Application of a soft rule-based model to storm surge and sea level variability

Sam Royston*,†, Kevin Horsburgh*, Jonathan Lawry†

*National Oceanography Centre, Liverpool, UK; †Engineering Mathematics, University of Bristol, UK

samyst@noc.ac.uk

Abstract

A rule-based model, founded in Bayesian statistics, is applied to sea level prediction with two aims. The data-driven model uses information theory metrics to prioritise model inputs to create a tree structure of IF-THEN rules. One aim of this method, therefore, is to interpret these rules linguistically, to identify dominant relationships between input data and the data of interest in the physical system and to identify key mechanisms for individual events. Secondly, the model makes probabilistic predictions, and so we aim to utilise the natural error estimates made by the method. Application of the model to sea level problems has been proven on the problem of short-term forecasting of storm surge in the North Sea. The same probabilistic and transparent approach is then applied to the prediction of local annual mean sea level from atmospheric and analogue tide gauge data. The algorithm identifies the data fields providing most information about the system and the probabilistic nature of the model allows for a natural data-driven error estimation in the predictions, which bounds the variability.

1 Decision tree model

An entropy based decision tree algorithm, referred to as LID3[1], is used here to forecast water levels using available input data, for a range of sea level problems. The decision tree algorithm is fuzzy and lies in a Bayesian framework, where the attribute and target data are 'fuzzified' into membership functions on descriptors, as presented in Figure 1 for one of the data sets used in this study, the skew surge at Sheerness.

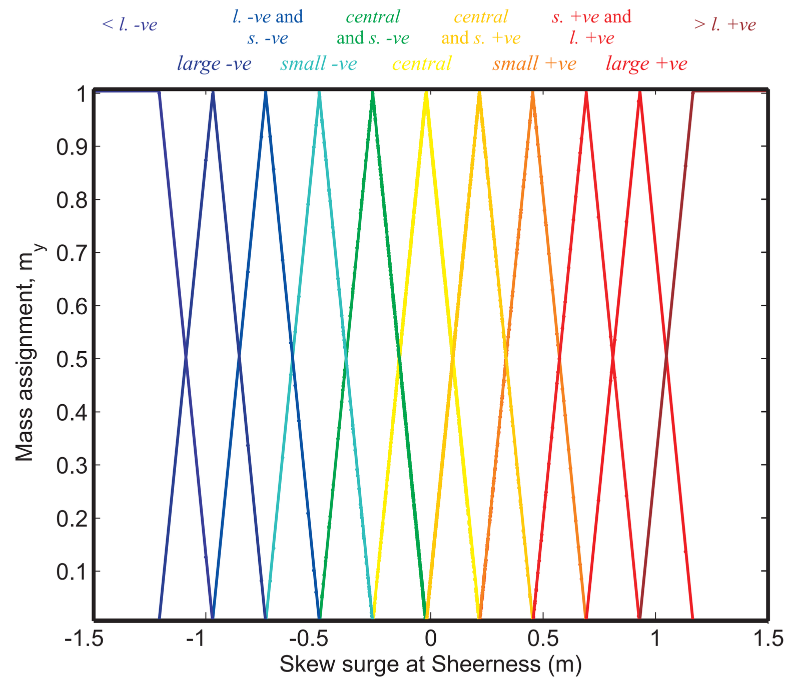

Figure 1: Fuzzy label sets for skew surge at Sheerness (11 no. uniformly-distributed).

The decision tree algorithm uses a 'training' subset of the available data to update non-informative prior probabilities with posterior probabilities based on past events. Given a new input data vector, the tree structure provides a probability distribution of the target descriptions which can subsequently be 'defuzzified' into a real-valued forecast.

2 Storm surge

The applicability of the model is proven on the problem of short-term forecasting of storm surge, in a well understood region (the North Sea). Storm surge may enter the North Sea basin from the north, from external sources, or may develop internally due to wind stress acting on the sea surface, progressing as a coastally-trapped gravity wave cyclonically around the basin.

Table 1: Predictive Accuracy

Method	All Data			Upper 5^{th} Percentile		
	AAE (m)	RMSE (m)	r^2	AAE (m)	RMSE (m)	r^2
Operational model[2]	0.076	0.097	0.52	0.130	0.157	0.21
LID3 decision tree	0.078	0.107	0.42	0.168	0.203	0.06
LLS regression	0.073	0.102	0.45	0.147	0.192	0.09

Tide gauge data from the UK's north-east coast and meteorological data (wind speed and atmospheric pressure) are used as model inputs, to forecast storm surge at Sheerness tide gauge on the Thames Estuary with approximately 7.5 to 8 hours lead-time. Data is taken from 1999 to 2008 (a total of 3332 complete data vectors; 50% for training and 50% for testing).

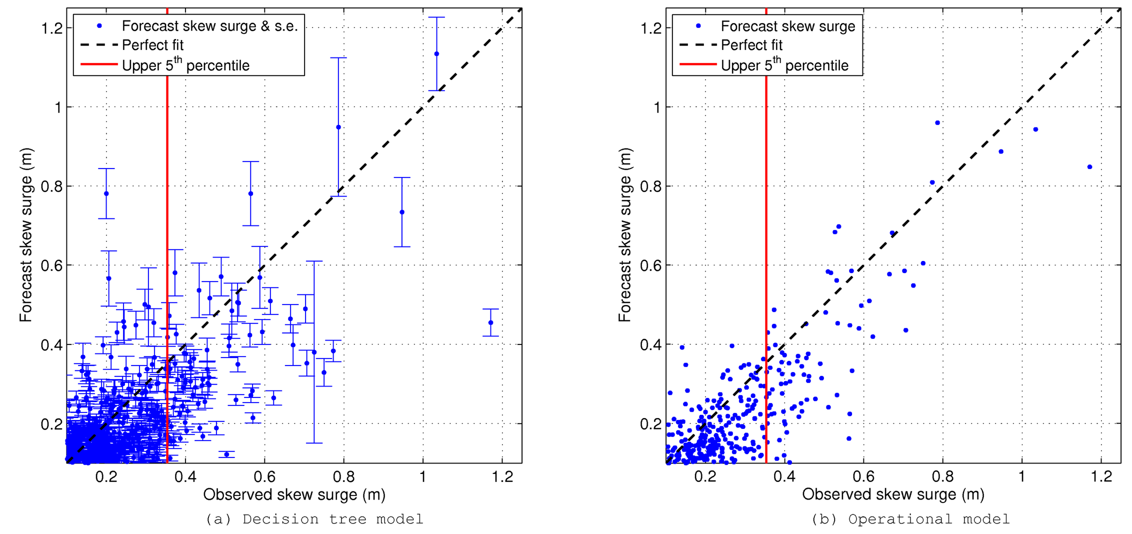

Figure 2: Scatter plot of forecast against observed skew surge from (a) the decision tree model, inc. standard error bars, and compared against (b) the operational forecast.

The decision tree structure can be schematised. For example, the tree identifies the most informative attribute to be skew surge at Whitby (the closest gauge to Sheerness), with additional information then obtained from the north-south wind speed data, the key driver in developing internal surge.

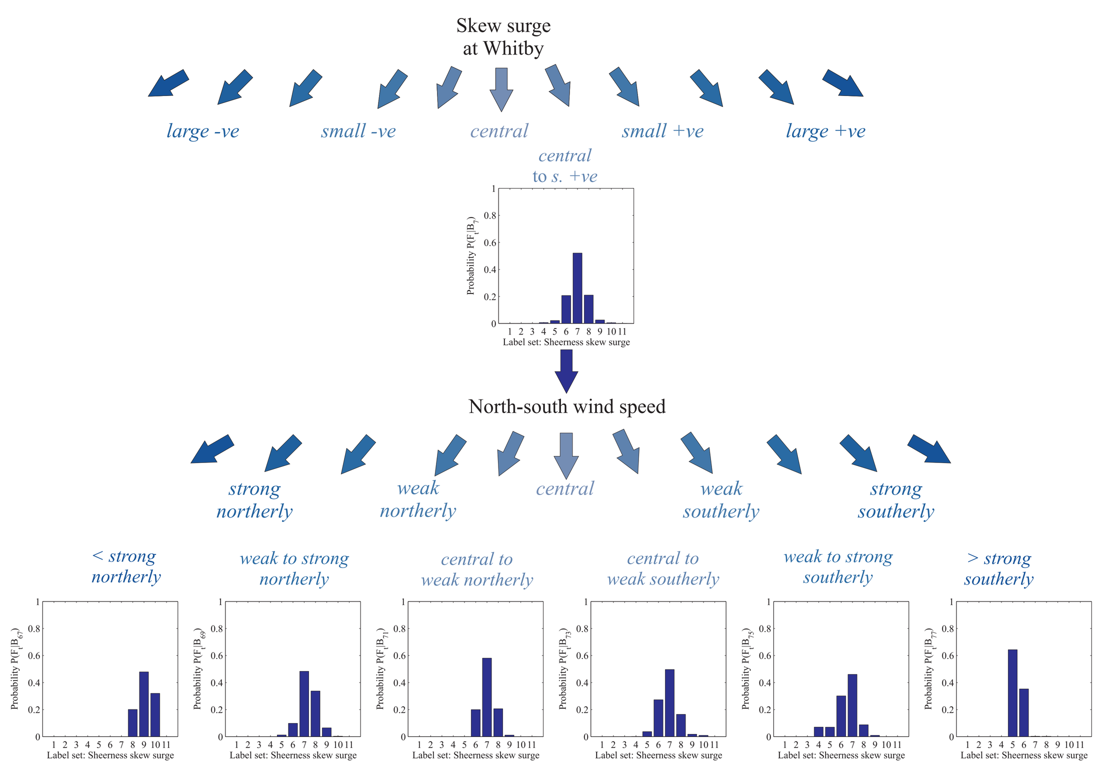

Figure 3: Decision tree structure for the central branches of the decision tree, consistent with the key mechanisms of storm surge generation in the North Sea.

3 Mean sea level variability

3.1 Local Scale Variability

To investigate the success of the model to long timescale problems, the decision tree model is applied to the problem of determining the forcing factors in local scale variability of annual mean sea level. The model hindcasts the rate of change in annual MSL at Brest from the rate of change in annual MSL at Newlyn; local atmospheric pressure and precipitation at Brest; and, large-scale atmospheric indices with high correlation with the Brest data series. Coincident data from 1953 to 2006 is used, and the annual MSL reconstructed from the integral of the rate of change.

The model exhibits comparable accuracy to alternative methods, with bias in the reconstructed MSL giving rise to the majority of the error.

Table 2: Predictive Accuracy

Method	Rate of change		MSL		
	RMSE (mm/yr)	r^2	MAE (mm)	RMSE (mm)	r^2
LID3 decision tree	10.1	0.91	-11.8	23.8	0.80
LLS regression	13.7	0.83	+7.8	18.1	0.84

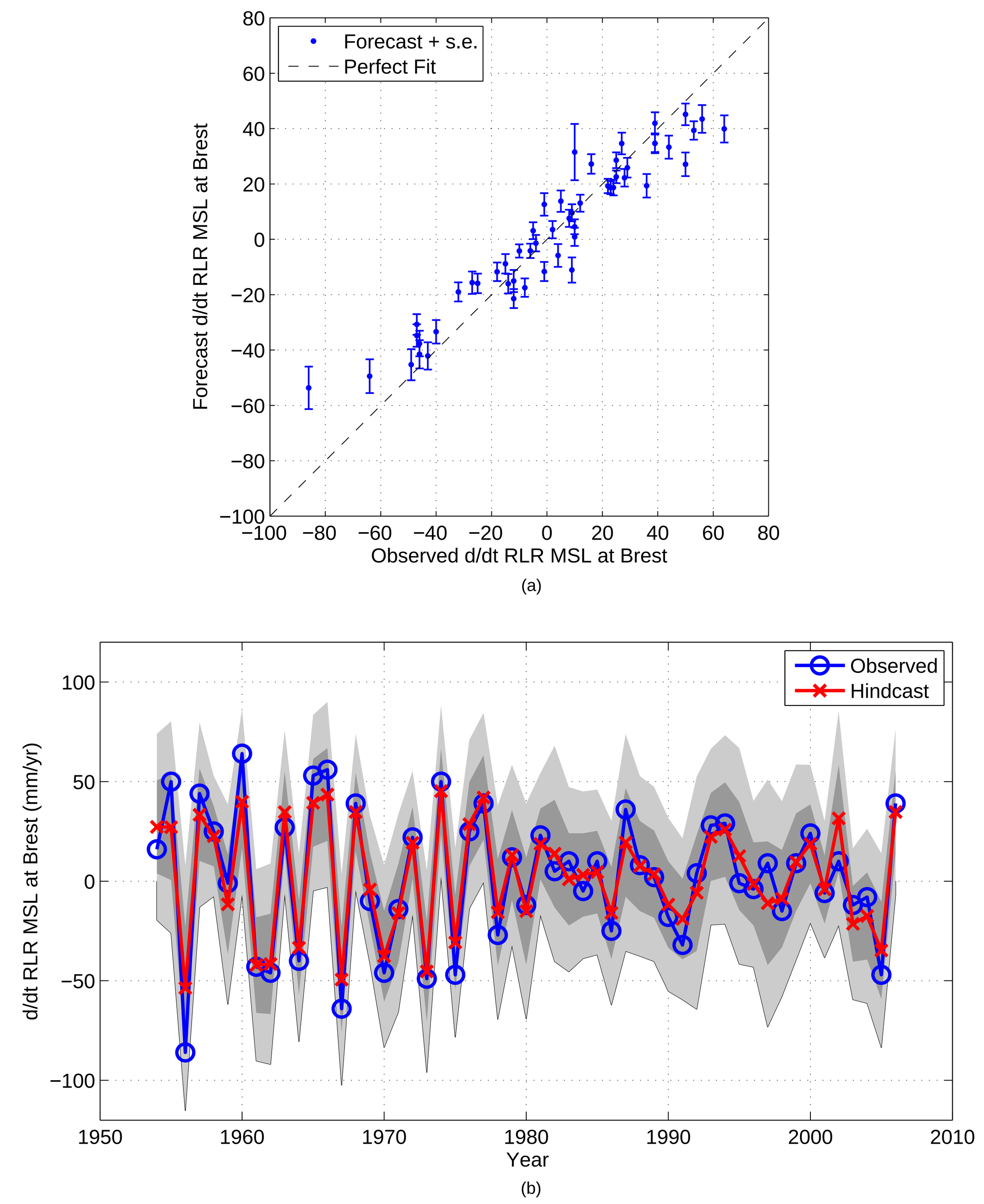

Figure 4: Hindcast rate of change in the Brest annual mean sea level and error estimates made by the decision tree model, (a) as a scatter plot, and (b) as a timeseries showing 1σ and 2σ errors.

The tree structure identifies the Scandinavia pattern, North Atlantic Oscillation and East Atlantic / West Russia pattern as most informative of the atmospheric indices, given the rate of change in annual MSL at Newlyn and local sea level pressure at Brest.

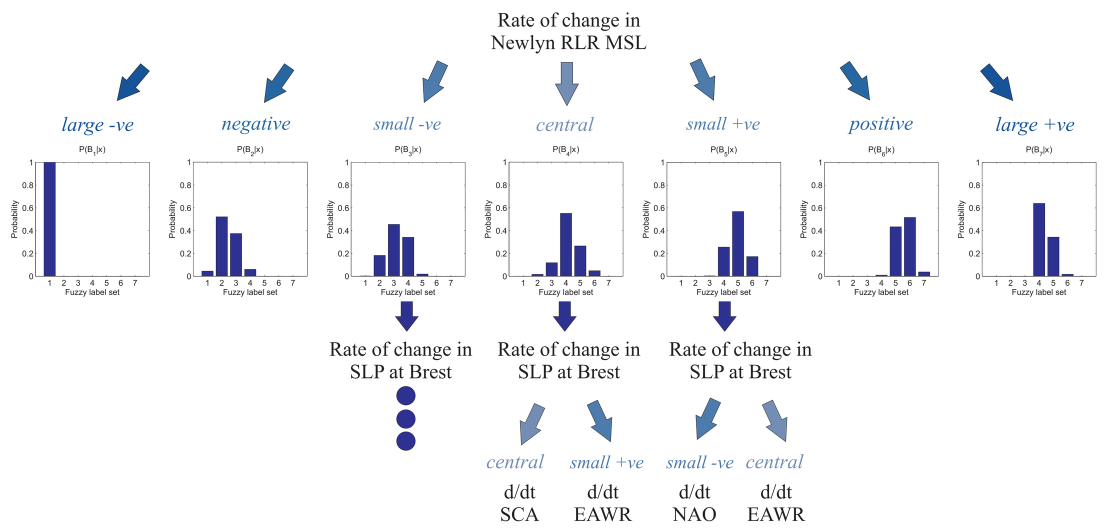

Figure 5: Schematic of the decision tree structure for the rate of change in annual MSL at Brest (significant to 90th %ile).

3.2 Regional scale variability

Going forward, the applicability of the rule-based method presented here will be assessed in a study of the forcing factors in regional sea level variability, by applying the model to the hindcast of north-east Atlantic MSL variability, for example from satellite altimetry.

Initially, annual MSL records have been investigated to identify whether a timeseries exists that is of sufficient length to provide useful information from the decision tree model. Tide gauge and atmospheric index data will be averaged where similarities meet appropriate criteria; for example using correlation coefficients, spatial proximity[3] and hierarchical clustering[4,5] methods to identify optimal merging of data, as exemplified in Figure 6.

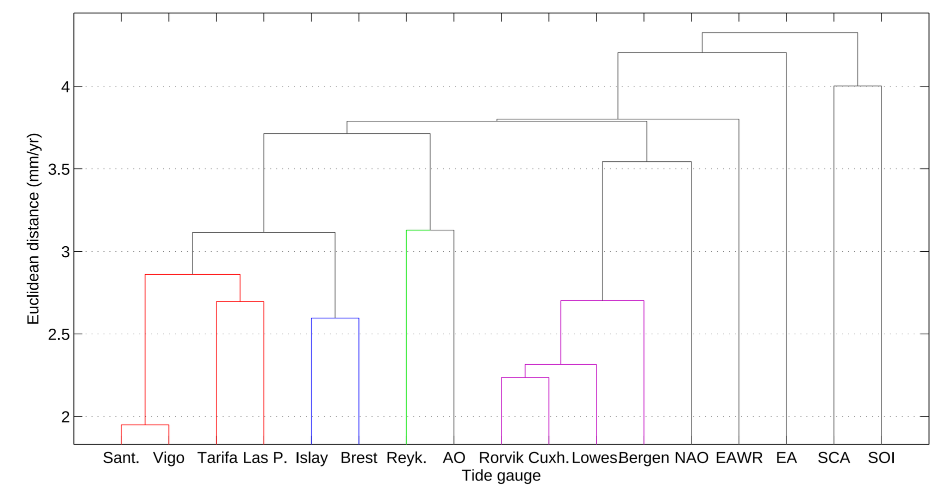

Figure 6: Merging of tide gauge records may be determined by hierarchical clustering, for example, by this dendrogram.

We hope to identify key relationships between the coastal tide gauge data, large-scale atmospheric motions and the regional MSL in this statistically robust method. In addition, it is hoped that natural error estimates can be made of the hindcasts, to identify natural bounds of the regional MSL variability.

4 Conclusions

The data-driven model has comparative accuracy to alternative methods in hindcasting both short-term (storm surge) and long-period (annual mean) sea level variability. The model provides probabilistic forecasts, which subsequently provide natural error bounds on the predictions. This is particularly useful for flood warning purposes but may aid our understanding of natural sea level variability on longer timescales.

In addition, the model tree structure can be interrogated to ensure that the model is consistent with our understanding of the physical system. This interpretation of the IF-THEN rules may prove useful in determining the key drivers of sea level variability in regions around the world.

References

[1] Z. Qin and J. Lawry (2005) *Information Sciences* **172**, pp.91–129.

[2] UK Coastal Monitoring and Forecasting service. UK Meteorological Office. *http://www.metoffice.gov.uk/publicsector/environmental*.

[3] S. Jevrejeva *et al.* (2004) *JGR*, **111**, C09012.

[4] S. H. You and J.-W. Seo (2009) *Natural Hazards*, **51**, pp. 97–114.

[5] M. G. Scotto *et al.* (2009) *Applied Ocean Res.*, **31**, pp. 4–11.

Acknowledgements

National Tide Gauge Network data was provided by BODC (www.bodc.ac.uk) and meteorological data was taken from NOC's archive of the UK's operational storm surge forecast model. Annual MSL data is provided by the PSMSL (www.psmsl.ac.uk), satellite altimetry by CSIRO (www.cmar.csiro.au/sealevel) and atmospheric indices are available from UEA's Climate Research Unit (www.cru.uea.ac.uk) and NOAA (www.cpc.ncep.noaa.gov). S. Royston's PhD was funded by the UK EPSRC's Flood Risk Management Research Consortium.