Coastal Engineering 58 (2011) 119–130

Contents lists available at ScienceDirect

Coastal Engineering

journal homepage: www.elsevier.com/locate/coastaleng

Prediction and assimilation of surf-zone processes using a Bayesian network Part I: Forward models

Nathaniel G. Plant [a,*], K. Todd Holland [b]

[a] U.S. Geological Survey, 600 4th St. S., St. Petersburg, FL 33701, USA
[b] Naval Research Laboratory, Stennis Space Center, MS 39529, USA

ARTICLE INFO

Article history:
Received 9 October 2009
Received in revised form 27 August 2010
Accepted 3 September 2010

Keywords:
Wave height
Bathymetry
Field data
Duck94

ABSTRACT

Prediction of coastal processes, including waves, currents, and sediment transport, can be obtained from a variety of detailed geophysical-process models with many simulations showing significant skill. This capability supports a wide range of research and applied efforts that can benefit from accurate numerical predictions. However, the predictions are only as accurate as the data used to drive the models and, given the large temporal and spatial variability of the surf zone, inaccuracies in data are unavoidable such that useful predictions require corresponding estimates of uncertainty. We demonstrate how a Bayesian-network model can be used to provide accurate predictions of wave-height evolution in the surf zone given very sparse and/or inaccurate boundary-condition data. The approach is based on a formal treatment of a data-assimilation problem that takes advantage of significant reduction of the dimensionality of the model system. We demonstrate that predictions of a detailed geophysical model of the wave evolution are reproduced accurately using a Bayesian approach. In this surf-zone application, forward prediction skill was 83%, and uncertainties in the model inputs were accurately transferred to uncertainty in output variables. We also demonstrate that if modeling uncertainties were not conveyed to the Bayesian network (i.e., perfect data or model were assumed), then overly optimistic prediction uncertainties were computed. More consistent predictions and uncertainties were obtained by including model-parameter errors as a source of input uncertainty. Improved predictions (skill of 90%) were achieved because the Bayesian network simultaneously estimated optimal parameters while predicting wave heights.

Published by Elsevier B.V.

1. Introduction

The coastal environment is characterized by extreme variability. Physical variables that exhibit significant spatial and temporal evolution include waves, currents, and bathymetry. This is particularly true in the surf zone, where waves transition from non-breaking to breaking conditions, transferring momentum to drive currents, sediment transport, and bathymetric change. Numerous numerical geophysical models are available that make predictions of these processes. These process models can also be inherently statistical, wherein only quantities that are formally averaged over several wave periods (and, often, averaged over the water column) are simulated. For instance, temporally averaged statistical properties of waves can be accurately predicted by SWAN (Simulating Waves Nearshore, Booij et al., 1999; Ris et al., 1999) given accurate bathymetry, water levels, description of the frequency-directional spectra at boundary conditions, and specification of a number of tuning parameters. Similarly, wave-averaged currents can be predicted by models such as Delft-3D (Lesser et al., 2004; Reniers et al., 2007) or ADCIRC (Westerink et al., 2008) that have clearly demonstrated predictive skill. At higher resolution, wave-resolving Boussinesq models also have excellent predictive skill (Chen et al., 2003) and an ability to simulate at the shorter time scales given correspondingly high-resolution time-series data on the model boundaries. This increased fidelity may be necessary for predicting sediment transport and bathymetric evolution (Henderson et al., 2004). However, the higher temporal resolution of Boussinesq models comes with increased computational cost along with the more demanding specification of boundary conditions.

At the present level of hydrodynamic modeling capability, improved predictive skill typically depends on improving the accuracy of model-boundary conditions (Plant et al., 2009), rather than on refinements of the model parameterizations, indicating that the geophysical theory governing nearshore processes is relatively mature. Modeling of sediment-transport processes is an exception to this statement, where skillful predictions can depend strongly on choice of model parameterization (Ruessink et al., 2007). Parameter dependence is primarily due to the use of wave-averaged models that do not resolve physically important processes such as the details of the wave boundary layer and higher-than-second-order moments of the near-bed velocities (Henderson et al., 2004; Hsu et al., 2006).

* Corresponding author. Tel.: +1 727 803 8747x3072; fax: +1 727 803 2032.
E-mail addresses: nplant@usgs.gov (N.G. Plant), todd.holland@nrlssc.navy.mil (K.T. Holland).

0378-3839/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.coastaleng.2010.09.003

These neglected details must be absorbed into the available free parameters. The bottom line is that the predictive skill of the present modeling capability will largely depend on uncertainties in model inputs, model parameters, or both.

Finally, real-world implementation of coastal-process models, useful to emergency managers, the military, lifeguards, and beach-goers who need to make decisions based on rapid environmental forecasts, is becoming a necessity. In this scenario, the fidelity of boundary and initial conditions required by numerical models may be so poor that the inadequacies of the input data can only be improved by additional (and often costly) observations of processes in the interior of the model domain. This sets up a complex situation where optimal solutions require the use of some sort of assimilation model to (1) map incomplete and inaccurate data to the boundary conditions, (2) correct model errors in the interior of the domain, and (3) extend intrinsic model outputs (e.g., wave height) to related observables (e.g., video or radar observations, Bell, 1999; van Dongeren et al., 2008) that can contribute to model improvement. Presently, formal data-assimilation schemes that have been applied to nearshore coastal processes have focused primarily on scientific evaluation of model physics (Feddersen et al., 2004), studying the sensitivity of inverse solutions to the scale of variability (Kurapov et al., 2007), and evaluating unknown model parameters (Plant et al., 2004; Ruessink et al., 2007), rather than to solving the operational problem of inadequate boundary conditions.

The lack of regularly applied data assimilation to nearshore coastal models is likely due to computational constraints of the state-of-the-art models, which often perform simulations at computational times that exceed the simulation time. Furthermore, the coupled wave, current, and sediment-transport problem is nonlinear and requires significant model iteration to converge toward the best inverse solutions (Kurapov et al., 2007). To get an idea for the dimensionality of the problem, consider a coastal region that might span 100 km^2, with a model resolution of 100 m^2, which requires evaluation at 1 million grid positions. At each position, there may be roughly 10 variables of interest (e.g., depth, wave height, energy dissipation, etc.). The models might have a time step of 1 s, and a forecast would estimate hourly averaged conditions requiring evaluation of a total of 10^{10} values to describe the coastal environment. A typical assimilation scheme might require 10 iterations, pushing the total bookkeeping load to 10^{11} values per coastal-process forecast.

Although formal data assimilation is possible, its implementation is difficult in coastal and surf-zone environments because (1) it is computationally intensive; (2) it requires specification of often unknown quantities (such as the model error); (3) it is unwieldy in the face of a large number of uncertainties; and (4) it is unmanageable in the face of a large number of observations. Nonetheless, in practical applications of coastal prediction, forecasts will be expected even from boundary and forcing data with substandard accuracy (e.g., outdated bathymetric survey and hydrodynamic observations, or forecasts with significant uncertainty). At a minimum, model initialization will require some form of data assimilation to correct the imperfect model inputs. For instance, data interpolation of one form or another is a commonly used (and abused) form of data assimilation (Ooyama, 1987; Plant et al., 2002). Regardless of approach, practical applications can be supported only if the data-assimilation method offers some reduction in observation and model errors and, most importantly, provides a useful estimate of prediction uncertainties.

In this two-part paper, we present a new application of Bayesian statistical modeling to surf-zone modeling and assimilation. Part I is devoted to describing the modeling framework and applying it to several case studies focused on forward modeling. Part II is devoted to inverse problems and extending the framework to a very flexible set of assimilation problems. We continue with Part I in Section 2 (Model formulation) by reviewing Bayes Rule and then developing a specific application for surf-zone wave modeling. In Section 3 (Application), we describe a data set used for training the Bayesian network and provide hindcast and forecast prediction examples. These examples highlight the role that data and model uncertainty play in affecting prediction errors. Lastly, in Section 4 (Discussion), we explore the problem of obtaining required uncertainty estimates and demonstrate their impact on model skill.

2. Model formulation

2.1. Review of Bayes Rule

In contrast to variational data assimilation, mentioned in the previous section, we suggest that a Bayesian-inference approach can be used to combine available measurements with nearshore-process models to make statistically robust forecasts. The Bayesian approach is formally consistent with sensitivity-based variational assimilation (Wikle and Berliner, 2007). Bayes Rule is

$$p\left(F_i \mid O_j\right) = p\left(O_j \mid F_i\right) p(F_i) \,/\, p\left(O_j\right), \tag{1}$$

where the left-hand side of Eq. (1) is the updated conditional probability of a particular forecast, F_i, given a particular set of observations, O_j. The forecast might include both initial and boundary conditions. The observations might be obtained near the boundaries or in the interior of the model domain. Variational data assimilation strives to return the conditional mean value of the solution: $\overline{F_j} = \Sigma_i\, p\,(F_i \mid O_j)\, F_i$. A Bayesian approach is more general and strives to estimate the likelihood of many possible forecasts. The first term on the right-hand side of Eq. (1) is the inverse of the left-hand side and is the likelihood of the observations if the forecast is known. This term can include both model and observation errors. That is, if the model and measurements were error free, then an observation would be likely only if it equaled the forecast value. In reality, there are numerous errors causing spread in the likelihood function.

The next term on the right-hand side of Eq. (1) is the prior probability of each forecast. This is what is known about the problem before new data are available. It might be the result of a prior assimilation cycle or derived from climatology. The mean value obtained from this distribution is equivalent to the "background" or "best-guess" solution used in variational ocean data assimilation (Benett, 2002) or in optimal interpolation (Ooyama, 1987). Finally, the last term is a normalization factor to account for the total likelihood of the observations. In variational data assimilation and optimal interpolation, this term is solved by inverting the data covariance and is often responsible for large computational costs. Here, it is estimated through integration over all forecast possibilities.

A primary advantage to the Bayesian approach to data assimilation is that the probability distribution of the forecast is estimated. This allows for both data and forecast to have non-Gaussian probability distributions, which may be crucial to describing strongly nonlinear nearshore processes. In principle, this means that the Bayesian approach will lead to more accurate forecast statistics, including estimates of the most likely forecast and its uncertainty. A possible disadvantage of the Bayesian approach is that it may increase the dimensionality of the problem. If each F_i is a single forecast with 10^{10} quantities to describe it (i.e., the typical numerical model system mentioned earlier), then the complete Bayesian approach requires that we track the joint probability of these variables against observations, which might also be any of the 10^{10} quantities. The problem would have exploded to a dimension of 10^{100}!

In practice, however, the utility of the Bayesian approach is most evident when the problem of interest can be boiled down to a much-reduced dimensionality. (This statement actually applies to most other data-assimilation schemes, which, for instance, do not attempt