2.4

# PREDICTION OF SKEW SURGE BY A FUZZY DECISION TREE

Samantha J. Royston [a,b1], Kevin J. Horsburgh [a], Jonathan Lawry [b]
[a] Proudman Oceanographic Laboratory, Liverpool, UK
[b] University of Bristol, Bristol, UK

## 1 INTRODUCTION

Storm surge resulting from mid-latitude weather systems have the potential to cause considerable damage and fatalities as a result of coastal flooding (Lavery and Donovan, 2005; Lescrauwaet et al., 2006; Jonkman and Vrijling, 2008). The real-time, accurate prediction of storm surge is of primary importance in flood forecasting and warning systems. Whilst existing deterministic, hydrodynamic forecast models are highly successful at predicting storm surge (for example, Horsburgh et al. (2008)) the development of new models and further improvement to existing models is expensive. Therefore, opportunities exist for alternative, complementary approaches to improve upon predictive accuracy and our understanding of the system dynamics.

Artificial intelligence (AI) and probabilistic techniques have been applied to the real-time or hindcast prediction of water level including meteorological effects with some success. For example, artificial neural network (ANNs) models have been applied to this problem in the Gulf of Mexico (Tissot et al., 2001; Cox et al., 2002; Tissot and Zimmer, 2007), Western Australia (Makarynskyy et al., 2004), Taiwan (Lee, 2008b,a), the Bohai Sea (Liang et al., 2008) and the North Sea (Ultsch and Röske, 2002; Prouty et al., 2005). A real-time forecast system now successfully operates in the Gulf of Mexico (Tissot et al., 2008). Alternative approaches have included Support Vector Machines (SVMs) applied to the Adriatic Sea (Canestrelli et al., 2007) and Taiwan (Rajasekaran et al., 2008); Genetic Programming (GP) applied to the Gulf of Mexico (Charhate and Deo, 2008); and, chaos and fuzzy Naive-Bayes models applied to the North Sea (Siek et al. (2008) and Randon et al. (2008) respectively).

For mid-latitude, coastally-trapped storm surges in shelf seas with moderate tidal ranges, such as those that occur in the North Sea, Adriatic Sea and Sea of Azov, it has been shown that surges tend to cluster on the rising limb of the tide due to a small phase shift in turn driven by the change in water depth due to the meterologically-driven surge (Horsburgh and Wilson, 2007). Storm surge predictions by data-driven techniques, such as those referred to, tend to use the residual water level record, which displays this clustering and may include a leaked tidal component in the signal resulting from non-linear interactions. It is therefore proposed that instead an alternative metric, the skew surge record, be used to train and test a storm surge predictor. Skew surge, is defined as the difference in elevation between the maximum observed water level in a tidal cycle and that predicted by tide tables and is considered to be the most appropriate measure of storm surge for flood warning purposes (since the peak total water level is simply reconstructed from the skew surge plus the high water level for that tidal cycle). The components of total water level are schematised in Figure 1.

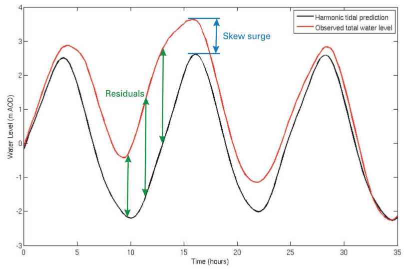



Figure 1: A schematic of total water level and it's constituents.

Furthermore, there has only been very limited use of data-driven models to provide insight into the system dynamics (for example, Siek and Solomatine (2007); Liu et al. (2007); Lee et al. (2008)). This is primarily due to the choices of technique, which were likely to have been chosen for their ease of implementation and computational efficiency given the weight of evidence for good predictive accuracy. An alternative, probabilistic data-driven technique is proposed - specifically, a fuzzy, linguistic, entropy-based decision tree - which offers a transparent structure in combination with proven predictive accuracy for similar environmental problems (McCulloch et al., 2007, 2008).

The fuzzy decision tree algorithm is applied to the problem of predicting skew surge at the entrance to the Thames Estuary, UK, given harmonic tidal predictions, skew surge from remote tide gauges around the North Sea and meteorological reanalysis and forecast data.

## 2 DATA AND METHOD

### 2.1 Fuzzy Decision Tree Method

The continuous data sets are first transformed into fuzzy sets. The method can therefore also use cat-

[1] *Corresponding author address: Proudman Oceanographic Laboratory, Joseph Proudman Building, 6 Brownlow Street, Liverpool, L3 5DA, UK; email: samyst@pol.ac.uk*

egorical or descriptive data sets. An entropy-based probabilistic decision tree then splits the training record to maximise homogeneity in the target fuzzy sets (the skew surge at Sheerness), with each node subsequently assigned a probability distribution for the target fuzzy sets. The predicted probability distributions can subsequently be 'defuzzified' to give real-valued predictions, given test input data.

### 2.1.1 Fuzzy Discretisation

In order to construct a model allowing for inaccuracies and noise in the observational data, the input and output data sets are discretized into trapezoidal fuzzy sets. In this problem, it is useful to think of the discretisation or partitioning in terms of labels, such as $\{$small$\}$, $\{$medium$\}$ and $\{$large$\}$, and to consider the fact that in the real world different voters might consider it appropriate to label any given value by different labels (that is, one voter may consider it appropriate to label a value $\{$small$\}$ whilst another might label it $\{$medium$\}$). By applying a full fuzzy covering with 50% overlap, appropriate label sets (combinations of labels) can be defined by probability distributions or mass assignments on the fuzzy sets, $m_{x_r}$ or $m_y$, for the inputs and target respectively. The label sets are then both exclusive and exhaustive and are derived only from adjacent labels (i.e. for the example given, appropriate labels sets are $\{$small$\}$, $\{$small or medium$\}$, $\{$medium$\}$, $\{$medium or large$\}$ and $\{$large$\}$ only; the set $\{$small or large$\}$ is not appropriate). This process is presented in Figure A for a model example.

The partitions can be chosen by applying a statistical or percentile distribution to the data such as a uniform distribution applied in this case; or the partitions can be assigned to minimise entropy. Entropy, $E$, is a measure of the homogeneity or 'purity' of the output partition interval or 'bin' given the input data; it is a measure of the spread of class or partition boundaries across each database and is derived from the probability of a particular partition across a database:

$$E = -\sum_{t=1,\ldots,m} P(F_t) log_m P(F_t) \tag{1}$$

where

$$P(F_t) = \frac{\sum_{i=1}^{N} m_y(i)}{\sum_{i=1}^{N}\sum_{t=1}^{m} m_y(i)} \tag{2}$$

and $F_t$ are the label sets of the database for each target fuzzy set, $t = 1, \ldots, m$, and $i = 1, \ldots, N$ are the data vectors at each timestep. The entropy function is illustrated for a binary case (for example, where the output classification can be true or false) in Figure 2 which shows that entropy is maximised if the data is uniformly spread with a probability of 0.5 for each output class and minimised if the data set tends lies purely in one class or the other. Minimising the entropy maximises the information gained from a given partition, since a low entropy suggests the data set of interest tends towards being homogenous in the output class.

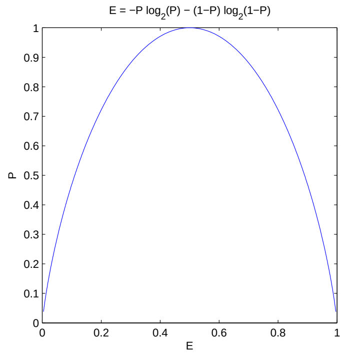



Figure 2: Entropy, $E$, of a binary classification system against the probability of each output class, $P(F_t)$.

The data can be discretised by entropy by predefining the desired number of fuzzy sets (Qin and Lawry, 2005) or by applying the Minimum Description Length Principle (Quinlan and Rivest, 1989; Fayyad and Irani, 1992).

The predictor is highly sensitive to the choice of fuzzy discretisation (method and number of fuzzy sets) (Qin, 2005) and so was thoroughly tested with uniform, normal and percentile distribution and entropy-based discretisation methods for a range of number of fuzzy sets. It has been found that a uniform discretisation is most successful for predicting events throughout the domain. Additionally, in this case where we are interested in extreme values, additional label sets have been defined at the end of each universe to improve predictive accuracy here (Randon, 2004). This can be thought of as including linguistic label sets of $\{$very small$\}$ and $\{$very large$\}$ at the extrema.

### 2.1.2 Probabilistic Decision Tree

The probabilistic decision tree, hereafter denoted DT, is based on the ID3 algorithm, which develops the tree structure using an entropy-based approach whereby the input variable used at each depth of the tree is chosen to maximise information gain (by minising entropy) and thus the tree should tend towards the smallest possible tree to a given threshold for the data sets presented to it. At each depth, the expected entropy of developing a branch with each of the remaining available input data sets is calculated and each branch is extended using the data that provides most information. In the first instance, the entropy for the entire database can be calculated from equation 1. The entropy for subsequent branches, $E(B)$, is derived from equations 1 and 2 by summing the mass assignments over the subsets of data at each branch node, $i \in B$, rather than the entire database. For the input variables, $x_{r=1,\ldots k}$, each fuzzy label set (such as {small or medium}) is denoted $F_j$ where $j = 1, \ldots, n$ and for the target variable, $y$, each fuzzy label set is denoted $F_t$ where $t = 1, \ldots, m$. The probability of an output classification, $F_t$, given a branch, $B$, is denoted $P(F_t|B)$. Similarly, the entropy of a branch extended by an unused input variable, $x_r$, is determined from the mass assignments summed over the new subsets of the data at each new node and is denoted $E(B \cup F_{jr})$. The subsequent information gain, $IG$, from extending a branch, by each fuzzy set $j$ of each input variable $x_r$ is given by:

$$IG(B, x_{rj}) = E(B) - EE(B, x_{rj}) \tag{3}$$

where the expected entropy from extending a branch, $EE$, is defined as:

$$EE(B, x_{rj}) = \sum_{F_j} E(B \cup F_{rj}) P(F_{rj}|B) \tag{4}$$

and the probability distribution within a node can be given by the standard frequentist view:

$$P(F_t|B) = \frac{\sum_{i \in DB} m_{x_r}(i)(F_j) m_y(i)(F_t)}{\sum_{i \in DB} m_{x_r}(i)(F_j)} \tag{5}$$

or extended using Laplace's law of succession with a Dirichlet distribution as a prior:

$$P(F_t|B) = \frac{\sum_{i \in DB} m_{x_r}(i)(F_j) m_y(i)(F_t) + 1}{\sum_{i \in DB} m_{x_r}(i)(F_j) + m} \tag{6}$$

As the DT expands, the branch may be terminated at a leaf node if there are no examples in the 'training' data set of a combination of fuzzy sets in a given branch and so a null set exists at the node. In this case, $P(F_t|B)$ cannot be estimated from the 'training' data and is therefore assigned an equal probability in each fuzzy set:

$$P(F_t|B) = \frac{1}{m} \tag{7}$$

In addition, the user may set termination criteria which creates leaf nodes based on confidence or significance testing or on homogeneity criteria in the target fuzzy sets.

#### 2.1.3 Prediction and 'Defuzzification'

Given an input vector, the DT model will predict a set of probabilities of the output variable falling within each of the nodes. The probability of following a branch given any specific input data vector is equal to the joint probability (or product) of the mass assignments along that branch (defined from the prior probabilities of the 'training' dataset):

$$P(B|\mathbf{x}) = \prod_{r=1}^{k} m_{x_{jr}}(F_{jr}) \tag{8}$$

The probability of obtaining each fuzzy set in the output variable, $F_t$, is then determined using Jeffery's rule:

$$P(F_t|\mathbf{x}) = \sum_{v} P(F_t|B_v) P(B_v|\mathbf{x}) \tag{9}$$

where $v$ branches exist in the tree structure.

To then 'defuzzify' the predicted probability to a real-valued prediction of the output variable, given a specific input, the estimate or expected value is given by:

$$\hat{y_t} = \sum_{F_t} a_i P(F_t|\mathbf{x}) \tag{10}$$

where $a_i$ can be determined from some distribution on the target fuzzy sets (Randon, 2004). Appropriate defuzzification methods include the use of the mode of each fuzzy set and the use of the expected value, given by:

$$a_i = \frac{\int_{\Omega_t} y_t m_{y_t}(F_t) dy_t}{\int_{\Omega_t} m_{y_t}(F_t) dy_t} \tag{11}$$

### *2.2 Site of Interest and Data*

Coastally-trapped tides and surges, resulting from disturbances in the North Atlantic Ocean and/or meteorological forcing within the North Sea basin, propagate cyclonically around the North Sea, following the UK's east coast from north to south. Therefore, observed sea levels from the north-east coast of the UK can inform of storm surges progressing towards the Thames Estuary (Darbyshire and Darbyshire, 1956; Rossiter, 1959). The UK Tide Gauge Network gauge at Sheerness, the whereabouts of which is given in Figure 3, is of particular importance because predicted extreme sea levels here are used to determine whether or not to close the Thames Barrier which protects London from flooding. Hence this gauge is chosen as the target for the prediction model. UK Tide Gauge Network data was provided by the British Oceanographic Data Centre (BODC, 2008).

Yearly data is provided in ASCII fixed-format files giving the date and time of observations, total observed water level and the residual water level (being the observed total water level minus the astronomical tidal prediction which can therefore be back-calculated). The residual water level includes storm surge from meteorological forcing, mean sea level variations or drift resulting from effects such as density variations, phase shift in the tidal predictions due to non-linear effects, particularly in shallow water, and noise from shorter period forcing. The skew surge is subsequently calculated for each tidal cycle, reducing the number of data records and removing some of these unwanted effects. Data are flagged by BODC to indicate where values are null or missing, improbable, or interpolated. The null and improbable values are set to null values in the original data sets, resulting in corresponding null values in the skew surge record.

The correlation between the surge signal at Sheerness and remote gauges along the UK's east coast improves with proximity to Sheerness, as expected. However, in order to provide a sufficient lead-time to allow for the closure of the Thames Barrier, the most southerly tide gauge that is appropriate as an input to the predictive model is Whitby, with a lead time of approximately 8 hours (Randon et al., 2008). Potential model inputs consist of the observed skew surge, harmonic prediction for high water and timing between the observed and harmonic prediction of high water, from five remote tide gauges from Lerwick to Whitby, also shown in Figure 3, located along the UK east coast with wave travel times to Sheerness in the approximate range of 8 to 14 hours. Additionally, barometric pressure and easterly and northerly wind speed components for an area over the southern North Sea are derived from archive data from the UK's operational storm surge model operated by the UK Meteorological Office's Storm Tide Forecasting Service (STFS, 2009).


Figure 3: The North Sea region, highlighting UK tide gauges used in the analysis.

Available data was taken from 1980-2008 for the tide gauge locations, and from 1999-2008 for the meteorological data. The available data for each period of data was collected to create two sets of model inputs:

1. Skew surge and timing (between the observed peak water level and harmonic prediction for high water) at Lerwick, Wick, Aberdeen, North Shields and Whitby and tidal high water at Sheerness.
2. The water level data as above, and additionally barometric pressure at Sheerness and northerly and easterly components of wind speed averaged over the southern North Sea for that tidal cycle.

The first model learns from the first 80% of the available data set and is tested on the unseen latter 20%, as a long record exists, whereas for the second model (which has less data), a 50% split is used for training and testing. Where data is missing, the record is neglected so a fair comparison can be made with a linear least squares model. However, it is noted that the method can be used with missing data, which is assigned a uniform prior probability for each fuzzy set.

## 3 RESULTS

The real-valued predictions for the unseen test data from the fuzzy DT model are compared against a linear least squares regression (LS) model for each of the two input structures (without and including meteorological data). The root mean squared error (RMSE), correlation, $r$, and coefficient of determination, $r^2$, are determined for the two models. A comparison can be also be made against a reference forecast of persistence, which assumes the same observed skew surge from the previous tidal cycle persists at Sheerness. The mean squared error skill score (MSE-SS) compares the mean squared error of the model to that of a persistence forecast and is given by:

$$\text{MSE-SS} = 1 - \frac{\text{MSE}_{\text{model}}}{\text{MSE}_{\text{persistence}}}$$

Hence, the MSE-SS is a measure of the reduction of variance by the model compared with the reference forecast.

Figures 4 and 5 present comparative scatter plots of the predicted skew surge at Sheerness against that observed for the test data sets, for the DT and LS models for the two input structures respectively. Table 1 presents the error and variance statistics of the model predictions. It is noted that the error of the UK's operational storm surge forecats model are of the order of 0.1m (STFS, 2009).

Table 1: Comparison of fuzzy decision tree (DT) model predictions with linear least squares regression (LS)

| Model | Without Meteorological Data | |
|---|---|---|
| Predictor | Fuzzy DT | Linear LS |
| RMSE (m) | 0.121 | 0.121 |
| MSE-SS | 0.63 | 0.64 |
| $r$ | 0.71 | 0.73 |
| $r^2$ | 0.50 | 0.53 |

| Model | Including Meteorological Data | |
|---|---|---|
| Predictor | Fuzzy DT | Linear LS |
| RMSE (m) | 0.108 | 0.102 |
| MSE-SS | 0.69 | 0.72 |
| $r$ | 0.79 | 0.82 |
| $r^2$ | 0.62 | 0.67 |

It can be seen that the real-valued predictions from the DT model are comparable in accuracy to predictions from a linear LS regression model, although the linear LS regression prediction appears to have less scatter than the DT prediction. This is most notable in the second extended model where less training data is available for the model structures. The DT model is considerably more skillful at predicting a large positive skew surge at Sheerness than a persistence forecast or random chance. The decision tree predictions at the extremes are biased towards the more central values of skew surge due to the natural leptokurtic distribution of the data and the transform of real-valued data into fuzzy sets and back. However, for largest positive skew surge events at Sheerness in the test data set for the second model, including meteorological data, the DT model accurately predicts the skew surge where the linear LS regression model fails.

Given the comparable accuracy to other data-driven prediction methods, a significant benefit of the DT method is that the tree structure can be interrogated as sets of rules, allowing insight into the key physical drivers of surges at this critical location.

Figure 6 schematises the tree branches where the DT model predicts strong probabilities of a large positive skew surge at Sheerness, for the first model with only water level input data. These branches can be interpreted as rules which are given below for both the first water level model and the second, extended model with meteorological data respectively.

Model 1: Without meteorological data

Rule 1:
IF Whitby skew surge IS large positive
AND Aberdeen skew surge IS medium positive
AND ( Wick skew surge IS small positive )
OR ( Wick skew surge IS medium positive
AND skew surge timing at Aberdeen IS slightly advanced )
THEN there is a strong probability (P $>$0.95) that Sheerness skew surge will be large positive.

Rule 2:
IF Whitby skew surge IS small positive
AND tidal range IS neap or mid-range
AND skew surge timing at Wick IS slightly delayed
AND skew surge timing at Lerwick IS central
THEN there is a strong probability (P $>$0.95) that Sheerness skew surge will be large positive.

Rule 3:
IF Whitby skew surge IS medium positive
AND Lerwick skew surge IS small positive
AND skew surge timing at Wick IS slightly delayed
AND skew surge timing at Lerwick IS slightly advanced
THEN there is a strong probability (P $>$0.95) that Sheerness skew surge will be medium or large positive.

Rule 4:
IF Whitby skew surge IS medium positive
AND Lerwick skew surge IS central
AND North Shields skew surge IS central
AND skew surge timing at Whitby IS significantly advanced or slightly delayed
THEN there is a strong probability (P $>$0.95) that Sheerness skew surge will be medium or large positive.

Rule 5:
IF Whitby skew surge IS medium positive
AND Lerwick skew surge IS small negative
AND North Shields skew surge IS medium positive
AND tidal range IS between mid-range and neap tide
THEN there is a strong probability (P $>$0.95) that Sheerness skew surge will be medium or large positive.

Model 2: With meteorological data

Rule 1:
IF Whitby skew surge IS small positive
AND north-south wind component IS strong northerly
AND timing of skew surge at North Shields IS central or slightly delayed
THEN there is a strong probability (P = 0.87) that Sheerness skew surge will be medium or large positive.

Rule 2:
IF Whitby skew surge IS small positive
AND north-south wind component IS northerly
AND timing of skew surge at North Shields IS delayed
THEN there is a moderate probability (P = 0.62) that Sheerness skew surge will be medium or large positive.

Rule 3:
IF Whitby skew surge IS central
AND north-south wind component IS strong northerly
AND atmospheric pressure at Sheerness IS mid-range
THEN there is a moderate probability (P = 0.50) that Sheerness skew surge will be medium or large positive.

The two rules leading to the highest probability of a large positive skew surge at Sheerness (labelled Rules 1 and 2 for Model 1) are given by an observed skew surge amplifying as it progresses from north to south. The rules generally identify a phase shift from delays in the skew surge at the more northerly gauges to an advance in the time of expected peak water level at the more southerly gauges, which itself can be explained by an increase in wave celerity with increased depth caused by the surge within this shallow shelf sea. The propagation speed ($c$) of all shallow water waves is given by $c = (gh)^{1/2}$ where $g$ is Earth's gravitational constant and $h$ is water depth. The entropy-based algorithm also identifies that information can be gained from the state of the tide, with large skew surges at Sheerness more likely to occur during smaller (neap) tidal ranges. This is consistent with earlier examination of the data for this site (Horsburgh and Wilson, 2007).

The second extended model, including meteorological data, is dominated by the inputs with largest correlation with the skew surge at Sheerness; the skew surge at Whitby and the north-south wind speed component.

## 4 CONCLUSIONS AND PROSPECTS

The accuracy of the method is comparable to alternative prediction methods whilst offering the major benefit of transparency. Potential improvements to predictive accuracy at the extremes may be gained from more sophisticated data sampling into training and test sets (such as bagging or clustering), prior and post transformation of the data to reduce its leptokursis (for example by a sigmoid function) and more sophisticated approaches to the fuzzy set discretisation.

The tree predictions and interpretation are encouraging for further development and application of this technique to this problem, and other tidal and sea level applications. The probabilistic approach avoids optimization procedures, and it also allows null values in some or all of the input variables, where measured data may sometimes be missing for operational reasons. The resulting rules from the decision tree structure can be interpreted and preliminary results are consistent with our understanding of the physical system. The method is fast to implement on standard PCs and could therefore be utilized in a real-time application.

## ACKNOWLEDGEMENTS

This work is supported by the Flood Risk Management Research Consortium, who are funded by the Engineering and Physical Sciences Research Council and UK and Rebublic of Ireland Government Departments and agencies. The tidal data were supplied by the British Oceanographic Data Centre as part of the function of the National Tidal & Sea Level facility, hosted by the Proudman Oceanographic Laboratory and funded by the Environment Agency and the Natural Environment Research Council.

## References

BODC, 2008: British Oceanographic Data Centre UK Tide Gauge Network data. Available online at `www.bodc.ac.uk`.

Canestrelli, E., P. Canestrelli, M. Corazza, M. Filippone, S. Giove, and F. Masulli, 2007: Local learning of tide level time series using a fuzzy approach. *Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN)*, 1813–1818, iSBN:978-1-4244-1380-5.

Charhate, S. B. and M. C. Deo, 2008: Storm surge prediction using NN and GP. *Sixth Conference on Artificial Intelligence and its Applications to the Environmental Sciences*, American Meteorological Society.

Cox, D. T., P. Tissot, and P. Michaud, 2002: Water level observations and short-term predictions including meteorological events for entrance of Galveston Bay, Texas. *Journal of Waterway, Port, Coastal and Ocean Engineering*, **128 (1)**, 21–29.

Darbyshire, J. and M. Darbyshire, 1956: Storm surges in the North Sea during the winter 1953-4. *Proceedings of the Royal Society of London Series A*, **235 (1201)**, 260–274.

Fayyad, U. M. and K. B. Irani, 1992: On the handling of continuous-valued attributes in decision tree generation. *Machine Learning*, **8**, 87–102.

Horsburgh, K. J., J. A. Williams, J. Flowerdew, and K. Mylne, 2008: Aspects of operational forecast model skill during an extreme storm surge event. *Journal of Flood Risk Management*, **1 (4)**, 213–221.

Horsburgh, K. J. and C. Wilson, 2007: Tide-surge interaction and its role in the distribution of surge residuals in the North Sea. *Journal of Geophysical Research: Oceans*, **112**.

Jonkman, S. N. and V. K. Vrijling, 2008: Loss of life due to floods. *Journal of Flood Risk Management*, **1 (1)**, 43–56.

Lavery, S. and B. Donovan, 2005: Flood risk management in the Thames Estuary looking ahead 100 years. *Phil. Trans. Royal Soc. Mathematical, Physical and Engineering Sciences*, **363 (1831)**, 1455–1474.

Lee, B.-C., C.-C. Huang, C. C. Kao, and L. Z.-H. Chuang, 2008: Study on the characteristics of storm surge over Taiwan eastern waters by wavelet transform. *Proceedings of the* $18^{th}$ *International Offshore and Polar Engineering Conference*, Vol. 3, 753–758.

Lee, T.-L., 2008a: Back-propagation neural network for the prediction of the short-term storm surge in Taichung harbor, Taiwan. *Engineering Applications of Artificial Intelligence*, **21 (1)**, 63–72.

Lee, T.-L., 2008b: Prediction of storm surge and surge deviation using a neural network. *Journal of Coastal Research*, **24 (sp3)**, 76–82.

Lescrauwaet, A. K., J. Mees, and C. R. Gilbert, 2006: State of the Coast of the Southern North Sea: An indicators-based approach to evaluating sustainable development in the coastal zone of the Southern North Sea. Tech. Rep. 36, Flanders Marine Insistute (VLIZ) Special Publication.

Liang, S.-X., M.-C. Li, and Z.-C. Sun, 2008: Prediction models for tidal level including strong meteorologic effects using a neural network. *Ocean Engineering*, **35 (7)**, 666–675.

Liu, X.-L., J.-X. Zhang, and C.-F. Huang, 2007: Recognition on the relativity between typhoon and storm surge using fuzzy interval information matrix. *Advances in Intelligent System Research: Proceedings of the $1^{st}$ International Conference on Risk Analysis and Crisis Response*, Vol. 2, 39–44.

Makarynskyy, O., D. Makarynska, M. Kuhn, and W. E. Featherstone, 2004: Predicting sea level variations with artificial neural networks at Hillarys Boat Harbour, Western Australia. *Estuarine, Coastal and Shelf Science*, **61**, 351–360.

McCulloch, D. R., J. Lawry, and I. D. Cluckie, 2008: Real-time flood forecasting using updateable linguistic decision trees. *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, IEEE, 1935–1942, iSBN:978-1-4244-1818-3.

McCulloch, D. R., J. Lawry, M. A. Rico-Ramirez, and I. D. Cluckie, 2007: Classification of weather radar images using linguistic decision trees with conditional labelling. *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 1–6, iSBN:1-4244-1210-2.

Prouty, D. B., P. E. Tissot, and A. A. Anwar, 2005: Using artificial neural networks for predicting storm surge propagation in North Sea and the Thames Estuary. *Fourth Conference on Artificial Intelligence Applications to Environmental Science*, American Meteorological Society.

Qin, Z., 2005: Learning with fuzzy labels: A random set approach towards transparent intelligent data mining techniques. Ph.D. thesis, University of Bristol, Faculty of Engineering.

Qin, Z. and J. Lawry, 2005: Decision tree learning with fuzzy labels. *Information Sciences*, **172**, 91–129.

Quinlan, J. R. and R. Rivest, 1989: Inferring decision trees using the minimum description length principle. *Information and Computation*, **80**, 227–248.

Rajasekaran, S., S. Gayathri, and T.-L. Lee, 2008: Support vector regression methodology for storm surge predictions. *Ocean Engineering*, **35 (16)**, 1578–1587.

Randon, N. J., 2004: Fuzzy and random set based induction algorithms. Ph.D. thesis, University of Bristol, Faculty of Engineering.

Randon, N. J., J. Lawry, K. J. Horsburgh, and I. D. Cluckie, 2008: Fuzzy Bayesian modelling of sea-level along the East Coast of Britain. *IEEE Transactions on Fuzzy Systems*, **16 (3)**, 725–738.

Rossiter, J. R., 1959: Research on methods of forecasting storm surges on the east and south coasts of Great Britain. *Q. J. Royal Met. Soc.*, **85 (365)**, 262–277.

Siek, M. B. and D. P. Solomatine, 2007: Recurrence plots in the analysis of extreme ocean storm surges. *2nd International Workshop on Recurrence Plots*, DELFT Cluster.

Siek, M. B., D. P. Solomatine, and S. Velickov, 2008: Multivariate chaotic models vs neural networks in predicting storm surge dynamics. *Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2112–2119, iSBN:978-1-4244-1820-6.

STFS, 2009: Storm Tide Forecasting Service operational storm surge model forecasts. Available online at `http://www.metoffice.gov.uk/publicsector/environmental.html`.

Tissot, P. E., D. T. Cox, and P. Michaud, 2001: Neural network forecasting of storm surges along the Gulf of Mexico. *Proceedings of the The Fourth International Symposium on Ocean Wave Measurement and Analysis (WAVES)*, ASCE / US Army Corps of Engineers.

Tissot, P. E., J. Davis, N. Durham, and W. G. Collins, 2008: Implementation of a neural network based surge prediction system for the Texas coast. *Sixth Conference on Artificial Intelligence and its Applications to the Environmental Sciences*, American Meteorological Society.

Tissot, P. E. and B. Zimmer, 2007: ANN training methods targeting performance during extreme events. *Fifth Conference on Artificial Intelligence and its Applications to the Environmental Sciences*, American Meteorological Society.

Ultsch, A. and F. Röske, 2002: Self-organising feature maps predicting sea level. *Information Sciences*, **144**, 91–125.

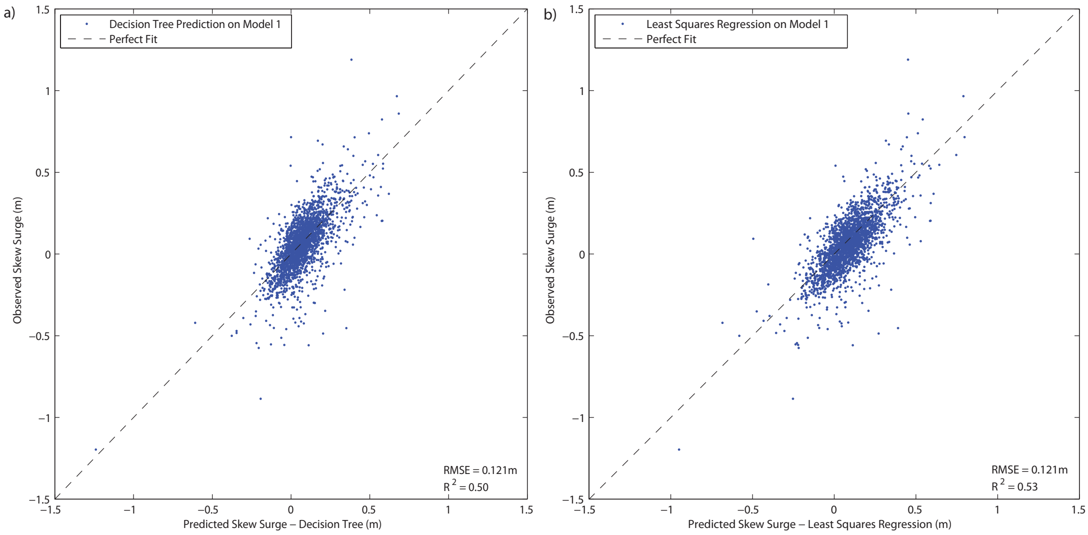



Figure 4: Scatter plots of predicted against observed skew surge at Sheerness for the first model structure without meteorological data, for a) the decision tree model and b) linear least squares regression.

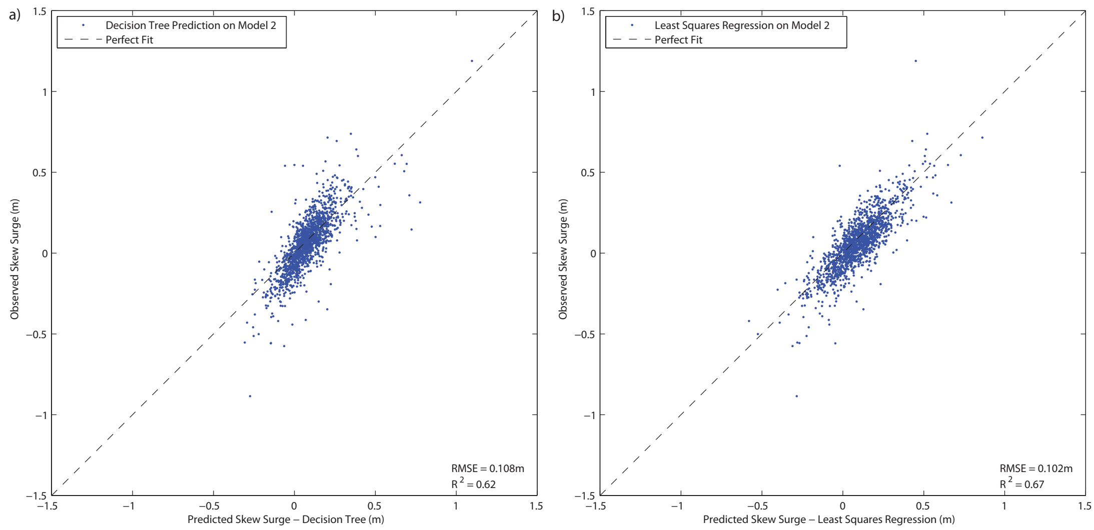



Figure 5: Scatter plots of predicted against observed skew surge at Sheerness for the second model structure including meteorological data, for a) the decision tree model and b) linear least squares regression.

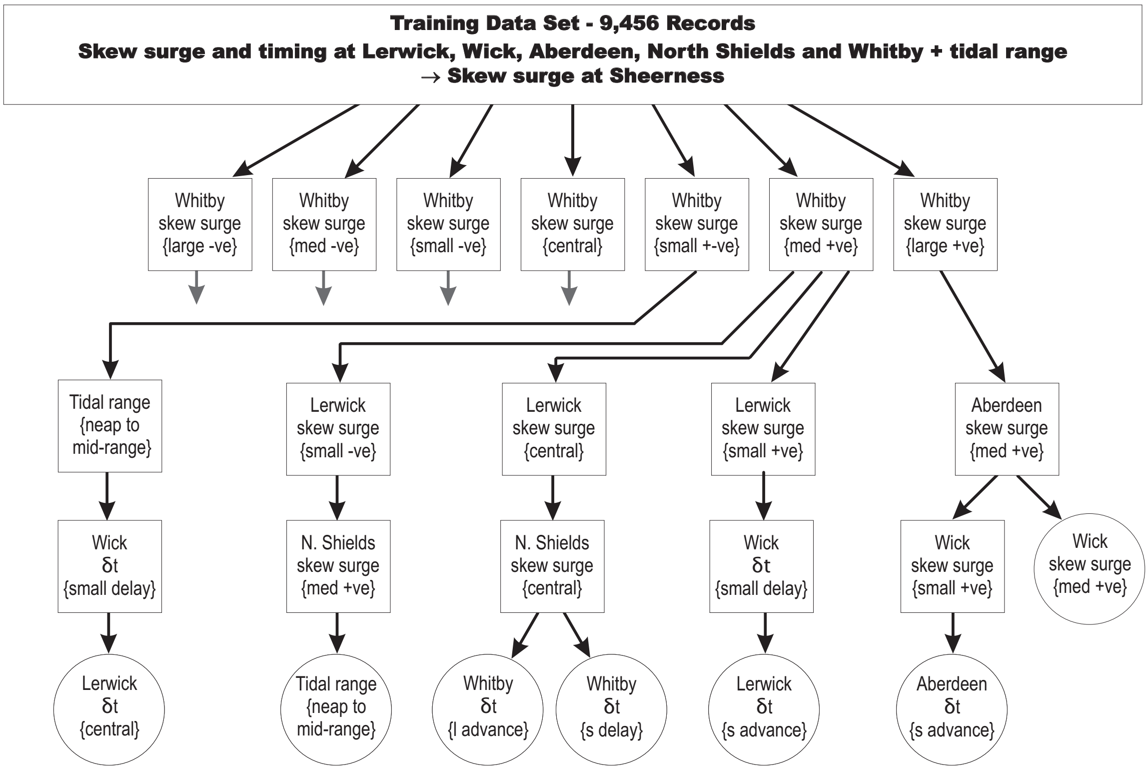
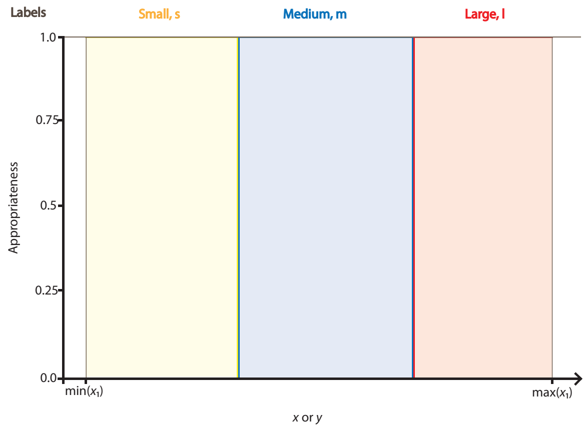
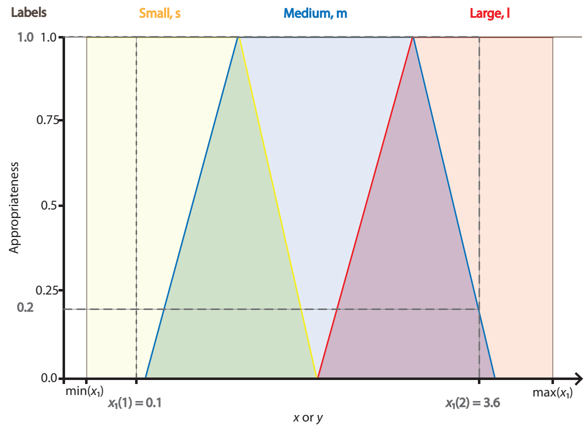


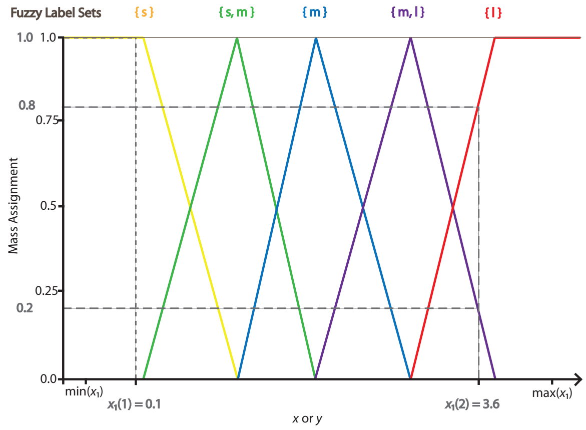

Figure 6: Schematic of the fuzzy decision tree model branch structures for large skew surge at Sheerness, for the first model structure without meteorological data.




(a) A model example of data partitioned into crisp bins.




(b) Labels with fuzzy edges corresponding to the crisp bins shown in Figure (a).




(c) Trapezoidal fuzzy sets applied to continuous data with label sets corresponding to Figure (b). Example mass assignments are given for the label sets.

$$
\begin{bmatrix} x_1 \\ \hline x_1(1) = 0.1 \\ x_1(2) = 3.6 \\ \vdots \\ x_1(N) = 1.6 \end{bmatrix}
\rightarrow
\begin{bmatrix}
\langle & \{s_1\} & \{s_1, m_1\} & \{m_1\} & \{m_1, l_1\} & \{l_1\} & \rangle \\ \hline
\langle & 1 & 0 & 0 & 0 & 0 & \rangle \\
\langle & 0 & 0 & 0 & 0.2 & 0.8 & \rangle \\
 & & & \vdots & & & \\
\langle & 0 & 0.4 & 0.6 & 0 & 0 & \rangle
\end{bmatrix}
$$

(d) An example of transforming continuous valued data into mass assignments on the fuzzy label sets, for an example data set.

Figure A: The process of fuzzy discretisation of a data vector for a model example.