DEVELOPING IMPROVED METHODOLOGIES FOR SEA-LEVEL TREND ESTIMATION

Phil Watson¹ and Ron Cox¹

One of the most fundamental, critical environmental issues confronting mankind into the foreseeable future remains the ominous spectre of climate change, in particular the pace at which impacts will occur and our capacity to adapt. Sea level rise is one of the key artefacts of climate change that will have profound impacts on global coastal populations. It is estimated that some 600 million people live within the Low Elevation Coastal Zone (contiguous areas along the coastlines of the world less than 10m above Mean Sea Level) and considered vulnerable to storm surges and projected sea level rise. Although extensive research has been undertaken into sea level rise, there remains considerable conjecture and scientific debate about the temporal changes in mean sea level and the climatic and associated physical forcings responsible for them. The improvement of analytical techniques to isolate mean sea level (or trend) from the contamination of the decadal and inter-decadal (and longer) cyclical influences and noise remains the overwhelming aim of researchers in the sea level rise field. In effect, comparison from one estimate of mean sea level to another, in part has become an indirect qualitative view of the merit of the analytical approach applied. An innovative and transparent process by which to identify the most appropriate analytical technique for isolating the mean sea level signal is to test such approaches against “synthetic” (or custom built) data sets with a known MSL signal. Testing of contemporary analytics against a “synthetic”, physics based data set will substantially improve the rigour and confidence around current estimates of MSL and its temporal characteristics to better inform projection modelling endeavours and improve public education around the issue of sea level rise. This paper provides a summary of the process associated with the development of the core synthetic data set used for testing purposes proposed as part of a post graduate research program. A more detailed discussion will be published in an upcoming issue of the Journal of Coastal Research.

Keywords: Synthetic data set, mean sea level, climate change.

INTRODUCTION

One of the most fundamental, critical environmental issues confronting mankind into the foreseeable future remains the ominous spectre of climate change, in particular the pace at which impacts will occur and our capacity to adapt (Watson, in press). Sea level rise is one of the key artefacts of climate change that will have profound impacts on global coastal populations. Although higher sea level only directly impacts coastal areas, these are the most densely populated and economically active land areas on Earth (McGranaghan, Balk, and Anderson, 2007; Nicholls, 2011; Sachs, Mellinger, and Gallup, 2001).

It is estimated that some 600 million people live within the Low Elevation Coastal Zone (contiguous areas along the coastlines of the world less than 10m above mean sea level) and considered vulnerable to storm surges and projected sea level rise. This threatened population is growing significantly (McGranaghan, Balk, and Anderson, 2007) and it will almost certainly increase in the coming decades, especially if the strong tendency for coastal migration continues (Nicholls, 2011).

Watson (in press) advises that although extensive research has been undertaken into sea level rise, there remains considerable conjecture and scientific debate about the temporal changes in mean sea level and the climatic and associated physical forcings responsible for them. In particular, significant debate has centred around the issue of a measurable acceleration in ocean water level records (see Baart, Van Koningsveld, and Stive, 2012; Donoghue and Parkinson, 2011; Houston and Dean, 2011a,b; Rahmstorf and Vermeer, 2011; Rhein et al., 2013; Watson, 2011), a feature central to projections based on the current knowledge of climate science (IPCC, 2013).

The complexity of the broad range of physical influences embedded within monthly and annual average ocean water level data sets used for sea level research deems that such records are not able to be definitively deconstructed or parametrically modelled to precisely estimate mean sea level at a given point in time or location. Inevitably, the corroborations of a range of alternative techniques is relied upon to converge on an estimate of mean sea level and associated trends, velocities and accelerations. With so much critical reliance on accurate estimates of these physical parameters to understand climate change and improve future projections, there is increased urgency in identifying the better performing analytics for defining the temporal characteristics of mean sea level.

The improvement of analytical techniques to isolate mean sea level (or trend) from the contamination of inter-annual to inter-decadal (and longer) cyclical influences and noise, remains the overwhelming aim of researchers in the sea level rise field. The complexity of the dynamic influences

¹ School of Civil and Environmental Engineering, University NSW, Sydney, NSW 2052, AUSTRALIA
and noise embedded within ocean water level data sets has led sea level research toward successively more sophisticated time series analytical techniques to improve estimates of the trend. In particular over recent decades, the emergence and rapid improvement of data adaptive approaches to isolate trends from nonlinear, non-stationary and comparatively noisy environmental data sets such as Empirical Mode Decomposition (Huang et al., 1998; Wu and Huang, 2009), Singular Spectrum Analysis (Broomhead and King, 1986; Golyandina, Nekrutkin, and Zhigljavsky, 2001; Vautard and Ghil, 1989) and Wavelet analysis (Daubechies, 1992; Grossmann and Morlet, 1984; Grossmann, Kronland-Martinet, and Morlet, 1989) are theoretically encouraging.

In the absence of an absolute knowledge of the mean sea level signal (or trend) for a particular record, the accuracy of the trend has increasingly been inferred from the assumed sophistication of the underpinning analytical approach. An innovative and transparent process by which to identify the most efficient analytical technique for isolating the mean sea level signal is to test such approaches against “synthetic” (or custom built) data sets with a known mean sea level signal.

From the extensive literature available and through consultation with some of the world’s leading oceanographers, sea level researchers and subject matter experts (refer Acknowledgements section), ocean water level records can be considered to be a complex composite of a mean sea level signal and a range of key dynamic components. These dynamic components include: seasonal influences, pole tide, cyclical longer-term tidal harmonics (e.g., nodal tide), climate mode influences; and random environmental noise. For a detailed discussion on these components refer Watson (in press).

METHODS
This section summarises the methodology applied to the development of the synthetic data set, in particular, the construction of each of the core elements within the data set.

Overview
In order to be effective, the synthetic data set developed for this research has been specifically designed to mimic the key physical characteristics embedded within real-world ocean water level data. Hence, the synthetic data set comprises a range of known dynamic components added to a nonlinear, non-stationary time series of mean sea level.

A schematic representation of the elements comprising the core synthetic data set is depicted at Figure 1. This data set has been designed as a monthly average time series spanning a 160 year period from 1850 – 2010. This time period has been selected to reflect the predominant date range for the longer Permanent Service for Mean Sea Level (PSMSL) data holdings.

Figure 1: Schematic representation of core synthetic data set.
The core synthetic data set has been designed to be as generically applicable as possible (i.e., reflective of the environmental attributes and signals captured by a tide gauge located anywhere worldwide). In order to do this, each of the key dynamic influences are represented by a bin of monthly time series spanning the full 160 year period, reflecting the range of real-world measured influences for each particular component. The synthetic data set contains 20,000 separate time series, with each time series generated by successively adding a randomly sampled signal from within each of the dynamic components to the fixed mean sea level signal.

The selection of 20,000 randomly generated time series represents a reasonable balance between optimising the widest possible set of complex combinations of real-world signals and the extensive computing time required to analyse the synthetic data set. Further, the 20,000 generated trend outputs from each analysis applied to the data set provides a robust means of statistically identifying the better performing techniques for extracting the trend.

In addition to the published literature, the data from 43 selected PSMSL sites have been analysed and decomposed to estimate genuine seasonal signals and noise components using the respective methods detailed below pertaining to these components. The selected sites (Figure 2) were based on maximising a range of factors including global spatial coverage, length and quality of records and range of environmental influencing factors. The other dynamic components and the fixed mean sea level signal have been specifically based on the scientific literature and the methods used to generate the bins of real-world signals are detailed in the following sections dedicated to each component.

Figure 2: Location of PSMSL stations used to derive seasonal components. Co-ordinate system is degrees latitude and longitude.

Mean Sea Level
Within the synthetic data set, the mean sea level has been developed as a smoothed, non-linear time series signal. This has been achieved by applying a broad cubic smoothing spline to a range of points over the 1950-2010 time horizon reflective of the general characteristics of the global trend of mean sea level (Bindoff et al., 2007; Church et al., 2013) accentuating the key positive and negative “inflexion” points evident in the majority of long ocean water level data sets (Woodworth et al., 2009). The cubic smoothing spline has been used to predict the mean sea level time series for each respective month over
the designated time span. It is noteworthy that the error margins on the global mean sea level reconstructions broaden significantly moving back in time owing to the quality and quantity of available records prior to 1900. Therefore, the portion of the mean sea level time series prior to 1880 has been generally assumed. This monthly time series signal (Figure 3) is the key fixed signal embedded within each of the 20,000 generated time series of the core synthetic data set.

Figure 3. Generated monthly time series signal of mean sea level (MSL).

**Seasonal Influences**

In order to build the seasonal component bin for the synthetic data set, the selected gauge sites (Figure 2) have been decomposed to isolate the seasonal signal. These records provide a broad mixture of seasonal signatures, involving locations which also encompass significant monsoonal and glaciological cycles. The seasonal signal has been extracted from each of the monthly average gauge records using three (3) separate established methodologies, including:

- **Method 1** - locally weighted polynomial regression smoothing (or “LOESS”) of the seasonal sub-series (i.e., the series of all January, February, March values, etc.);
- **Method 2** - spectral analysis using a de-seasonalising band stop filter available in the IDEOLOG software package (Pollock, 2008); and
- **Method 3** - fitting an autoregressive integrated moving average (or ARIMA) model to the data. This approach has been undertaken using the X-12-ARIMA seasonal adjustment software package developed by the US Census Bureau (US Census Bureau, 2009).

**Long Period Tidal Harmonic Influences**

The 18.61 year nodal cycle is the key tidal harmonic influence of relevance to the synthetic data set. The literature is replete with equilibrium theory and measured amplitudes and phase angles of signals at the nodal frequency from tide gauge records and satellite altimetry. Sinusoidal curves with a period of 18.61 years covering the range of nodal amplitudes and phase angles described in the literature, have been used to develop the bin of time series representing nodal tide signals.

**Pole Tide Influence**

Although the pole tide has separate annual and 433 day (Chandler “wobble”) components, the amplitude of the annual pole tide harmonic (less than ≈ 5 mm) is relatively small compared to that of the seasonal signal at the same frequency, and thus has not been considered further for addition to the synthetic data set. However, in order to represent the component of the pole tide at the Chandler frequency, sinusoidal curves with a period of 433 days have been developed with time varying amplitudes. The maximum amplitude of 18 mm in early 1993 determined by Desai (2002) from satellite altimetry data, has been fitted to the time varying time series of polar motions determined by Malkin and Miller (2010) spanning the period from 1850 to 2010. This maximum amplitude time varying signal has been factored to generate time varying sinusoids with a period of 433 days covering the phase and amplitude range of measured pole tide signals discussed in the literature. A small number of time series have also been generated to reflect the anomalously high amplitude signals at this frequency that have been measured in the North Sea region (up to 40 mm).

**Climate Mode Influences**

The significant work of Trenberth *et al.* (2007) confirm that of the many identified climate patterns, the majority of inter-annual variability in circulation and surface climate can be described by four key patterns (namely SAM, NAM, ENSO and PDO). From the numerous studies correlating climate mode
influences to sea level anomalies, there is also strong evidence of in particular, the dominant global ENSO signal superimposed on strong localised patterns of climate variability on shorter monthly, seasonal and inter-annual timescales of influence (such as SAM, NAM/NAO).

ENSO is the dominant, global signal with power in the inter-annual to decadal frequency band, with varying localised influence (Trenberth et al., 2007). Although closely correlated to ENSO with key influence in the north and western pacific region, the PDO has also been shown to have a global influence on mean sea level at bi-decadal and longer frequencies (Hamlington et al., 2013). In order to encompass the widest possible range of climate mode influences likely to be embedded within ocean water level data sets, the “global” climate mode influence has been constructed as a composite of signals reflective of both ENSO and PDO signatures added to a “regional” influence such as SAM and NAM (schematically indicated in Figure 4). The mechanics of how the respective bins of signals have been constructed are detailed in Watson (in press).

Figure 4. Schematic representation of climate mode component.

Noise

The 43 PSMSL gauge sites used for seasonal decomposition (refer Figure 2) have similarly been used to estimate the likely range of white noise attributes within conventional monthly averaged ocean water level data sets. The process applied to extract the white noise component is relatively straightforward, involving 3 key steps:

Step 1 - isolate and remove seasonal influences using “stl decomposition” function in the R analytical software package (R Core Team, 2014)(refer Method 1, Seasonal Influences section);
Step 2 - fit generalised least squares (GLS) linear regression model to “de-seasonalised” time series data to extract residuals; and
Step 3 - fit autoregressive (AR) model to the residuals from the fitted GLS regression model to remove the serial correlation. Analysis of the correlogram of the Partial Autocorrelation Function (PACF) has been used to determine the optimum lagged AR model to remove serial correlation from the residuals.

Only 5 sites (Arrecife, Canary Islands; Churchill, Canada; Jolo, Philippines; Mossel Bay, South Africa; Visakhapatnam, India) proved unsuitable for this process, largely due to either significance of gaps in the available data record or significance of the deviation from linear of the assumed underlying trend. From this analysis, the standard deviation of the monthly residuals ranged from a low of 23 mm (Magueyes Island, Puerto Rico) to a high of 290 mm (Trois-Rivières, Canada), the latter affected by significant fluctuations from year to year in glaciological discharges directed through the Saint Lawrence River system. The next highest standard deviations recorded were 144 mm (Helsinki, Finland) and 143 mm (Cuxhaven, Germany).

RESULTS

This section provides a brief summary of the key dynamic cyclical components (Figure 1) used to generate the 20,000 time series comprising the synthetic data set.
The seasonal signals extracted via Methods 1 and 2 produced near identical results for the respective monthly amplitude at each station. To avoid unnecessary duplication of results, only the Method 1 and Method 3 outputs have been used to compose the seasonal time series components in the synthetic data set. From the analysis, the phase and amplitude of the seasonal signals varied significantly ranging from the smallest amplitude signal (-24 to +3 mm) recorded at Argentine Islands (Antarctica) to the largest amplitude signal (-630 to +1220 mm) recorded at Trois-Rivières (Québec, Canada) situated at the confluence of the Saint-Maurice and Saint Lawrence Rivers and significantly affected by glaciological cycles. Based upon this analysis, the seasonal component of the synthetic data set consists of a bin of 82 separate time series representing a wide range of repetitive and time varying seasonal signals expected within ocean water level data sets.

There is a considerable body of literature dedicated to both the equilibrium theory of the nodal tide and associated measurements from both tidal records and satellite altimetry. In consideration of the extensive literature, the component of the synthetic data set representative of the lunar nodal tide signal has been randomly sampled from a bin of sinusoidal curves with an 18.61 year period, with amplitudes ranging from zero to 30 mm (in 0.5 mm increments) and phase angles ranging from zero to 180° (in 10° increments). In total, 1141 time series (including a zero time series) have been generated to represent the range of nodal tide influences expected within ocean water level data sets.

A peak amplitude time series signal has been used as the basis for developing time varying amplitude sinusoids with a period of 433 days to represent signals reflective of the pole tide influence in the synthetic data set. To accommodate the range of amplitudes of this signal that can vary from zero to the peak, the maximum time varying amplitude time series has been factored from zero to unity in 50 equal increments. These sinusoids have been determined both for phase angles of zero and 180° to represent the diurnal characteristics of the signal. In order to represent the larger amplitude pole tide signals experienced in the North Sea region, 5 additional time series have been generated with factors corresponding to maximum time varying amplitudes of 24, 28, 32, 36 and 40 mm with zero phase. In total, the pole tide component is represented by a bin of 106 time varying sinusoids.

Some of the most dynamic influences on ocean water levels are directly attributable to large scale modes of climate variability which are described commonly by patterns which cover a broad range of climatological variables on particular spatial and temporal scales. Although many teleconnections and patterns have been identified, much of the inter-annual variability in the circulation and surface climate can be related to a small number of patterns related to the SAM, NAM, ENSO and PDO (Trenberth et al., 2007).

The global climate mode influence within the synthetic data set has been based on fitting maximum measured amplitude signals to ENSO and PDO indices based on global characteristics of sea surface anomalies measured during the altimetry period (Zhang and Church, 2012). In each case, portions of the respective indices were recycled to pad the record over the full length of the synthetic data set. In total, the global climate mode influence component is represented by a bin of 160 complex time series signals.

The regional climate mode influence within the synthetic data set has been based on fitting maximum measured amplitude signals to SAM and NAO indices based on regional characteristics of sea surface anomalies measured by Aoki (2002) around Antarctica and by Woolf, Shaw and Tsimpolis (2003) from altimetry data around the North Sea, the Mediterranean and eastern parts of the North Atlantic, respectively. As with the global climate mode indices used, portions of the respective SAM and NAO indices were recycled to pad the record over the full length of the synthetic data set. In total, the regional climate mode influence component is represented by a bin of 127 complex time series signals.

From the analysis to isolate the white noise residuals from the 43 PSMSL guage sites (refer Figure 2), the standard deviation of the monthly white noise residuals ranged from a low of 23 mm (Magueyes Island, Puerto Rico) to a high of 290 mm (Trois-Rivières, Canada), the latter affected by significant fluctuations from year to year in glaciological discharges directed through the Saint Lawrence River system. The next highest standard deviations recorded were 144 mm (Helsinki, Finland) and 143 mm (Cuxhaven, Germany). In order to generate a white noise component reflective of real-world attributes, a Gaussian (normal) distributed set of residuals of length 1920 months has been randomly sampled for each time series in the synthetic data set. The scale of each set of normally distributed white noise residuals has been determined by randomly sampling from a bin of standard deviations ranging from 20 to 300 mm (stepped in increments of 1 mm to 150 mm, then 50 mm thereafter; 134 standard deviations in total) to reflect the results from analysis of the gauge records sampled.
CONCLUSION

Ocean water levels are a complex mix of numerous cyclical dynamic influences on differing physical, spatial and temporal scales, superimposed on a comparatively low amplitude signal of mean sea level rise over time. Numerous time series analysis techniques are available for estimating the trend component from the contamination of the many cyclical dynamic influences and noise. All have inherent strengths and weaknesses and differing capacity in resolving the range, scale and frequency of signals commonly embedded within ocean water level data sets used to estimate mean sea level.

The synthetic data set described in this paper is based on complex randomly sampled signals that mirror real-world attributes of key dynamic components embedded within ocean water level data, added to a fixed, non-stationary, nonlinear signal of mean sea level. This data set comprising 20,000 unique monthly average time-series, provides a robust tool to test analytics for their utility to isolate the known mean sea level signal (or trend) with improved temporal resolution. Similarly, this monthly data set can be subdivided into shorter sections and annualized to investigate additional issues associated with record lengths and annual versus monthly records.

By identifying the better performing class of analytics for estimating trends from ocean water level data sets, we can gain improved insight into key temporal signatures in these records that in turn can better inform the calibration of AOGCMs at finer scale (regional). Similarly, by identifying the mean sea level trend with improved temporal accuracy, one can be more confident that key change points evident in the mean sea level record will be genuinely identified rather than an artefact of inherent limitations imposed by various analytical approaches. With so much debate and perceived uncertainty surrounding the interpretation of mean sea level signals over recent years, this innovative and transparent process will provide a significant step forward that will improve sea level research and its understanding.

For a more detailed discussion on the development of the synthetic data set readers are referred to Watson (in press). Analysis is now well advanced testing a broad range of time series analytics against the synthetic data set for their utility to isolate the mean sea level (trend) component. The results of this work will be detailed in a forthcoming paper and will form the basis of the development of an analytical package designed specifically for sea level researchers.

ACKNOWLEDGEMENTS

This research has benefitted significantly from direct consultations with some of the world’s leading oceanographers, sea level researchers and subject matter experts relevant to various components of the proposed synthetic data set. By consequence, I am indebted to the following individuals whose contributions have helped considerably to shape the final product and have ranged from providing specific and general expert advice, guidance and review, as well as provision of original research data outputs for use (in alphabetical order): Dr Shailen Desai (Jet Propulsion Laboratory, NASA, USA); Dr Bruce Douglas (retired, Department of Geography, University of Maryland, USA); Dr Ivan Haigh (National Oceanography Centre, University of Southampton, UK); Dr Angela Hibbert and Professor Chris Hughes (National Oceanography Centre, Natural Environment Research Council, UK); Professor Rob Hyndman (Department of Econometrics and Business Statistics, Monash University, Australia); Professor Huseyin Baki Iz (Department of Land Surveying and Geo-Informatics, Hong Kong Polytechnic University, Hong Kong); Dr Zinovy Malkin (Laboratory of Radioastronomy and Geodynamics, Pulkovo Observatory, Russia); Dr Andrew Robinson (Department of Mathematics and Statistics, University of Melbourne, Australia); Professor Philip Woodworth (former Chair, Global Sea Level Observing System (GLOSS) of the Intergovernmental Oceanographic Commission); and Dr Xuebin Zhang (Centre for Marine and Atmospheric Research, CSIRO, Australia).

REFERENCES


