EPSRC Maths Foresees Network Feasibility study report

Project title: Forecasting river levels utilising non-stationarity

Investigators (names and affiliations): Rebecca Killick (Mathematics and Statistics Department and Data Science Institute, Lancaster University), Ann Kretzschmar, Wlodek Tych, Suzi Ilic (Lancaster Environment Centre and Data Science Institute, Lancaster University)

Other participants: Ian Gold (Environment Agency)

Project background: Flooding poses a major environmental, economic, and social threat around the world. As a consequence of climate change, which will bring more variable and extreme weather conditions on the top of the sea level rise, the Environment Agency (EA) is expecting almost double investment in flooding protection schemes from £800M per year (e.g. in 2010-2011). The cost of the December 2015 flooding reached £5 billion and thousands of people were affected (Guardian, December 2015). Despite significant improvements in modelling and prediction of flood inundation levels and extent of flooding (Kretzschmar et al., 2014; Hoitink and Jay, 2016) there are still large discrepancies between predictions and observations in certain places and during storm events.

Whilst several models and forecast methods exist for predicting river levels (Smith et al., 2013; Leedal et al., 2013; McMillan and Westerberg, 2015), they typically assume that the data (or more importantly errors) are independent, stationary and Normally distributed. It is easily shown that the model errors depend on the river levels and in particular change drastically during storm events, when accurate predictions are most needed. The non-stationarity of the error structure demonstrates that the models to date are not describing the drivers of river level appropriately. Further challenges are in forecasting water levels in downstream stretches of rivers influenced by tides. Over the length of this part of a river, river flow will interact with tidal flow. The challenge is how a tidal wave, which transforms as it propagates up the river, interacts with the river flow and how these are included in models (Potter et al., 2016). While stationary methods such as harmonic analysis of tides have been used in the past, it is proposed that non-stationary methods should be used instead (Hoitink and Jay, 2016). The aim of this project is not to complicate the model further by adding further variables but to accept the non-stationarity and use recent advances in the area to model the error process as locally stationary. One advantage to our approach is that the single proposed model will be capable of predicting water levels in tidal and non-tidal influenced rivers. By providing a powerful statistical technique to adaptively forecast real time river levels, this project aims to improve robustness and reliability of flood prediction leading to better evidence and understanding of the underlying mechanisms, particularly related to tidal influence up-stream.

Project objectives: We propose to develop a flexible methodology able to forecast river levels at a particular location using data from both up- and down-river in addition to the standard meteorological information e.g.,forecast of rainfall. These data, provided by the Environment Agency, will be river levels taken from gauges along a river. We will extend the methodology developed by the PI (Killick et al., 2016), which forecasts non-stationary time series using a data-driven wavelet decomposition, to account for the added information contained in the up- and down-river data. The methodology will automatically choose which locations are informative at each time and forecast horizon. We will hold-out data for the most recent 18 months (which includes significant flooding) from the River Lune and use these as a basis for rolling testing in the wild. We expect to provide a different approach to Smith et al. (2013), and will compare it to the proposed methodology on the same synthetic series. We will consider confidence interval coverage (Verkade and Werner, 2011) as the main performance indicator as this is an important metric when using the tool in a real life scenario.

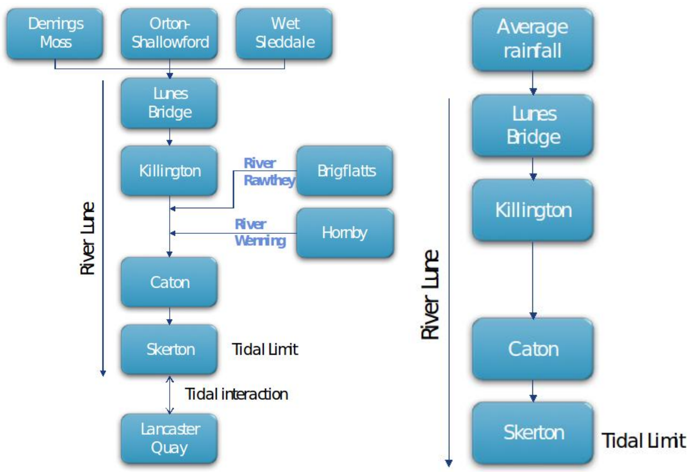
Key project outcomes: The key project outcomes are:

- A real-time model for the River Lune incorporating rainfall, river levels upstream and tidal data from down stream.
- A real-time hybrid model combining traditional hydrological models with locally stationary residuals.

Flooding is a worldwide problem but its severity was brought home to those of us living in the north-west UK on 5th/6th December 2015 when storm Desmond hit. There was severe flooding and in Lancaster the electricity sub-station was flooded leaving 61K homes without power for 3 days. Many rainfall and river flow records were broken. 16 months on, the damage is still visible. It is likely that the next time it happens, it will worse due to the effects of climate change.

Forecasting these extremes is essential and although modelling techniques have improved there are often still discrepancies between predictions and observations most significantly during storm events when these are most important. There is a general assumption that the errors are independent, stationary and normally distributed which is often not the case especially under storm conditions and even more so when tidal influences come into play as they did in Lancaster.

This project aimed to compare two methods for real-time forecasting a modification of the traditional Kalman Filter approach and a statistical technique called LPACF forecasting based on a data driven wavelet decomposition. A trial model was set up for the River Lune catchment that flows through Lancaster and caused flooding during storm Desmond. Figure 1(a) shows a schematic of the proposed model.



(a) (b)

Figure 1: Flow charts relating the different datasets. (a) Idealized. (b) Implemented.

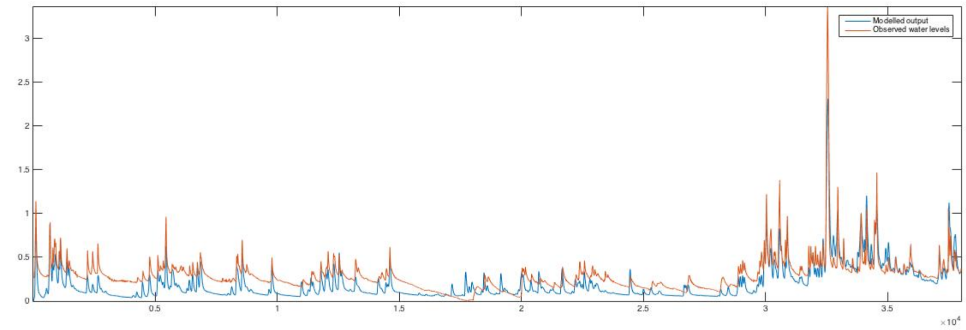
The Lune rises in the North Pennines and reaches the sea at Lancaster. It has 2 major tributaries that gather water from the Bowland Fells to the East. The Lune flows out to the sea through Lancaster with the tidal limit being Skerton Weir just upstream of the town. Rainfall, water level and tidal data for the catchment was made available by the Environment Agency. The methodology employed is similar to that used by Romanowicz et al. (2006) on the river Severn. It was proposed that the tidal influence at Skerton - which maybe over-topped at very high tides - be modelled using Dynamic Harmonic Regression as suggested by Smith et al. (2013). For the purpose of this feasibility study a simplified model was employed - shown in Figure 1(b).

A series of cascading transfer function models were set up linking rainfall in the upper catchment to water level forecasts at Skerton Weir the tidal limit and giving an absolute lead time of 12hrs, far in excess of the Environment agency requirements. A model has been fitted to the data for 2015 and the early part of 2016. As can be seen (Figure 2(a)), the extreme conditions in Winter 2015 have biased the model fit so that one size does not fit all. It would be possible to fit one model to Spring, one to Summer and one to Winter but in reality the conditions are changing continuously and with this the model parameters also.

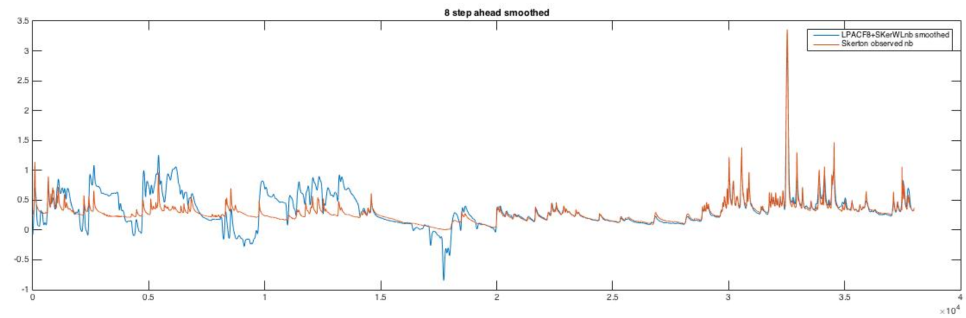
An alternative approach to the Normal distributed errors is to assume a locally stationary wavelet model error structure (Nason et al., 2000). A recent approach (Killick et al., 2016) introduces the local partial autocorrelation function as a tool for dynamically forecasating non-stationary time series. This creates the analogue of the stationary partial autocorrelation function and, akin to stationary time series, uses it to choose the AR order for forecasting purposes. This approach is brought to the non-stationary setting through the introduction of the local partial autocorrelation function which gives an indication of the local AR order used for forecasting.

We applied both methods to the data from the Environment Agency in order to predict the river levels at Skerton. The local partial autocorrelation forecasting was applied to the residuals of the original transfer function model to take account of the nonstationary residual structure. The forecasts were then a combination of the original (transfer) model forecast and the locally stationary forecast given in Figure 2(b).

What is clear from Figure 2 is that the transfer function model is biased to the larger river levels thus under estimating the smaller values earlier in the year. In contrast, the lpacf residual forecast sometimes over estimates and underestimates in the first half of the year. Arguably a large part of the performance of this forecast is due to the poor initial transfer function model estimates coupled with a “settling in” period that the wavelet approach



(a)



(b)

Figure 2: Actual and forecast measurements of river levels at Skerton for 2015 (a) Transfer Function model forecast. (b) lpacf forecast.

needs. This is due to the fact that we have run this in an online setting, thus at the start of the data the method only has a small number of data points from which to forecast. In reality the wavelet approach would be run for several months until past the "settling in" period before using the forecasts in anger. From around the middle of the data stream the lpacf forecast estimates appear to be much more stable and closer to the truth than the transfer function approach. We believe this to be a combination of stabilization of the lpacf approach and arguably more accurate transfer function model estimates. In particular, the estimates of the peak river levels of almost 3.5m during Storm Desmond are almost attained by the lpacf forecasts whereas the transfer function model forecasts the peak at below 2.5m although the difference could be due to tidal influence which we have not accounted for. Both forecasts reported are computed as 8-step ahead forecasts which is roughly 3.5 hours lead time.

An alternative competitor is the Kalman Filter. The Kalman filter is a powerful tool for predicting the state of a system based on the previous state. It is an iterative process that quickly homes in on the correct value of a process that contains random error or uncertainty as new values are added. It is efficient and can be used in real-time because the estimate of the current state is only dependent on the previous step not the whole history. Figure shows how the adaptive Kalman filter homes in on the correct value as more data is added and lead time gets shorter.

To compare the Kalman Filter forecast and the transfer function model with lpacf errors forecast we concentrate on the period around Storm Desmond. The results can be seen in 4. Due to the fact that the Kalman Filter relies on previous observations for predictions it estimates the peak river levels too late, although accurately estimates the timing of the majority of the smaller peaks in the data. In contrast the lpacf forecasts accurately estimate the time of the peak but just miss the full severity of the flooding. The timing of the breaching of the bank is accurately estimated. Whilst the lpacf forecast is less smooth than the Kalman Filter, with its 95% confidence intervals it appears to capture almost all the smaller peaks in the data. The confidence intervals of the lpacf forecast are typically wider than that of the Kalman Filter which reflects the uncertainty due to the missing tidal influence inducing nonstationary.

The results shown here are encouraging. Future work includes investigating the affect of bandwidth and "stabilization" lengths for the lpacf methodology. Additionally investigating where the methods differ and when they over

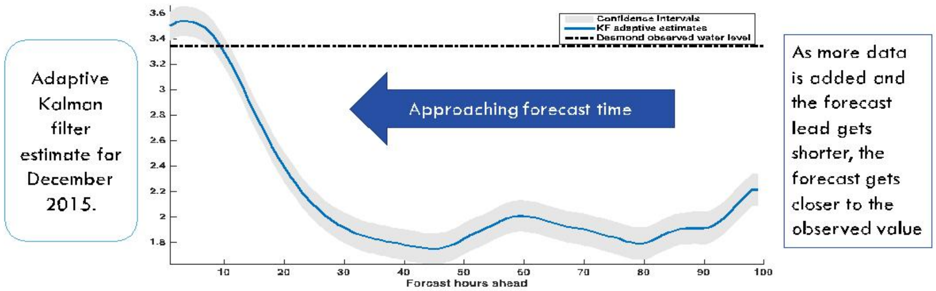
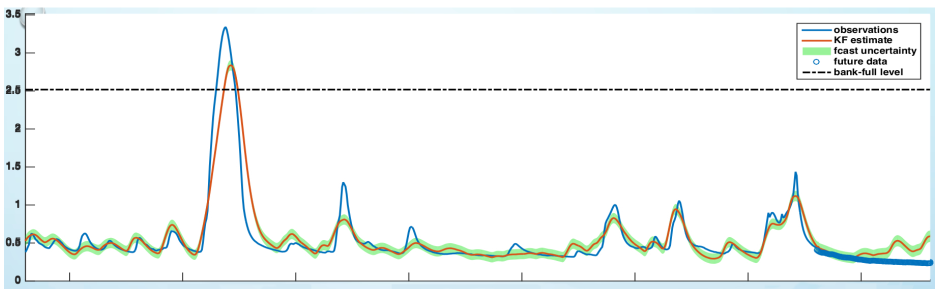
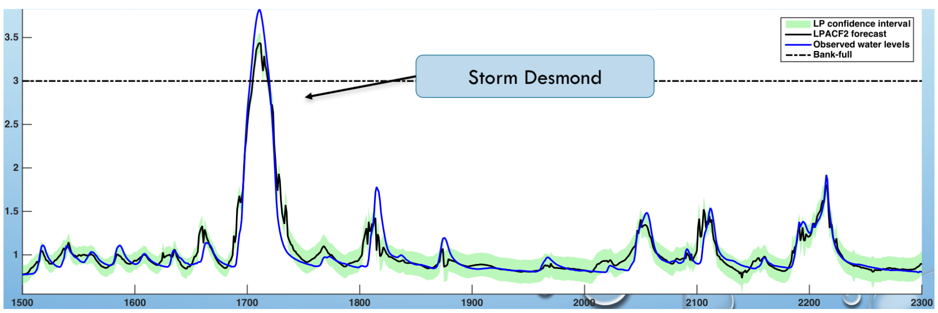


Figure 3: Kalman Filter forecast.



(a)



(b)

Figure 4: Actual and forecast measurements of river levels at Skerton for around Storm Desmond (a) Kalman forecast. (b) lpacf forecast.

or under estimate as it would be useful to provide guidance on when it would be better to use one method rather than the other. Unfortunately the project did not get enough time to consider the interaction of river water and tides which is important to forecasts at Lancaster Quay. The influence of the river on the tidal cycle is evident in the fast rise and slow tidal recession and will be included as further work. Finally, the lpacf forecasting approach relies on the model residuals as it assumes a mean-zero process. It would be interesting to see how the model performs if it could be extended to processes with non-zero mean which would allow the original data to be modelled directly.

Potential for initiating or developing future multidisciplinary collaborations: The project employed a post-doctoral researcher with a PhD from Environmental Science that had a strong deterministic computational component. The project exposed the researcher to the stochastic interpretation of the problem and trained them in interdisciplinary research. The nature of this project has given the researcher the demonstrable interdisciplinary skills to build a successful career in this area. Additionally the project has brought together Dr Killick and Dr Tych who plan to submit a joint funding bid to EPSRC (Living with Environmental Change) / NERC. Furthermore, Dr. Killick will deliver a workshop at the International Glaciology Society British Branch Meeting in September 2017.

References

Hoitink, A. J. F. and Jay, D. A. (2016). Tidal river dynamics: Implications for deltas. *Reviews of Geophysics*, 54(1):240–272.

Killick, R., Knight, M., Nason, G., and Eckley, I. (2016). The local partial autocorrelation function and its application to the forecasting of locally stationary time series. *(under revision).*

Kretzschmar, A., Tych, W., and Chappell, N. A. (2014). Reversing hydrology: Estimation of sub-hourly rainfall time-series from streamflow. *Environmental Modelling & Software*, 60:290 – 301.

Leedal, D., Weerts, A. H., Smith, P. J., and Beven, K. J. (2013). Application of data-based mechanistic modelling for flood forecasting at multiple locations in the Eden catchment in the National Flood Forecasting System (England and Wales). *Hydrology and Earth System Sciences*, 17(1):177–185.

McMillan, H. K. and Westerberg, I. K. (2015). Rating curve estimation under epistemic uncertainty. *Hydrological Processes*, 29(7):1873–1882.

Nason, G. P., von Sachs, R., and Kroisandt, G. (2000). Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum. 62:271–292.

Potter, D., Folkard, A., and Ilic, S. (2016). Compound tide and overtide generation by tidal-stream turbines. *(to be submitted).*

Romanowicz, R., Young, P., and Bevan, K. (2006). Data assimilation and adaptive forecasting of water levels in the river severn catchment, united kingdom. *Water Resources Research*, 42(6).

Smith, P. J., Beven, K. J., and Horsburgh, K. (2013). Data-based mechanistic modelling of tidally affected river reaches for flood warning purposes: an example on the River Dee, uk. *Quarterly Journal of the Royal Meteorological Society*, 139(671):340–349.

Verkade, J. S. and Werner, M. G. F. (2011). Estimating the benefits of single value and probability forecasting for flood warning. *Hydrology and Earth System Sciences*, 15(12):3751–3765.