, in the context of d-separation, the direction of causality has important ramifications on the propagation of uncertainty from observations to forecasts.

[29] When computational conditions and problem size permit, a conditional probability table (CPT) can be created that directly enumerates the conditional probabilities of all nodes in the BN. This becomes impractical rapidly, however, because the size of the CPT scales on the order of $n \times d^{k+1}$ where n is the number of nodes, d is the number of bins, and k is the number of parents for a node. In the case, where full enumeration is impractical due to this rapid increase in computational expense with complexity, an iterative expectation-maximization (EM) algorithm is used [Dempster et al., 1977] to calculate approximate probabilities and maximum-likelihood values for the BN without

full enumeration of the CPT. The EM algorithm iterates between estimating the maximum log likelihood of the function and finding the set of parameters resulting in that maximum log likelihood.

4.2. Performance Metrics

[30] Performance metrics can be applied to evaluate the performance of a BN and to guide in BN design.

[31] Formal sensitivity analysis evaluating the variance reduction in forecasts due to updates of each parent node was used to understand the influence each input variable has on forecasts. This information can also help determine a trade off between sensitivity and optimal bin discretization. Sensitivity is calculated as the percent of variance reduction in a response variable due to updating of a finding.

$$Vr = \frac{V(F) - V(F|O)}{V(F)} \times 100\% \quad (2)$$

where Vr is variance reduction, $V(F)$ is the variance of a forecast prior to update with a finding (observation), and $V(F|O)$ is the variance of the forecast after updating with the observations. $V(F)$ and $V(F|O)$ are calculated as

$$V(F) = \sum_{j=1}^N p(f_j) (f_j - E(f_j))^2 \quad (3)$$

$$V(F|O) = \sum_{i=1}^M \sum_{j=1}^N p(f_j|o_i) (f_j - E(f_j|o_i))^2 \quad (4)$$

where $p(f_j)$ is the prior probability of the j th forecast, f_j is the actual value of the j th forecast, $E(f_j)$ is the expected value (forecast by the BN) of the j th forecast, $p(f_j|o_i)$ is the updated (posterior) probability of the j th forecast given the i th evidence datum, $E(f_j|o_i)$ is the expected value of the j th forecast given the i th evidence datum, M is the number of

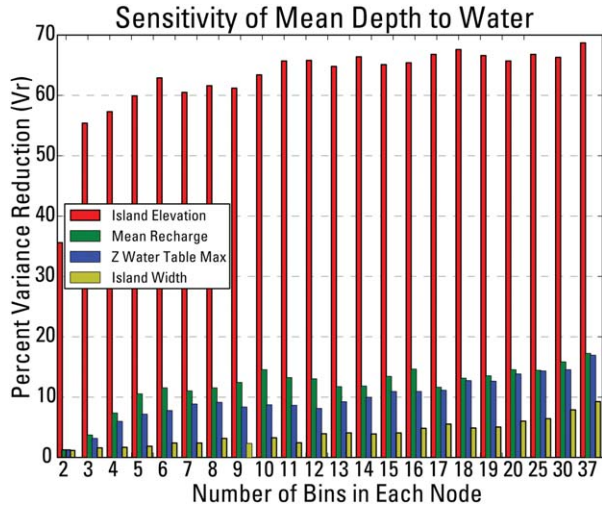


Figure 5. Bar chart illustrating the sensitivity of mean depth to water to updates of other nodes in the BN. The sensitivity metric is percent of variance reduction.

discrete evidence data, and N is the number of discrete forecasts. Finally, the percent variance reduction is calculated as the variance calculated by using observations O from an input node divided by the variance calculated by updating the response variable with findings of itself. By definition, then, Vr for the forecast node is 100%, and all other nodes are less than or equal to 100%.

[32] To systematically evaluate sensitivity as a function of BN complexity, we evaluated a series of BNs all trained on the same groundwater model but using an increasing number of bins in each node. In this systematic approach, bin thresholds were calculated by using auto-discretization with the goal of finding cutoffs between bins that resulted in equal levels of belief, even though bin sizes were variable. Each node was discretized into the same number of bins. Figure 5 shows that sensitivity generally improves with the number of bins (as a proxy for overall model complexity). This metric is prone to overfitting so other diagnostics discussed below—particularly metrics of forecasting power—are more important than sensitivity for BN design. Despite this limitation, the sensitivity provides some insight. For example, the relative rank of the various input variables for forecasting mean depth to water is consistent regardless of model complexity.

[33] Other aspects of complexity in the model could also be evaluated such as the number and nature of causal connections (edges) among the various nodes. In the example of this work, the causal connections are relatively well understood and such evaluation would be more difficult to quantify than the number of bins as a metric of complexity. Nonetheless, considering multiple hypotheses is generally encouraged [Chamberlin, 1890]. When multiple conceptualizations about the underlying process model are present, a more formal—even if qualitative—assessment of model structure is valuable [Pollino et al., 2007; Chen and Pollino, 2012].

[34] In addition to sensitivity, the BN can also be evaluated by using diagnostics of how well it captures the behavior of the model and the quality of forecasts made with it. The two key diagnostic metrics used in this work

are skill and likelihood ratio [e.g., Gutierrez et al., 2011; Plant and Holland, 2011a; Weigend and Bhansali, 1994]. Skill is calculated as

$$sk = \left[1 - \frac{\sigma_e^2}{\sigma_o^2} \right] \times 100\% \quad (5)$$

where σ_e^2 is the mean squared error between observations and BN forecasts, and σ_o^2 is the variance of the observations. Skill expressed the correspondence between observed data and collocated forecasts from the BN and can range from 0 to 100%. This meaning of skill is consistent with the Nash-Sutcliffe model efficiency coefficient [Nash and Sutcliffe, 1970].

[35] The likelihood ratio is calculated as

$$LR_j = \log_{10} \left\{ p(F_i | \tilde{O}_j)_{F_i=O'_j} \right\} - \log_{10} \left\{ p(F_i)_{F_i=O'_j} \right\} \quad (6)$$

where F_i is a forecast, $\sim O_j$ is a subset of observations in the network, withholding the observations directly correlated with the forecast, \tilde{O}_j is an independent observation withheld from the forecast. LR_j , then, expresses the change in the likelihood due to the observations $\sim O_j$ relative to the prior probability of the forecast without the benefit of the network. When $LR_j > 0$, improvement in the forecast due to information in the BN is indicated. Conversely, $LR_j < 0$ indicates degradation of forecast performance due to use of the BN [Plant and Holland, 2011a]. This calculation can be repeated for all forecasts made with the BN and the LR values summed over each class of forecast.

4.3. K-Fold Cross Validation

[36] The diagnostics discussed above are valuable for characterizing how well the BN summarizes the existing data. However, like any system of variable complexity, these diagnostics may be prone to overfitting and exaggerate quality of forecasts. To mitigate this problem, the diagnostics were also calculated in conjunction with K -fold cross validation [e.g., Hastie et al., 2009; Marcot, 2012]. In this approach, the training data are divided evenly into K -folds or partitions, randomly selected (without replacement) from the entire training set. Each fold is made up of n/K values where n is the total number of data points. The sum of all folds thus constitutes the entire data set.

[37] The data are partitioned into two groups: the retained data ($n - n/K$ values) and the left-out fold of n/K values. The BN is then trained to the retained data and diagnostics are calculated only on the left-out data. This approach is similar to partitioning a data set into a training and validation set, but has the advantage of making the most use of the entire data set when calibrating the selected model for use in forecasts. In this work, we analyzed the data using tenfolds. Diagnostics are calculated over the entire calibration data set (leaving out one observation but not retraining the BN, as used in equation (6)) and also over the left out validation folds with the retrained BN. The former is referred to as calibration diagnostics and the latter as validation diagnostics. Table 1 shows the skill metrics as calculated for the 4-bin, 10-bin, 30-bin, and optimal models. The selection of the “optimal” model is discussed

Table 1. Summary of Performance Diagnostics for the *K*-Fold Cross Validation of the BN

Calibration or Validation	Response Variable	Skill (<i>sk</i>) 4 Bin Model (%)	Likelihood Ratio (<i>LR</i>) 4 Bin Model	Skill (<i>sk</i>) 10 Bin Model (%)	Likelihood Ratio (<i>LR</i>) 10 Bin Model	Skill (<i>sk</i>) 30 Bin Model (%)	Likelihood Ratio (<i>LR</i>) 30 Bin Model	Skill (<i>sk</i>) Optimal Model (%)	Likelihood Ratio (<i>LR</i>) Optimal Model
Calibration	Mean Depth to Water	68	336	82	854	97	1488	69	369
Validation	Mean Depth to Water	64	237	55	−95	13	−22	65	204
Calibration	Z Water Table Max	11	155	43	763	86	1455	12	245
Validation	Z Water Table Max	7	1	6	−347	7	−6	6	−105

below. The 4-bin, 10-bin, and 30-bin models were constructed such that each node was automatically discretized into 4, 10, and 30 bins, respectively. The validation results reflect *K*-fold cross validation using tenfolds. The principal model forecast of interest is the mean depth to water, but these metrics were also calculated for the intermediate variable of the maximum water table elevation.

5. Results and Discussion

[38] Emulation of a numerical groundwater model at Assateague island by using a BN serves two main purposes. First, it allows forecasts to be made, including propagation of uncertainty from inputs to forecasts, in an efficient manner, obviating the runtime requirements of the full numerical model in making forecasts. Second, enabled by casting results in terms of probability distributions, the salient characteristics of the groundwater system can be coupled with other process models in a decision support framework. The main response variable—mean depth to groundwater calculated along cross sections transecting the island—was selected for its potential importance in both forage habitat control for Piping Plovers and other ecological impacts including the presence of freshwater ponds that support a wild-horse population and the vegetative cover of the island [Kirwan *et al.*, 2007; O’Connell *et al.*, 2012].

5.1. Performance Metrics

[39] The diagnostics described in section 4 provide metrics on model performance both in terms of describing the training data and evaluating the quality of forecasts outside the training data.

[40] In areas of higher island elevation, the depth from the surface to the water table is generally greater. The dominance of island elevation in model sensitivity (Figure 5) reflects this; the water table position is greatly impacted by boundary conditions at the sea and bay edges of the island. As a result, water table position does not mound in the interior areas of the island in direct connection with land surface, but rather is muted because the boundaries are dominant. As island elevation increases, then, the depth to water is expected to also increase. Recharge (which is tied to land cover) also impacts water table position to a lesser extent than the boundaries so its sensitivity is similar to that of water table elevation itself. However, the importance of these variables remains much lower than island elevation. Island width is the least sensitive variable in the BN. Island width is not as directly connected to water table position and varies somewhat independently from island height.

[41] Figure 6 shows the improvement in skill obtained by increasing the number of bins for each node, as illus-

trated above for sensitivity analysis. For the tenfolds, *sk* values are averaged over all validation folds for comparison with the skill over the calibration set. For consistency, the *LR* values are summed. The ability of the BN to correctly assign results into bins increases rapidly at the lower number of bins and levels off with an increasing number of bins. Table 1 shows a subset of metrics for both calibration and validation data. Figure 6 graphically depicts the skill metrics over a wide range of bins. Improvement over the calibration set continues with adding bins—at the limit of a number of bins equal to the number of unique values in each bin, *sk* over the calibration should increase to 100%. This perfection of categorization, however, comes at a cost in terms of forecasting power.

[42] In Figure 6, for the 2-bin and 3-bin models, forecasting skill (as measured by *sk* over the validation set) increases with calibration skill. With increasing model complexity, these metrics diverge as closer adherence to the calibration data results in poorer performance in forecast. The model that best fits the calibration data is not useful in making forecasts. Choosing an optimal model in terms of complexity involves finding the model with the optimal trade off between calibration and validation skill. The optimal model—in the spirit of Occam’s Razor—is the simplest model that adequately explains the data. This is similar to seeking the minimum message length (MML) in information theory [Wallace and Boulton, 1968]. The longer a message is, the more information it contains (less error of description) but the more prone to error is its transmission (greater error of forecast). In the MML context, the simplest model is sought that has an acceptable trade off between calibration and validation skills. Expressed differently, the optimal model minimizes the combination of descriptive error and forecasting error. Figure 7 shows a schematic diagram of this tradeoff. The “U” shaped curve is upside-down in Figure 6 because we are measuring skill rather than error, but the optimal point is similar.

[43] Using this preliminary result as a guide, several candidate bin discretization results were evaluated, starting with auto-discretized bins of various number between 3 and 6 in each node, and adjusting bin cutoffs ad hoc. This process resulted in six additional models which were then evaluated by using the sensitivity and skill metrics.

[44] Figures 8 and 9 show the metrics of sensitivity and skill calculated for the six ad hoc BNs created manually with a level of complexity approximately commensurate with the 4-bin BN. The performance of these BNs is similar, with model 3 outperforming the others modestly in terms of both skill metrics and sensitivity. Using this two-step screening approach, model 3 was selected as optimal. Model 3 used auto-discretized bins ranging between

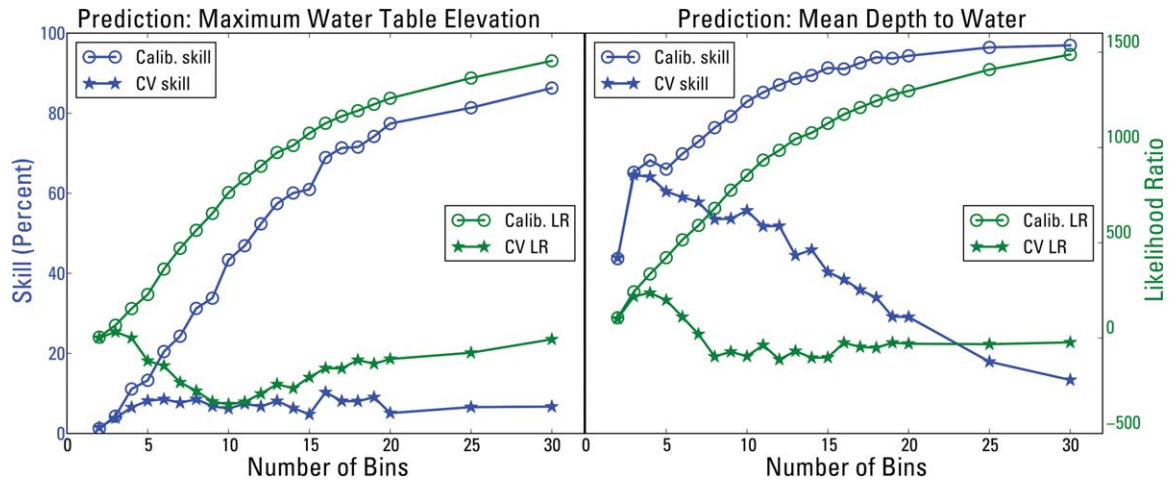


Figure 6. Figures showing performance diagnostics for two response variables. On the left plot, Z Water Table Max is shown and on the right plot is Mean Depth to Water. Skill values correspond to the leftmost (blue) vertical axis and likelihood ratio (LR) values correspond to the rightmost (green) vertical axis. In the legend, “Calib.” and “CV” refer to calibration and validation, respectively.

four and six bins per node. All six candidate models had similar characteristics of approximately four bins per node and in a couple cases manual adjustment was made of bin thresholds. Deciding among these options highlights the value of the two-step approach using the skill and sensitivity metrics at each step.

[45] Figure 10 shows a comparison between the observed and forecasted mean depth to water along the long axis of Assateague island. With minor exceptions, the greatest discrepancies between observed and modeled values occur where the forecast uncertainty is the greatest. Thus, not only is the forecasting capability optimized, but also the forecast uncertainty reflects actual forecast errors. Figure 11 highlights the improved agreement between the observed and forecast values. The great improvement in the 30 bin case compared with the 4 bin case highlights the overfitting potential with a large number of bins.

[46] The groundwater model only explicitly considers the land surface elevation in evapotranspiration calculations. Recharge is applied at the groundwater table and is a function of land cover, subdivided into wetlands, pine for-

est, grass, cultural, and open sand. The vegetative cover is correlated with island elevation.

[47] Using the BN, it is possible to explore the response of mean depth to water to any input or combination of input conditions. For example, in Figure 12a shows the forecast of mean depth to water where the island is wide. Comparing this to Figure 4a, the low sensitivity of island width relative to island elevation, as discussed above, is noted as the spread of the belief bars among the bins changes little from the prior condition (Figure 3) due to updating island width. In Figure 12b, when knowledge of both island elevation and island width are included, the certainty of a specific forecast (high mean depth to water) is greater than 90%, suppressing the likelihood of a forecast in any of the other three bins. Using the BN in this way, the response of mean depth to water to future island morphologic changes due to storms and sea level rise can be readily evaluated.

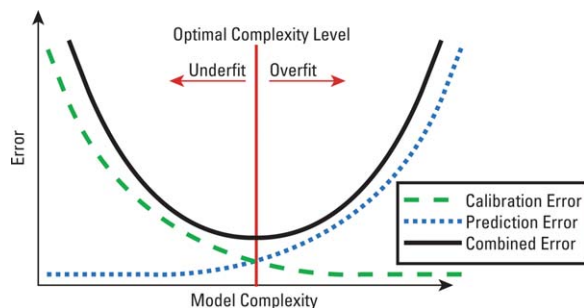


Figure 7. Schematic diagram of the minimum message length (MML) concept showing the optimal tradeoff between calibration error and forecasting error.

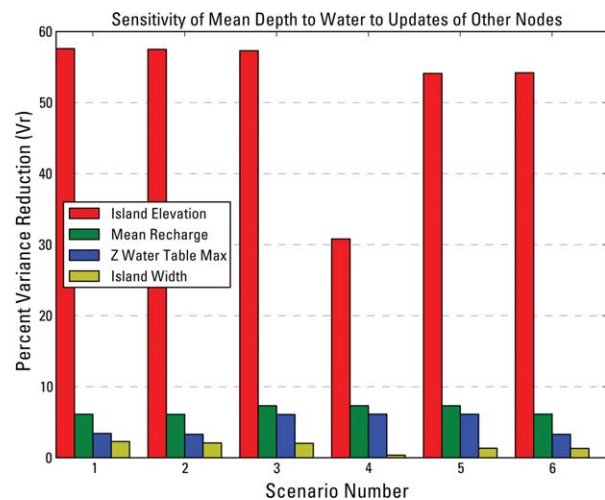


Figure 8. Bar chart illustrating the sensitivity of mean depth to water to updates of other nodes in the BN. The sensitivity metric is percent of variance reduction.

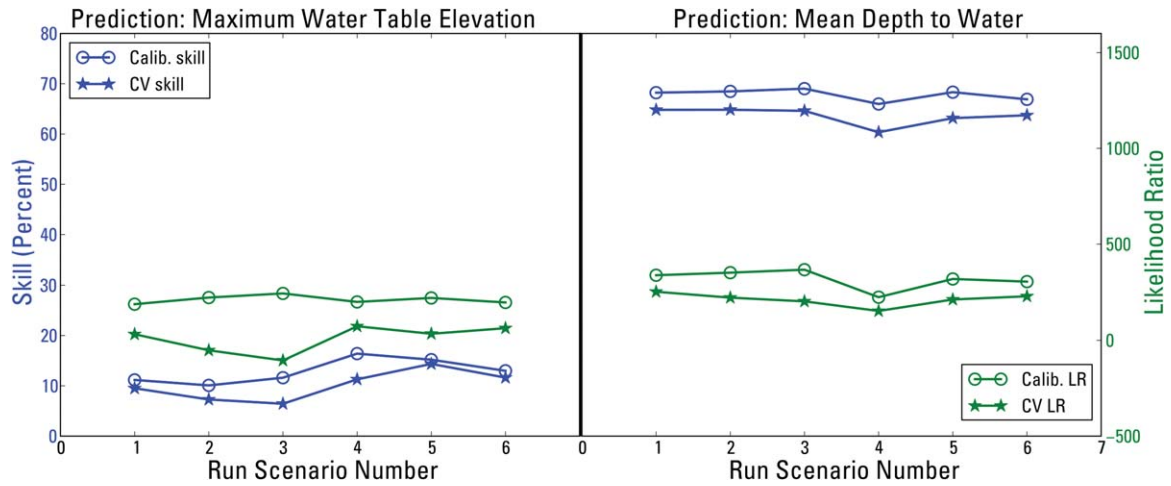


Figure 9. Figures showing performance diagnostics for two response variables for six scenarios focused around 4–6 bins in each node. On the left plot, Z Water Table Max is shown and on the right plot is Mean Depth to Water. Symbology is the same as Figure 6.

5.2. Bin Selection

[48] Our analysis with respect to complexity relies heavily on autodiscretization of bins because the only motivation for bins was the discretization of a continuous variable. However, in many cases, natural thresholds (discontinuities) are evident from the physical environment in which the problem is set and explicit consideration of these thresholds can be extremely important [e.g., *Fienen et al.*, 2004, 2009]. Management decisions may also be based on specific thresholds and, in such cases, bin selection is motivated by those thresholds.

5.3. Implications for Sea Level Rise

[49] Waves and currents associated with future sea level rise will impart clear morphological changes on the island,

including changes in width and elevation that we identify here as principal controls on Vadose zone thickness. Some areas will become narrower as the ocean-side shoreline erodes, and where marsh growth is not sustained [e.g., *Cahoon et al.*, 2009]. Others may become wider as storm overwash and ephemeral tidal inlets move sediment across the island and into the backbarrier lagoon. Island height may increase or decrease depending on sediment availability, and changes in storm climatology that determine the length of time over which features like dunes can become established [*FitzGerald et al.*, 2008; *Gutierrez et al.*, 2009].

[50] As the sensitivity analysis presented above shows, the depth to water is less sensitive to island width than to island elevation. These relationships are based on an

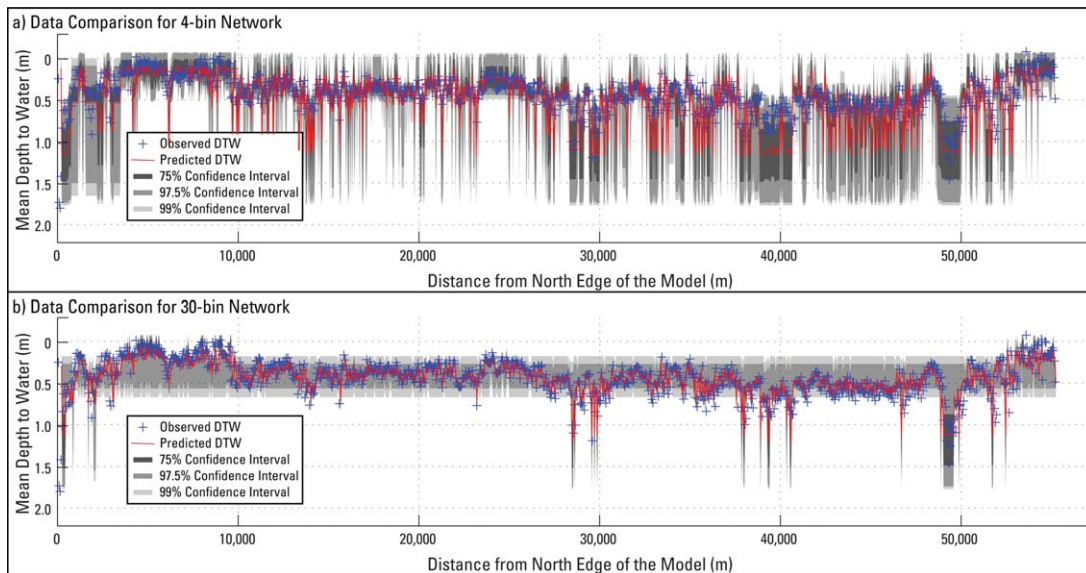


Figure 10. Comparison of observed and forecast mean depth to water along the long axis of Assateague island from North (left) to South (right). The uncertainty of forecasts is also depicted showing 75, 97.5, and 99% confidence intervals. Comparisons are made over the calibration data set (hindcast) with (a) four auto-discretized bins in each node and (b) 30 auto-discretized bins in each node.

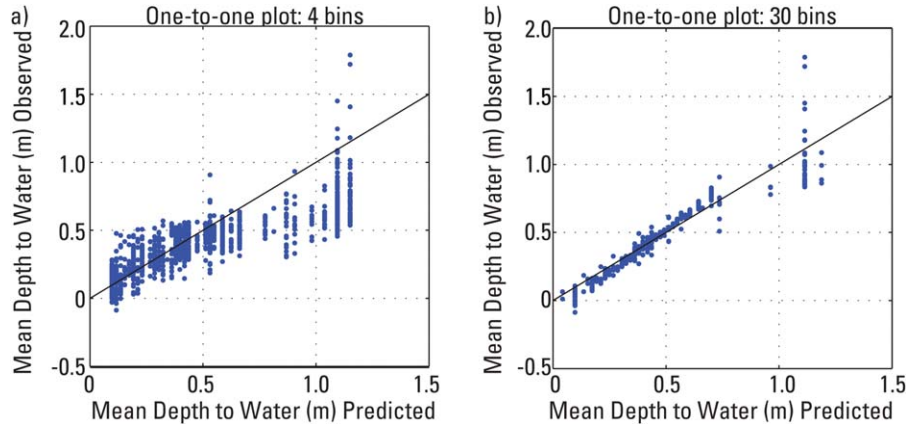


Figure 11. Plot of observed versus forecast mean depth to water. The solid line indicates one-to-one correspondence between observed and modeled results.

ergodic assumption. This assumption may be violated in some ways. For example, in addition to recharge changing from one current category of land cover to another, the recharge within those categories may change due to climate variability. Given a significant enough departure from the current characteristics, a BN would need to be built on a model incorporating those changes that stray beyond the ergodic range of current variability captured by sampling various parts of the island. Nonetheless, the power of the BN to forecast future conditions is enhanced at Assateague island because areas of the island reflect various stages in barrier island evolution. Therefore, our BN is well suited for sampling potential past, current, and future conditions, thus widening the ergodic range. Incorporating more detail about hydrogeomorphic response will enhance the ability of the BN to forecast a wider range of potential conditions reacting to rising sea level. Specifically, this holistic decision support system can serve as a management tool for predicting vegetative and other habitat responses to sea level rise in which depth to groundwater is an important element.

[51] An important aim of this work is to validate the emulation of a model for which intuitive confirmation

would be possible. If the behavior of a numerical model can be successfully emulated with this approach, it provides confidence that more complicated processes may also be emulated in this way.

6. Conclusions

[52] The BN for Assateague island captures the numerical groundwater model with descriptive (calibration) skill of 69% and forecasting skill of 68% for forecasts of mean depth to water. *K*-fold cross validation was implemented to parse forecasting skill from descriptive skill and to guide selection of an appropriate level of complexity for the BN. Indeed, as complexity increases, descriptive skill will converge on perfection (100%) but at great cost to forecasting skill. The estimated uncertainty of forecasts made using the BN is consistent with forecast errors.

[53] The level of forecasting power achieved in this work makes it possible to evaluate groundwater responses to changes in island morphology and recharge characteristics within the ergodic range upon which the model was based. With the ultimate goal of forecasting groundwater responses to sea level rise, the island width and elevation

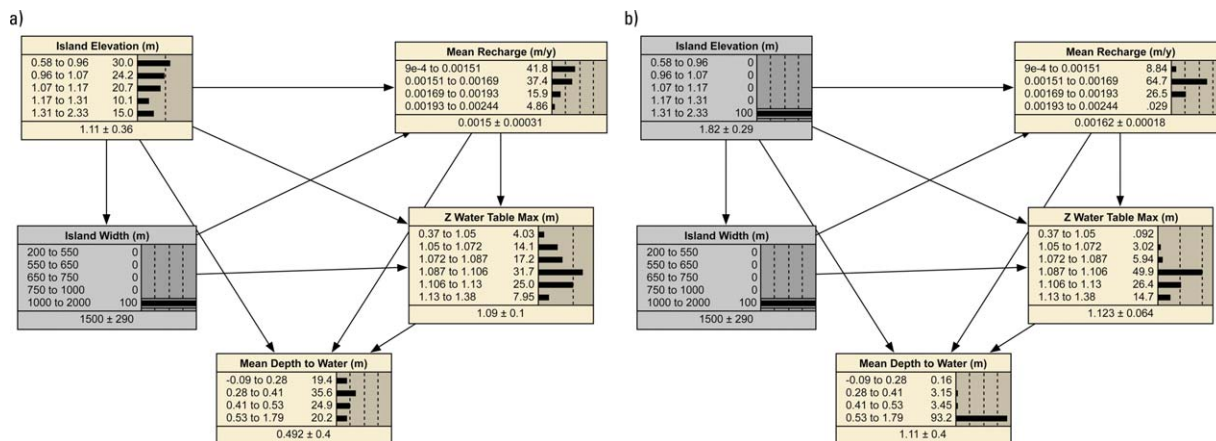


Figure 12. Example scenario evaluation using the optimal Assateague island Bayesian network: (a) the response to the greatest island width is evaluated and (b) this condition is further refined by also selecting the greatest island elevation.

are expected to show the most dramatic and rapid changes (decreases in both) with the rising of sea level around the island. The island elevation was the most sensitive parameter in forecasting depth to water using the BN.

[54] The high sensitivity of island elevation for forecasting depth to water highlights the importance of island morphology to assessing responses to future conditions. Assateague Island—like all barrier islands—is shaped by storms, washover events, and sea level rise. This is just one example of the feedbacks among processes that have important linkages to be considered for a holistic forecasting model that can include geomorphic and climate changes. By representing multiple process models with BNs, they can be linked using edges connecting their nodes in the same way nodes are connected within each individual process BN. Using a BN in this way bridges the gap between the full detail (and computational expense) of a process model and the efficiency and interconnections with decision metrics and other processes needed in a DSS.

[55] Assateague island provides an opportunity to sample conditions representative of current, past, and future conditions. Similar barrier islands throughout the world are morphologically similar to conditions found on Assateague island. As a result, a BN trained to conditions on Assateague island is extensible to a wide range of barrier island systems. Nonetheless, long-term forecasts including both climate change and sea level rise will eventually leave the ergodic range on which the current model was built. For example, changes in temperatures, precipitation, and land cover will impact evapotranspiration and recharge in the groundwater model. The nature of storms including frequency and intensity may also change in the future, impacting the hydrogeomorphic processes. Running each process model with loose coupling will address the ergodicity issue such that the various models exchange their mutual feedbacks at specified intervals over time. This can be done using explicit connections and casting the entire set of process models as a directed acyclic graph. The suite of results, ultimately, will be represented by a linked BN providing an integrated forecasting tool considering all the relevant processes and their uncertainties.

[56] **Acknowledgments.** This work was funded by the USGS Climate and Land Use Mission Area, Research and Development Program and the USGS Natural Hazards Mission Area, Coastal and Marine Geology Program. The authors are grateful to the National Park Service and U.S. Fish and Wildlife Service for collaboration, data, and access to Assateague Island. Constructive reviews by Stacey Archfield, Tony Jakeman, two anonymous reviewers, and the journal editors improved the manuscript.

References

- Banks, W., J. Masterson, and C. Johnson (2012), Well network installation and hydrogeologic data collection, Assateague Island National Seashore, Worcester County, Maryland, 2010, U.S. Geol. Surv. Open File Rep., 2012–5079, 20 p.
- Barlow, P. (2003), Ground water in freshwater-saltwater environments of the Atlantic coast, *Circ. 1262*, 113 pp., U.S. Geol. Surv., Reston, Va.
- Barthel, R., S. Janisch, N. Schwarz, A. Trifkovic, D. Nickel, C. Schulz, and W. Mauser (2008), An integrated modelling framework for simulating regional-scale actor responses to global change in the water domain, *Environ. Modell. Software*, 23(9), 1095–1121, doi:10.1016/j.envsoft.2008.02.004.
- Cahoon, D., D. Reed, A. Kolker, M. Brinson, J. Stevenson, S. Riggs, R. Christian, E. Reyes, C. Voss, and D. Kunz (2009), *Coastal Wetland Sustainability*, pp. 57–72, The U.S. Climate Change Science Program, U.S. Environ. Prot. Agency, Washington, D. C.
- Carmona, G., J. L. Molina, J. Bromley, C. Varela-Ortega, and J. L. García-Aróstegui (2011), Object-oriented Bayesian networks for participatory water management: Two case studies in Spain, *J. Water Resour. Plann. Manage.*, 137(4), 366–376, doi:10.1061/(asce)wr.1943-5452.0000116.
- Castelletti, A., S. Galelli, M. Ratto, R. Soncini-Sessa, and P. C. Young (2012), A general framework for dynamic emulation modelling in environmental problems, *Environ. Modell. Software*, 34, 5–18, doi:10.1016/j.envsoft.2012.01.002.
- Chamberlin, T. C. (1890), The method of multiple working hypotheses, *Science (Old Series)*, 15(92), 754–758, doi:10.1126/science.148.3671.754.
- Chen, S. H., and C. A. Pollino (2012), Good practice in Bayesian network modelling, *Environ. Modell. Software*, 37, 134–145, doi:10.1016/j.envsoft.2012.03.012.
- Condor Team (2012), Condor version 7.6.6 manual, Tech. Rep., University of Wisconsin, Madison.
- Davis, R. (1994), *Barrier Island Systems—A Geologic Overview*, pp. 435–456, Springer, New York.
- Dean, R., and M. Perlin (1977), Coastal engineering study of ocean city inlet, Maryland, *Proceedings of Coastal Sediments, American Society of Civil Engineers*, 77, pp. 520–542.
- Dempster, A., N. Laird, and D. Rubin (1977), Maximum likelihood from incomplete data via EM algorithm, *J. R. Stat. Soc., Ser. B*, 39(1), 1–38.
- Doherty, J. (2010), *PEST, Model-Independent Parameter Estimation—User Manual*, 5th ed., Watermark Numerical Computing in Brisbane, Australia.
- Doherty, J., and S. Christensen (2011), Use of paired simple and complex models to reduce predictive bias and quantify uncertainty, *Water Resour. Res.*, 47, W12534, doi:10.1029/2011WR010763.
- Ehrenfeld, J. G. (1990), Dynamics and processes of barrier-island vegetation, *Rev. Aquat. Sci.*, 2(3–4), 437–480.
- Essink, G. H. P. O., E. S. van Baaren, and P. G. B. de Louw (2010), Effects of climate change on coastal groundwater systems: A modeling study in the Netherlands, *Water Resour. Res.*, 46, W00F04, doi:10.1029/2009WR008719.
- Fienen, M., P. Kitanidis, D. Watson, and P. Jardine (2004), An application of Bayesian inverse methods to vertical deconvolution of hydraulic conductivity in a heterogeneous aquifer at Oak Ridge National Laboratory, *Math. Geol.*, 36(1), 101–126, doi:10.1023/B:MATG.0000016232.71993.bd.
- Fienen, M., R. Hunt, D. Krabbenhoft, and T. Clemo (2009), Obtaining parsimonious hydraulic conductivity fields using head and transport observations: A Bayesian geostatistical parameter estimation approach, *Water Resour. Res.*, 45, W08405, doi:10.1029/2008WR007431.
- Fienen, M., J. Doherty, R. Hunt, and H. Reeves (2010), Using prediction uncertainty analysis to design hydrologic monitoring networks: Example applications from the Great Lakes Water Availability Pilot Project, *Sci. Invest. Rep. 2010–5159*, 44 pp., U.S. Geol. Surv., Reston, Va.
- Fienen, M. N. (2013), We speak for the data, *Groundwater*, 51(2), 157, doi:10.1111/gwat.12018.
- FitzGerald, D. M., M. S. Fenster, B. A. Argow, and I. V. Buynevich (2008), Coastal impacts due to sea-level rise, *Annu. Rev. Earth Planet. Sci.*, 36(1), 601–647, doi:10.1146/annurev.earth.35.031306.140139.
- Guillaume, J. H. A., M. E. Qureshi, and A. J. Jakeman (2012), A structured analysis of uncertainty surrounding modeled impacts of groundwater-extraction rules, *Hydrogeol. J.*, 20(5), 915–932, doi:10.1007/s10040-012-0864-0.
- Gusyev, M. A., H. M. Haitjema, C. P. Carlson, and M. A. Gonzalez (2012), Use of nested flow models and interpolation techniques for science-based management of the Shenyne National Grassland, North Dakota, USA, *Ground Water*, 51, 414–420, doi:10.1111/j.1745-6584.2012.00989.x.
- Gutierrez, B., S. Williams, and E. Thieler (2009), *Ocean Coasts*, pp. 43–56, The U.S. Climate Change Science Program, U.S. Environ. Prot. Agency, Washington, D. C.
- Gutierrez, B. T., N. G. Plant, and E. R. Thieler (2011), A Bayesian network to predict coastal vulnerability to sea level rise, *J. Geophys. Res.*, 116, F02009, doi:10.1029/2010JF001891.
- Halsey, S. (1978), Late quaternary geologic history and morphological development of the barrier island system along the Delmarva Peninsula of the mid-Atlantic Bight, PhD dissertation, University of Delaware, Halsey.
- Hastie, T., R. Tibshirani, and J. H. Friedman (2009), *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer series in statistics*, 2nd ed., xxii, 745 pp., Springer, New York.

- Hayden, B. P., M. Santos, G. F. Shao, and R. C. Kochel (1995), Geomorphological controls on coastal vegetation at the virginia-coast-reserve, *Geomorphology*, 13(1–4), 283–300, doi:10.1016/0169-555x(95)00032-z.
- Hayes, M. O. (1979), *Barrier Island Morphology as a Function of Tidal and Wave Regime*, pp. 211–236, Academic, New York.
- Henry, M., and W. Dennis (1947), *Misty of Chincoteague*, 173 pp., Rand McNally, Chicago, Ill.
- Jaynes, E. T., and G. L. Bretthorst (2003), *Probability Theory: The Logic of Science*, xxix, 727 pp., Cambridge Univ. Press, Cambridge.
- Jenkins, C. T. (1968), Techniques for computing rate and volume of stream depletion by wells, *Ground Water*, 6(2), 37–46, doi:10.1111/j.1745-6584.1968.tb01641.x.
- Jensen, F. V., and T. D. Nielsen (2001), Bayesian networks and decision graphs, in *Statistics for Engineering and Information Science*, xv, 268 pp., Springer, New York.
- Jevrejeva, S., J. C. Moore, and A. Grinsted (2010), How will sea level respond to changes in natural and anthropogenic forcings by 2100?, *Geophys. Res. Lett.*, 37, L07703, doi:10.1029/2010GL042947.
- Kennedy, M. C., and A. O'Hagan (2001), Bayesian calibration of computer models, *J. R. Stat. Soc. Ser. B Stat Methodol.*, 63, 425–450, doi:10.1111/1467-9868.00294.
- Kirwan, M. L., J. L. Kirwan, and C. A. Copenheaver (2007), Dynamics of an estuarine forest and its response to rising sea level, *J. Coastal Res.*, 23(2), 457–463, doi:10.2112/04-0211.1.
- Kitanidis, P. K. (1995), Quasi-linear geostatistical theory for inverting, *Water Resour. Res.*, 31(10), 2411–2419, doi:10.1029/95WR01945.
- Korb, K. B., and A. E. Nicholson (2004), Bayesian artificial intelligence, in *Chapman & Hall/CRC Series in Computer Science and Data Analysis*, 364 pp., Chapman and Hall, Boca Raton, Fla.
- Krantz, D., C. Schupp, C. Spaur, J. Thomas, and D. Wells (2009), *Dynamics Systems at the Land-Sea Interface*, pp. 211–248, University of Maryland Center for Environmental Science, IAN Press, Cambridge, Md.
- Langevin, C., D. J. Thorne, A. Dausman, M. Sukop, and W. Guo (2007), Seawat version 4: A computer program for simulation of multi-species solute and heat transport, U.S. Geol. Surv. Tech. Methods Rep. Book 6, Chapter A22, Reston, Va.
- Leatherman, S. P. (1979), Migration of assateague island, Maryland, by inlet and overwash processes, *Geology*, 7(2), 104, doi:10.1130/0091-7613(1979)7<104:moaimb>2.0.co;2.
- Li, G. Y., C. Rosenthal, and H. Rabitz (2001), High dimensional model representations, *J. Phys. Chem. A*, 105(33), 7765–7777, doi:10.1021/jp010450t.
- Marcot, B. G. (2012), Metrics for evaluating performance and uncertainty of Bayesian network models, *Ecol. Modell.*, 230, 50–62, doi:10.1016/j.ecolmodel.2012.01.013.
- Markstrom, S., R. Niswonger, R. Regan, D. Prudic, and P. Barlow (2008), GSFLOW-coupled ground-water and surface-water FLOW model based on the integration of the Precipitation-Runoff Modeling System (PRMS) and the Modular Ground-Water Flow Model (MODFLOW-2005), U.S. Geol. Surv. Tech. Methods Rep. Book 6, Chapter D1, Reston, Va.
- Martin de Santa Olalla, F., A. Dominguez, F. Ortega, A. Artigao, and C. Fabeiro (2007), Bayesian networks in planning a large aquifer in Eastern Mancha, Spain, *Environ. Modell. Software*, 22(8), 1089–1100, doi:10.1016/j.envsoft.2006.05.020.
- Masterson, J. P., and S. P. Garabedian (2007), Effects of sea-level rise on ground water flow in a coastal aquifer system, *Ground Water*, 45(2), 209–217, doi:10.1111/j.1745-6584.2006.00279.x.
- Masterson, J. P., M. N. Fioren, D. Gesch, and C. S. Carlson (2013), Simulation of groundwater flow in the shallow aquifer system of assateague island, Maryland and Virginia, U.S. Geol. Surv. Open File Rep. 2013–1111.
- Michalak, A. M. (2008), A Gibbs sampler for inequality-constrained geostatistical interpolation and inverse modeling, *Water Resour. Res.*, 44, W09437, doi:10.1029/2007WR006645.
- Michalak, A. M., and P. K. Kitanidis (2003), A method for enforcing parameter nonnegativity in Bayesian inverse problems with an application to contaminant source identification, *Water Resour. Res.*, 39(2), 1033, doi:10.1029/2002WR001480.
- Molina, J. L., J. Bromley, J. L. García-Aróstegui, C. Sullivan, and J. Benavente (2010), Integrated water resources management of overexploited hydrogeological systems using object-oriented Bayesian networks, *Environ. Modell. Software*, 25(4), 383–397, doi:10.1016/j.envsoft.2009.10.007.
- Molina, J.-L., D. Pulido-Velázquez, J. L. García-Aróstegui, and M. Pulido-Velázquez (2013), Dynamic Bayesian networks as a decision support tool for assessing climate change impacts on highly stressed groundwater systems, *J. Hydrol.*, 479, 113–129, doi:10.1016/j.jhydrol.2012.11.038.
- Morgan, M. G., M. Henrion, and M. Small (1990), *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*, xii, 332 pp., Cambridge Univ. Press, Cambridge, N. Y.
- Najjar, R. G., et al. (2000), The potential impacts of climate change on the mid-Atlantic coastal region, *Clim. Res.*, 14(3), 219–233, doi:10.3354/cr014219.
- Nash, J. E., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models part i—A discussion of principles, *J. Hydrol.*, 10(3), 282–290, doi:http://dx.doi.org/10.1016/0022-1694(70)90255-6.
- National Research Council (2009), *Informing Decisions in a Changing Climate*, 188 pp., The National Academies Press, Washington, D. C.
- Nolan, B. T., R. W. Malone, J. A. Gronberg, K. R. Thorp, and L. Ma (2012), Verifiable metamodels for nitrate losses to drains and groundwater in the Corn Belt, USA, *Environ. Sci. Technol.*, 46(2), 901–908, doi:10.1021/es202875e.
- Norsys Software Corp. (1990–2012), Netica Version 4.16. [Available at <http://www.norsys.com/>].
- Nummedal, D. (1983), *Barrier Islands*, pp. 77–122, CRC Press, Boca Raton, Fla.
- O'Connell, M. J., H. H. Shugart, and G. S. Okin (2012), Measurement and monitoring of barrier island forest sensitivity to ecohydrological change using lidar remote sensing, *J. Coastal Res.*, 28(4), 793–810, doi:10.2112/jcoastres-d-10-00109.1.
- Oertel, G., and J. Kraft (1994), *New Jersey and Delmarva Barrier Islands*, pp. 207–232, Springer, New York.
- Oreskes, N., K. Shrader-Frechette, and K. Belitz (1994), Verification, validation, and confirmation of numerical models in the earth sciences, *Science*, 263(5147), 641–646, doi:10.1126/science.263.5147.641.
- Plant, N. G., and K. T. Holland (2011a), Prediction and assimilation of surf-zone processes using a Bayesian network part i: Forward models, *Coastal Eng.*, 58(1), 119–130, doi:10.1016/j.coastaleng.2010.09.003.
- Plant, N. G., and K. T. Holland (2011b), Prediction and assimilation of surf-zone processes using a Bayesian network part ii: Inverse models, *Coastal Eng.*, 58(3), 256–266, doi:10.1016/j.coastaleng.2010.11.002.
- Poch, M., J. Comas, I. Rodríguez-Roda, M. Sánchez-Marré, and U. Cortés (2004), Designing and building real environmental decision support systems, *Environ. Modell. Software*, 19(9), 857–873, doi:10.1016/j.envsoft.2003.03.007.
- Pollino, C. A., A. K. White, and B. T. Hart (2007), Examination of conflicts and improved strategies for the management of an endangered eucalypt species using Bayesian networks, *Ecol. Modell.*, 201(1), 37–59, doi:10.1016/j.ecolmodel.2006.07.032.
- Rahmstorf, S. (2007), A semi-empirical approach to projecting future sea-level rise, *Science*, 315(5810), 368–370, doi:10.1126/science.1135456.
- Rheinhardt, R. D., and K. Faser (2001), Relationship between hydrology and zonation of freshwater swale wetlands on lower Hatteras island, north Carolina, USA, *Wetlands*, 21(2), 265–273, doi:10.1672/0277-5212(2001)021[0265:rbhazo]2.0.co;2.
- Saddodini, A., R. A. Letcher, A. J. Jakeman, and L. T. H. Newham (2005), A Bayesian decision network approach for assessing the ecological impacts of salinity management, *Math. Comput. Simul.*, 69(1–2), 162–176, doi:10.1016/j.matcom.2005.02.020.
- Scavia, D., et al. (2002), Climate change impacts on us coastal and marine ecosystems, *Estuaries*, 25(2), 149–164, doi:10.1007/bf02691304.
- Schupp, C. A., N. T. Winn, T. L. Pearl, J. P. Kumer, T. J. B. Carruthers, and C. S. Zimmerman (2013), Restoration of overwash processes creates piping plover (*Charadrius melodus*) habitat on a barrier island (Assateague Island, Maryland), *Estuarine Coastal Shelf Sci.*, 116, 11–20, doi:10.1016/j.ecss.2012.07.003.
- Shao, G. F., H. H. Shugart, and D. R. Young (1995), Simulation of transpiration sensitivity to environmental-changes for shrub (*Myrica-Cerifera*) thickets on a Virginia barrier-island, *Ecol. Modell.*, 78(3), 235–248, doi:10.1016/0304-3800(93)e0091-g.
- Sherif, M. M., and V. P. Singh (1999), Effect of climate change on sea water intrusion in coastal aquifers, *Hydrol. Processes*, 13(8), 1277–1287, doi:10.1002/(sici)1099-1085(19990615)13:8<1277::aid-hyp765>3.0.co;2-w.
- Steinschneider, S., A. Polebitski, C. Brown, and B. H. Letcher (2012), Toward a statistical framework to quantify the uncertainties of hydrologic response under climate change, *Water Resour. Res.*, 48, W11525, doi:10.1029/2011WR011318.
- Stiber, N. A., M. J. Small, and M. Pantazidou (2004), Site-specific updating and aggregation of Bayesian belief network models for multiple experts, *Risk Anal.*, 24(6), 1529–1538, doi:10.1111/j.0272-4332.2004.00547.x.

- Therrien, R., R. McLaren, E. Sudicky, and Y.-J. Park (2012), HydroGeoSphere a three-dimensional numerical model describing fully-integrated subsurface and surface flow and solute transport, Tech. Rep., Groundwater Simulations Group.
- Tonkin, M., and J. Doherty (2009), Calibration-constrained Monte Carlo analysis of highly-parameterized models using subspace methods, *Water Resour. Res.*, *45*, W00B10, doi:10.1029/2007WR006678.
- Wallace, C. S., and D. Boulton (1968), An information measure for classification, *Comput. J.*, *11*(2), 185–194, doi:10.1093/comjnl/11.2.185.
- Weigend, A. S., and R. J. Bhansali (1994), Paradigm change in prediction, *Philos. Trans. R. Soc.–Math. Phys. Eng. Sci.*, *348*(1688), 405–420, doi:10.1098/rsta.1994.0100.
- Werner, A. D., and C. T. Simmons (2009), Impact of sea-level rise on sea water intrusion in coastal aquifers, *Ground Water*, *47*(2), 197–204, doi:10.1111/j.1745-6584.2008.00535.x.
- Zheng, C. (1990), MT3D, a modular three-dimensional transport model for simulation of advection, dispersion, and chemical reactions of contaminants in groundwater system, Tech. Rep., U.S. Environ. Prot. Agency. United States Environmental Protection Agency, Ada, Oklahoma.
- Zickfeld, K., A. Levermann, M. G. Morgan, T. Kuhlbrodt, S. Rahmstorf, and D. W. Keith (2007), Expert judgements on the response of the Atlantic meridional overturning circulation to climate change, *Clim. Change*, *82*(3–4), 235–265, doi:10.1007/s10584-007-9246-3.