Automatic quality-control on high-frequency tide-gauge data in the Caribbean

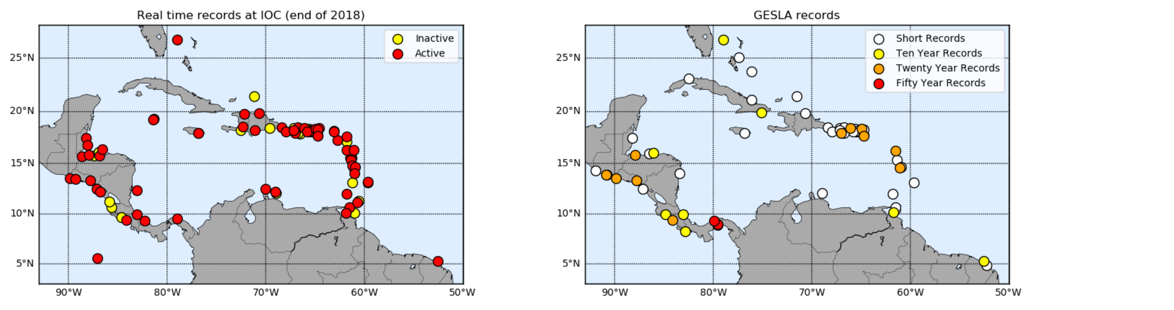
Joanne Williams, David Byrne, Svetlana Jevrejeva, Andrew Matthews, and Judith Wolf

joll@noc.ac.uk — National Oceanography Centre, Liverpool, UK — EGU Vienna 2019

National Oceanography Centre, Natural Environment Research Council

Introduction

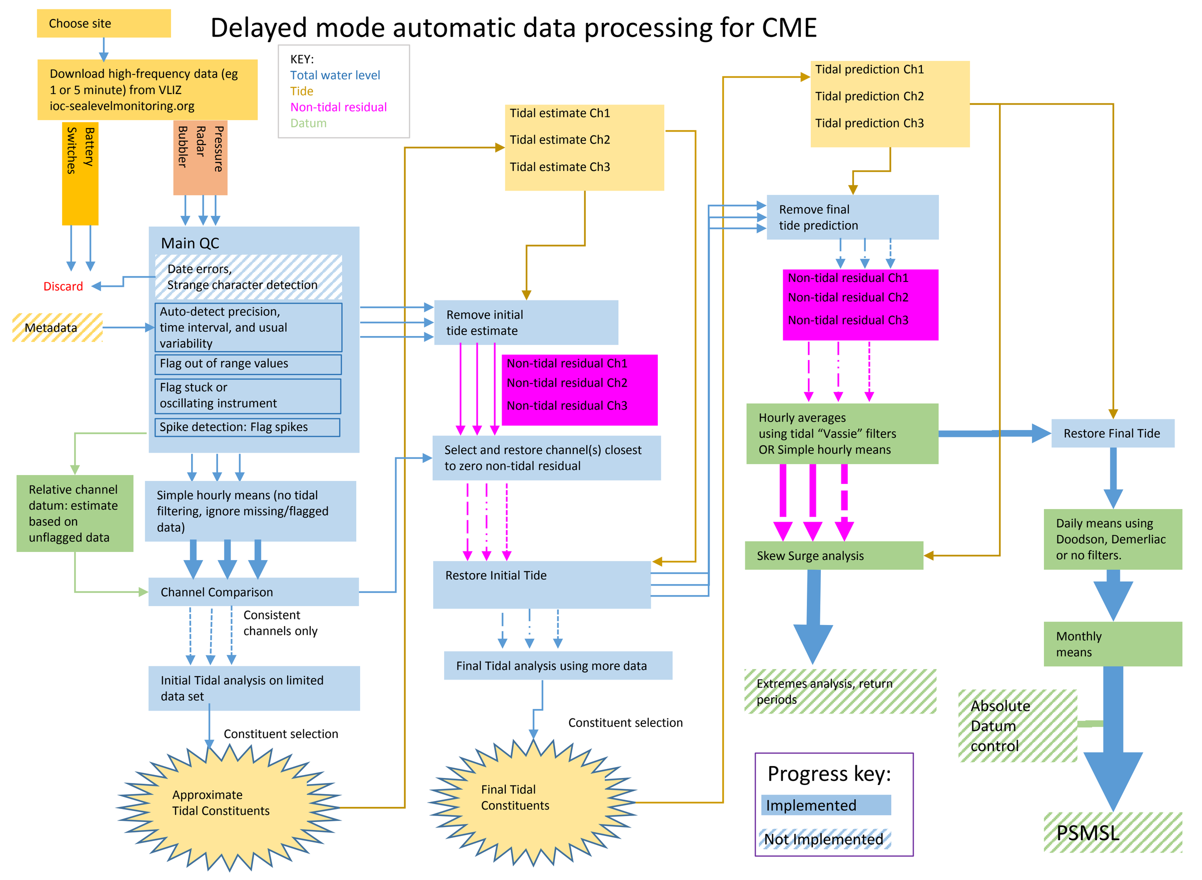
We present software for automatic quality control (QC) for tide gauges, including generalised comparison of instrument channels, fitting and predicting tides using irregular high-frequency data.
There are many tide gauges around the world for which research-quality data is not available for sea-level studies, including gauges that are maintained primarily for tsunami monitoring. In some cases high-frequency data is available for download through the Intergovernmental Oceanographic Commission (IOC) http://www.ioc-sealevelmonitoring.org, but manual QC has been too labour intensive.



In the Caribbean there are many more active tide gauges than are available in quality-controlled public repositories. US gauges omitted from map.

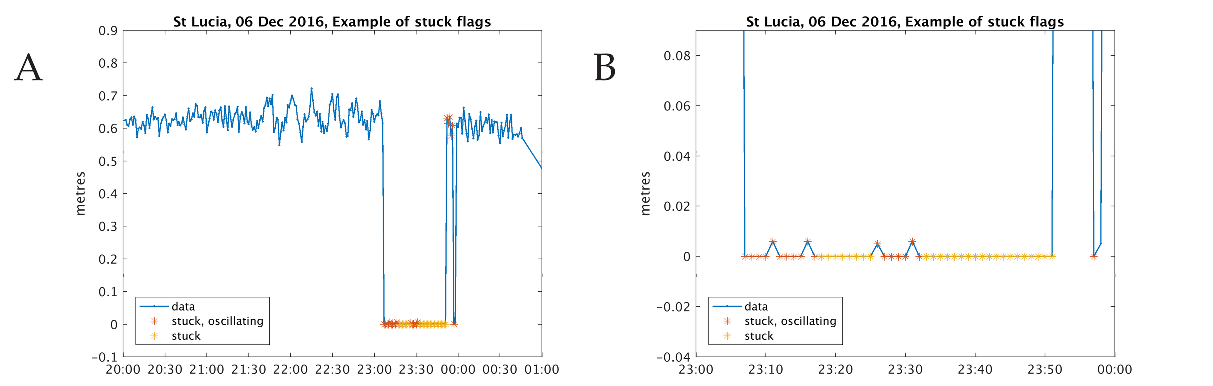
Outline of QC processs

We have developed a single Matlab package for QC of tide-gauge data downloaded from the IOC. A key element of this is the ability to fit tides to irregular data, at an early stage in the process. This enables comparison of channels using non-tidal residual, and without interpolation of missing data.



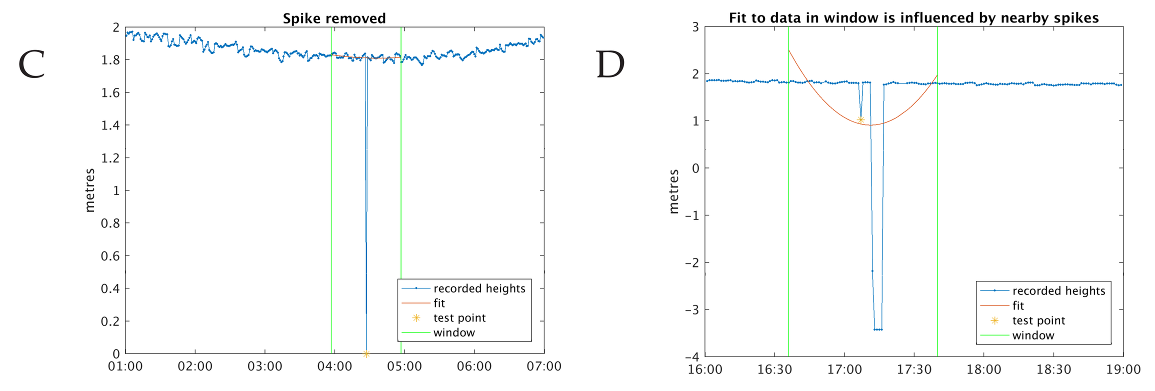
Stuck instrument test

The stuck instrument test (A) is to flag values when there is no change in the water level after a number of time steps. This test may miss faults where the instrument oscillates. The stuck-oscillating test (B) finds these, and flags intermediate points as also suspect.



Spike detection

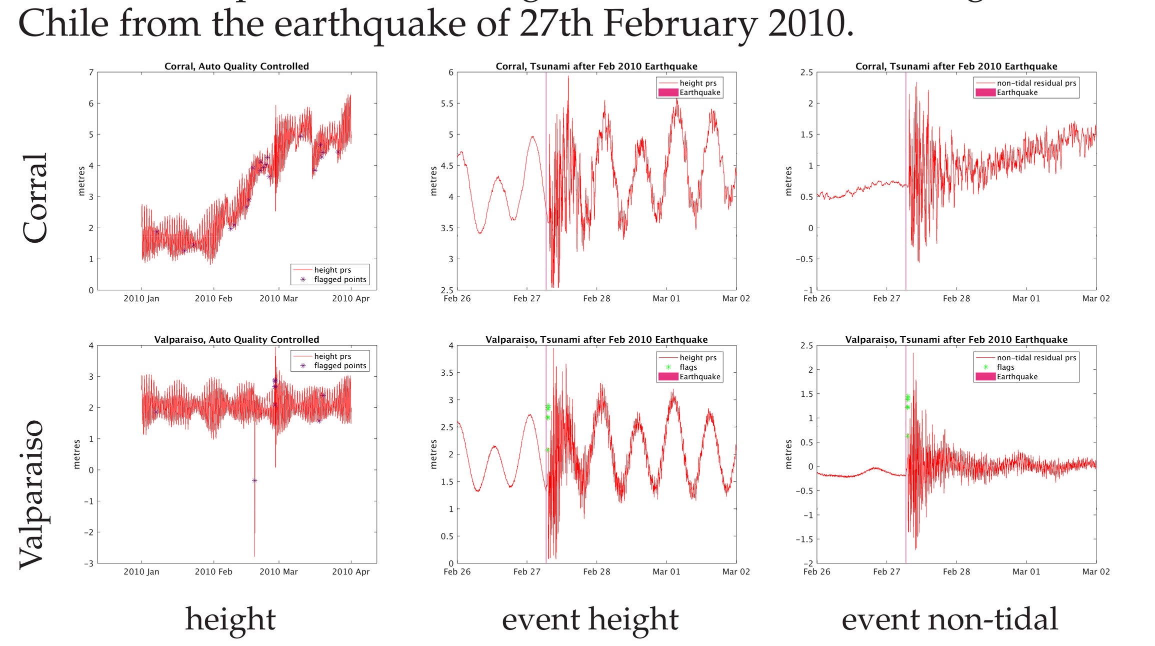
The spike detection algorithm (C) is an implementation of deviation from a spline. A polynomial is fitted over a short period of data, using an efficient least squares fit. The fit is not robust against outliers (other than the test point which is omitted), so problems may arise identifying spikes in the neighbourhood of other spikes or datum shifts (D).



Points are flagged if they are more than 6 scaled MAD (median absolute deviation) from the spline. This is more robust at identifying spikes in the region of other outliers than RMSE.

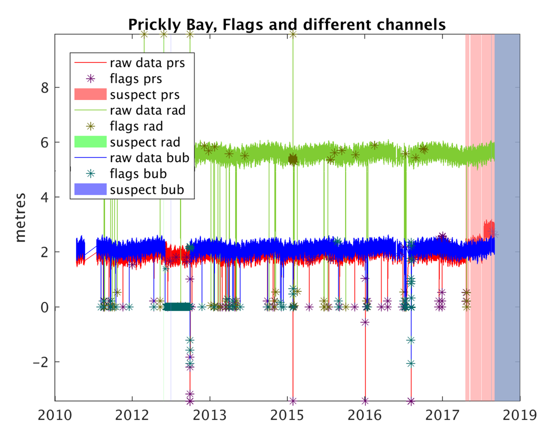
Do tsunamis pass spike detection?

We test the spike detection algorithm on tsunamis using records from Chile from the earthquake of 27th February 2010.



At Corral there was a low frequency signal in the pressure sensor during the first half of 2010 with a drift of around 4m. Despite this, the algorithm still succeeds in removing tidal frequencies and the tsunami itself is well preserved. At Caldera and Coquimbo also (not shown) the spike detection correctly passes the event. However at Valparaiso the first 6 records of the tsunami were incorrectly flagged as spikes, possibly in part due to adjacent data gaps.

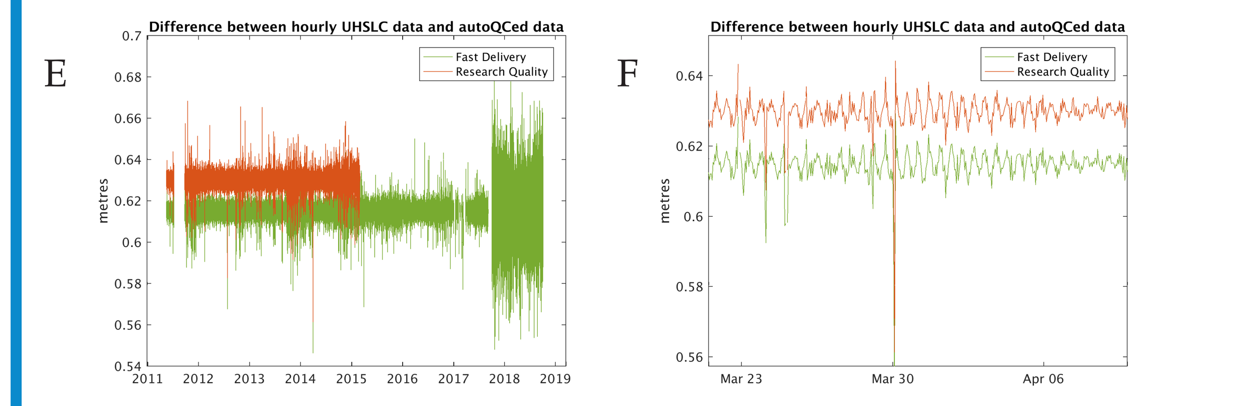
Channel Comparison



Comparing neighbouring channels is much easier after early tidal analysis. At Prickly Bay, Grenada, there is a drift in the pressure sensor clearly obvious by eye from around August 2018. Using the non-tidal residual it is clear that this begins earlier, about the beginning of April. Since the radar and bubble give consistent results over this period, it is very likely that there is a fault on the pressure sensor, and all the data from this channel is rejected for this period.
Generalised multiple channel comparison requires extremely careful handling of cases. In order to compare channels they must be at the same time interval, so we use hourly averages for this part of the process and assume that very short periods where channels overlap should also be flagged.

Comparison to UHSLC

Hourly and daily data for Prickly Bay, Grenada is available at UHSLC (University of Hawaii Sea-Level Center). The Fast Delivery (FD) service provides data up to recent months, but Research Quality (RQ) is only available up to 2015.
A 0.62m constant difference (E) is attributable to arbitrary datums applied during the autoQC. This is about 1cm greater on RQ, as if some manual datum correction from levelling information has been applied at UHSLC.

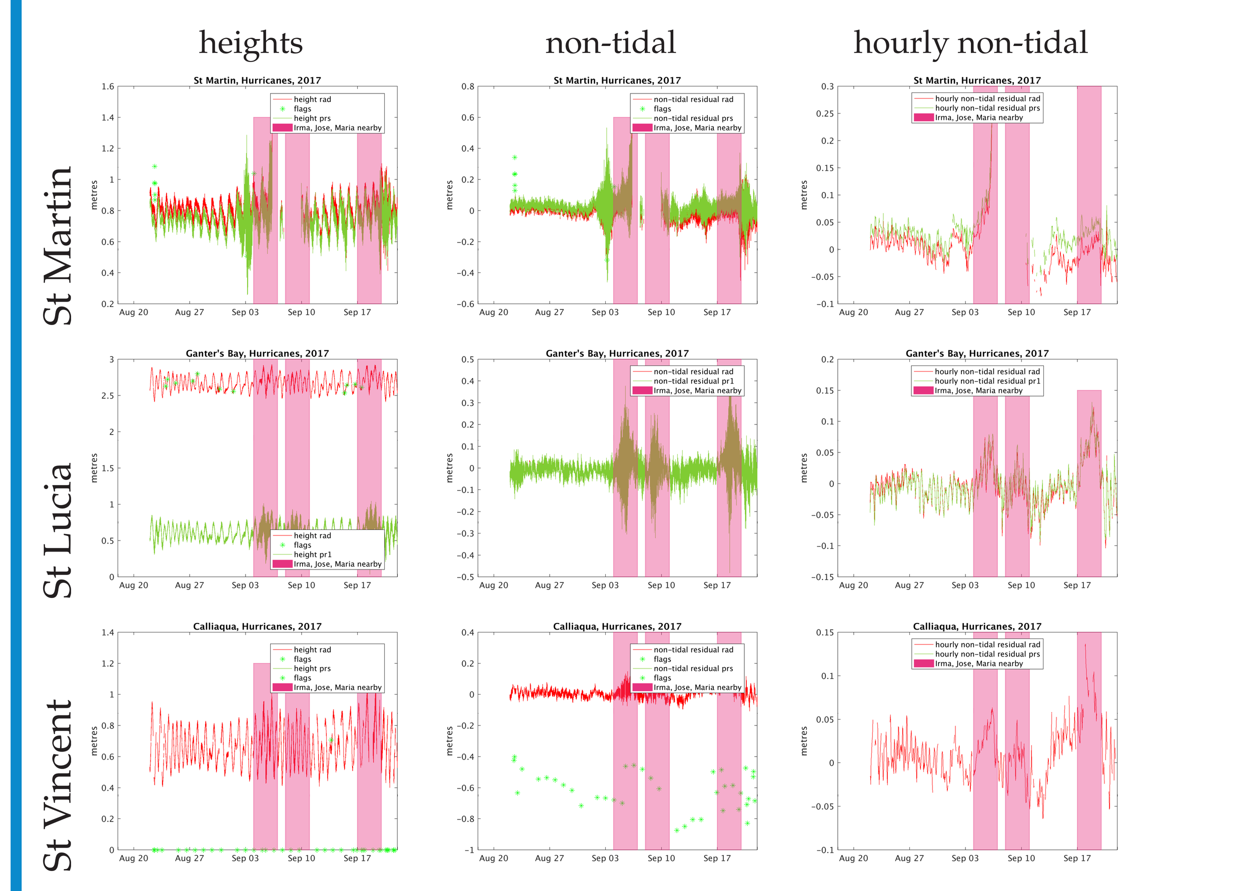


Difference between UHSLC hourly data, FD or RQ, and the single stitched channel from autoQC, 2011–2018

Until 2017 there is also a 1.5cm semi-diurnal periodic difference (F) between both FD and RQ and the autoQC data. This peaks on the rising tide and is consistent with a phase difference of 2.5 minutes between the two data sets. This could be a clock correction, or an error during the averaging from high frequency to hourly data, for example from a mis-aligned 5 minute filter in the UHSLC code. It remains for a selection of filter techniques in the autoQC and averaging on total water levels or non-tidal residuals accounts for less than 3mm.
After October 2017 the difference in the hourly data increases and is irregular. The cause is not certain.

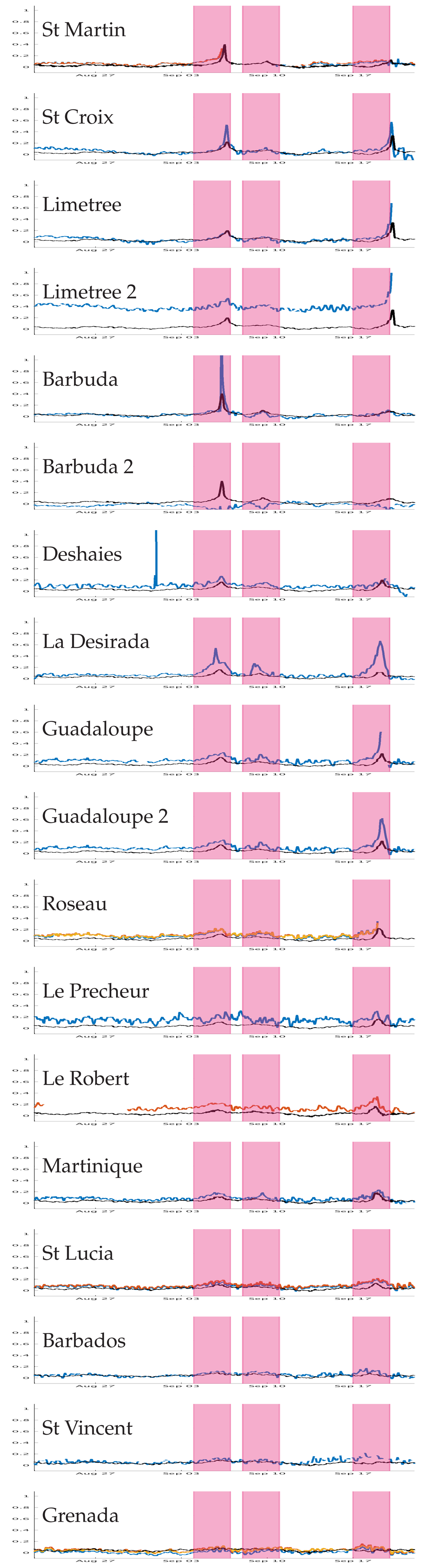
Observations of hurricanes

Even the small Caribbean tides disguise the average water level changes due to the hurricanes (Irma, Jose and Maria from 2017). But after removal of the tide the one-minute average of the waves is visible in the high-frequency data. The tides are fitted using harmonic analysis with the Matlab package NOCtide.



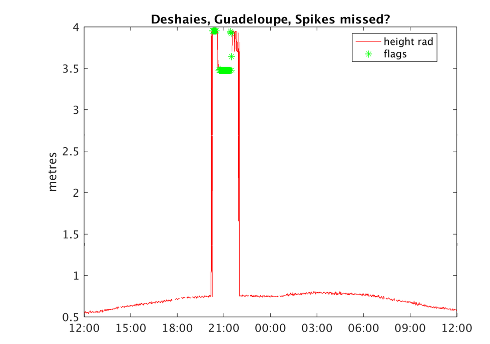
At Ganter's Bay, St Lucia and further north at St Martin, both events were more severe than at Calliaqua (St Vincent). At St Martin before the severity of the storm interrupted the record, we see hourly average non-tidal residuals of over 0.6m.

Comparison to model



NEMOSurge hourly non-tidal residual (black), observed hourly non-tidal residual after auto QC (other colours), August and September 2017. Approximate dates of Hurricanes Irma, Jose and Maria are indicated. Gauges are listed north-to-south. Vertical scale is 1 m.

With the QC tool, we can rapidly access and process data for a large number of gauges, providing a tool for surge model validation. These 18 gauges from the east Caribbean were processed for 2017 with no manual intervention (except to check which had data during the hurricane season).
At Limetree 2, there is a constant offset in the auto-QCed non-tidal residual, suggesting that there may be an instrument drift that the processing is not correcting. This may also be the case at Le Precheur and Le Robert, although without external confirmation we cannot tell if there is instead a process that is not captured by the surge model.
Barbuda 2 is obviously an observation error - it is an exact inversion of Barbuda. At these gauges, the surge is observed at 1.8m, far exceeding the model.
At Deshaies there is an extra peak that is not due to the hurricanes. In the high-frequency data, it can be seen that it is due to a group of spikes that have gone unidentified. This would be resolved by an extra pass of the spike detection algorithm, but it illustrates that results from the auto-QC still require some additional inspection.



Spike cluster at Deshaies

Acknowledgements

This work was funded under the Commonwealth Marine Economies Programme, project "Climate Change Impact Assessment: Ocean Modelling and Monitoring for the Caribbean CME states", workpackage 2.8. It was also supported by the National Environmental Research Council under National Capability Official Development Assistance (NC-ODA), ACCORD programme.
We thank colleagues at National Oceanography Centre for tidal analysis code, Begona Perez Gomez for assistance with spike detection and John d'Errico for Matlab file exchange contributions.
We are very grateful to VLIZ and the IOC for their provision of tide gauge data at so many gauges; and to the hundreds of teams around the world who fund, install, maintain those gauges, and who make their data available.

Website

For more information, see https://psmsl.org/cme/ , where the Matlab code will be made available in due course.