Environmental Modelling & Software 84 (2016) 482–493

Contents lists available at ScienceDirect

Environmental Modelling & Software

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

A Bayesian method for improving probabilistic wave forecasts by weighting ensemble members

Quillon Harpham*, Nigel Tozer, Paul Cleverley, David Wyncoll, Doug Cresswell

HR Wallingford, Howbery Park, Wallingford, Oxfordshire, OX10 8BA, United Kingdom

ARTICLE INFO

Article history:
Received 29 January 2016
Received in revised form
22 July 2016
Accepted 25 July 2016
Available online 25 August 2016

Keywords:
Bayesian
Ensemble
Feature types
Forecast
Observations
Waves

ABSTRACT

New innovations are emerging which offer opportunities to improve forecasts of wave conditions. These include probabilistic modelling results, such as those based on an ensemble of multiple predictions which can provide a measure of the uncertainty, and new sources of observational data such as GNSS reflectometry and FerryBoxes, which can be combined with an increased availability of more traditional static sensors. This paper outlines an application of the Bayesian statistical methodology which combines these innovations. The method modifies the probabilities of ensemble wave forecasts based on recent past performance of individual members against a set of observations from various data source types. Each data source is harvested and mapped against a set of spatio-temporal feature types and then used to post-process ensemble model output. A prototype user interface is given with a set of experimental results testing the methodology for a use case covering the English Channel.

Software/data availability

The software development was led by Paul Cleverley (paper co-author). Executable code for the extract, transform and load modules together with the statistical model was written in a windows environment with C# under. Net 4.0 using Visual Studio 2012 with a front end written in VB script by Naveed Hussain. Supporting scripts were written in Python using the Enthought EPD python implementation 7.2.1 (64 bit). Intermediate files were stored using CSV and XML. The relational database management system used was PostgreSQL 8.4 plus POSTGIS extensions. The execution environment was a windows server running Windows Server 2003 Standard Edition.

1. Introduction

Forecasts of wave conditions are required for planning of a wide range of weather sensitive maritime operations from construction and maintenance to decommissioning. There is an ever increasing requirement to minimise downtime in order to reduce costs and demand for increasingly more accurate forecasts is one aspect that can help planners make the most appropriate operational decisions. Two independent sets of innovations are now emerging, which offer new opportunities to improve forecasting services.

Traditional wave forecasts provide a single estimate of conditions with a typical outlook of 5–7 days, giving parameters such as significant wave height, maximum wave height, wave period and direction. Such deterministic forecasts provide limited or no information on the potential uncertainty in a given forecast. Probabilistic forecasts, in contrast, such as those based on an ensemble of multiple predictions, not only extend the range of the forecasts often out to 14 days, but also provide a measure of the uncertainty at any given time-step. With increasing computing power, probabilistic forecasts are becoming increasingly common and will no doubt become the norm.

Alongside the increase in availability of ensemble wave forecasts, new and innovative sources of observational data are emerging. Global Navigation Satellite System (GNSS) reflectometry (see for example Gleason et al., 2005) offers the potential for reflected signals from Global Positioning Satellites (GPS) to be used to interpret phenomena such as sea conditions (e.g. wave mean square slope from which other parameters can be inferred). The sensors must be situated in a low enough orbit to receive these signals and so cannot be geostationary, thereby producing a dataset of observations following the track of the satellite carrying the receiver. Back on the surface of the ocean, products such as the

* Corresponding author.
E-mail address: q.harpham@hrwallingford.co.uk (Q. Harpham).

http://dx.doi.org/10.1016/j.envsoft.2016.07.015
1364-8152/

Acronyms and abbreviations

ADCP	Acoustic Doppler Current Profiler
AR	Autoregressive
AWAC	Acoustic Wave and Current Profiler
AVHRR	Advanced Very High Resolution Radiometer
CRPS	Continuous Rank Probability Score
CTD	Conductivity, Temperature and Depth
DDM	Delayed Doppler Map
ECMWF	European Centre for Medium-range Weather Forecasts
GNSS	Global Navigation Satellite System
GPS	Global Positioning Satellites
NOC	National Oceanography Centre

“Ferrybox” (Chelsea Technologies 2012, Hydes et al., 2004; Hydes and Dunning, 2005; Dunning and Hand, 2005) allow sensors to be mounted on moving vessels collecting ocean data parameters, from which typical wave parameters (height, period and direction) can be interpreted. Like the GNSS receiver, such technology results in data along a track, but one far more variable and with a potentially much greater density of readings. As well as these moving devices, there are also an increasing number of static sensors producing oceanographic parameters in time series at fixed locations. Many of these static sensors offer real-time data streams (see for example the Channel Coast Observatory (2014)).

The use of ensembles aims to represent the uncertainty in a forecast using a population of individual ensemble members. Ensemble members may be perturbed instances of the same model – either global or downscaled – (e.g. Saetra and Bidlot (2004), Cao et al. (2009), Behrens (2015)), or of different models (e.g. Durrant et al., 2009) or a combination of the two (e.g. Alves et al., 2013). In weather forecasting historically, ensembles have often been focussed on the uncertainty at the tail end of the forecast window. The separation of ensemble members being relatively small at analysis time (e.g. Saetra and Bidlot (2004), Cao et al. (2009)) and growing as the forecasts progresses. Ensemble spread at analysis time can be achieved using Ensemble Transform techniques, e.g. Bunney and Saulter (2015), Alves et al. (2013), which is useful when the focus is on short term uncertainty.

Motivated by the potential spread of Wave Farms to harness the generating potential of the Ocean, Pinson et al. (2012) introduce a methodology for the probabilistic forecasting of wave energy flux. Using meteorological forecasts from the European Centre for Medium-range Weather Forecasts (ECMWF) and a log-Normal assumption for the shape of predictive densities, benchmarked improvements of between 6 and 70% are shown in terms of Continuous Rank Probability Score (CRPS). However, in studying the effectiveness of the spread of results presented by European ensemble wind speed forecasts, Saunders et al. (2014) concluded that “leading ensemble forecasts of European windspeed often represent uncertainty poorly” and, in particular, that “the miscalibration is worst at shorter lead times and improves at longer forecast lead times”. They also observed that the probabilistic information was very likely to be “erroneous and inaccurate for users”.

The purpose of this paper is to describe the WaveSentry system: a set of components for harvesting observed data sources with different identified characteristics and implementing an application of the Bayesian statistical methodology that modifies the probabilities of ensemble wave forecasts based on recent past performance of individual members against these observations. Portrayal of the result set is also briefly indicated. A set of data sources are introduced followed by characterisation of each from a set of spatio-temporal feature types which facilitates interoperability and extensibility at this level. Components for data collection and incorporation are then described. An example prototype user interface is then given with a set of experimental results testing the methodology for a use case covering the English Channel. The proposed methodology allows for the incorporation of the various different types of observed and modelled data to create ensemble forecasts with potentially enhanced accuracy, in both best estimate and uncertainty, providing added value and confidence to end users. The systems built on the methodology are capable of using ensemble data that may be very time consuming and computationally expensive to produce, while reacting swiftly and efficiently when fresh observations bring in new information. Indeed, the ideas presented here are not confined to ensemble wave forecasts with supporting marine data, they are appropriate to any situation where ensemble model output can be post-processed in this manner.

2. Methods

2.1. Data sources

Applications were written to allow three independent types of measured wave data to be incorporated into post-processing ensemble wave forecasts: GNSS Reflectometry, FerryBox mounted accelerometers and static fixed position devices.

2.2. GNSS reflectometry data

GNSS reflectometry (Gleason et al., 2005) uses a comparatively low cost receiver to pick up backscatter from GPS signals from which sea state data parameters such as ‘mean square slope’ can be derived. With an increasing number of GPS satellite transmitters being deployed, mounting receivers to make this additional use of the GPS signal will lead to a considerable increase in the potential data coverage currently available from the constellation of traditional satellites currently fitted with wave sensing instruments. Near real time observations are available from existing satellites and this technology will translate to the GNSS receivers. This is dependent on full implementation and validation of fast analysis methods that transform the backscatter signals, via a Delayed Doppler Map (DDM), into a useful parameter set including mean square slope and surface wind speed. All parameters derived are presented together along a data track following the satellite receiver path. As different GNSS signals come in and out of range the data received can be patchy with large stretches of the tracks giving no data.

2.3. FerryBox accelerometer

The use of FerryBox vessel mounted instrumentation for observing water quality (i.e. physical, chemical and biological content) is fairly widespread (Chelsea Technologies 2012, Hydes et al., 2004; Hydes and Dunning, 2005; Dunning and Hand, 2005). Adding a low cost accelerometer device from which wave conditions can be derived has the potential to provide relatively wide geographical coverage (Dunning, 2011). By recording vessel motion, application of an inversion routine is required to compute the associated wave parameters. One difficulty is in the derivation of the inversion routine which is vessel specific and is more likely to work well for relatively small vessels that respond to a wide range of sea states. Methods for near real-time signal transmission are