Uncertainty and bias in electronic tide-gauge records: evidence from collocated sensors

Abbreviated title: Bias in electronic tide-gauge records

Stella Pytharouli\textsuperscript{a}, Spyros Chaikalis\textsuperscript{b} and Stathis C. Stiros\textsuperscript{b}

\textsuperscript{a}Department of Civil and Environmental Engineering, University of Strathclyde, G1 1XJ Glasgow, United Kingdom, corresponding author, email: stella.pytharouli@strath.ac.uk

\textsuperscript{b}Geodesy and Geodetic Applications Lab., Department of Civil Engineering, University of Patras, GR 26500 Patras, Greece, email: stathis.stiros@strath.ac.uk

Abstract

Understanding noise and possible bias in tide-gauge sensors is important for determining the mean sea level, its fluctuations and their climatic, geophysical and engineering implications, but not an easy task. In the past, this problem has been examined through comparison of different sensors in the laboratory, or through correlations of neighbouring sensors. In this study we identified and studied 10 cases of harbours with fully collocated sensors. Transient differences were found between collocated records. Pressure gauges were found significantly more sensitive to noise than radar-type sensors, and with higher chances of long-term transient bias. The amplitude of the observed bias is important, of the same order of magnitude with tsunami waves in the open sea and with seismic ground displacements. Only 9\% of the sensors analysed were found to satisfy the 1 cm accuracy criterion imposed by thePermanent Service for Mean Sea Level (PSMSL).

Keywords: collocated tide-gauges, bias characteristics in tidal records, uncertainty limits in tide gauge data, van de Casteele test, drift, offset
1. INTRODUCTION

Tide gauges have been established in harbours since the 18th c. to provide information on the quasi-periodic tidal fluctuations of the sea level obstructing navigation. Early gauges were simple vertical metred poles fixed in quiet places of harbours. Because of their efficiency, they were also used to define the geodetic datum (zero elevation for maps) and were evolved into mechanical, continuously recording floaters. In the last 50 years, tide gauges have also been used for the study of large-scale water dynamics, as well as of various coastal, tectonic, meteorological and climatic processes (e.g. Church and White (2006), Menéndez and Woodworth (2010), Zerbini et al. (2017), Wöppelmann and Marcos, (2016); Becker et al (2016))). Recently, the new generation of electronic high-rate recording tide-gauges (with a sampling rate of a few minutes to a few seconds) have been used to monitor transient effects such as tsunami waves (e.g. Fujii and Satake (2007), Satake et al. (2013).

Two major problems with tide-gauge data are that their number at a global scale was till recently limited (at present there are approximately 2,300 operational stations worldwide; PMSL, 2017), and that the deterministic astronomic signal is mixed with an essentially stochastic meteorological signal; for this reason, there is no easy way to control noise and bias in their records, i.e. to understand what their uncertainty limits are (Chelton and Enfield, 1986). As a consequence, the uncertainty in the recordings of tide-gauges has not been resolved even for the modern, electronic tide-gauges.

The advent of electronic sensors (usually accompanied with tele-transmission of the recorded oceanographic signals) however, was a real revolution. The reduced cost and simplicity of the new instruments gave the possibility to significantly increase the number of mareographic stations at a global scale, to install tide-gauges in buoys, and hence, to collect and analyse real-time signals at various centres, e.g. the British Oceanographic Data Centre (BODC), the University of Hawaii Sea Level Centre (UHSLC) etc., and even to combine two or more tide-gauges in the same or adjacent sites. Such collocated tide-gauges give for the first time the opportunity to compare the signature of the same oceanographic effect in different sensors, and from their differences to estimate the error properties of tide-gauges. Currently, the percentage of sites with collocated sensors is of the order of a few percent of the total sites covered with tide-gauges, but with a tendency to increase. The new situation revolutionises the understanding of the uncertainty of these instruments, which so far, was only derived
from measurements in laboratory conditions, usually very different from those in port or open seas environments.

In this article we examined a large number of tide-gauges at global scale, identified those which can be regarded as collocated, and analysed their differences. This approach represents the optimal way to document the uncertainties/error properties of tide-gauges from real data in different hydraulic conditions and geographic environments, and is novel. This is because so far, differences between collocated operating time gauges have clearly been noticed, but they were discarded, readily assigned to outliers, a priori assuming one sensor reliable. For this reason, the statistical information provided by differences in collocated sensors was ignored.

The output of this study is to identify and classify the differences between the different types of collocated sensors using raw, unfiltered data. We aim to provide information on how each sensor type behaves, which type is the most prone to bias, how big this bias can be and whether it can be related to other external factors which, if controlled, could reduce errors.

1.1. MAIN TYPES AND FUNCTION OF TIDE-GAUGES

The most common type of tide gauge is the traditional mechanical floater. The float remains on the sea surface and is connected to a system of pulleys and weights. As the float follows the sea level fluctuations, its movement is translated to actual water level on a recording paper (IOC, 1985). A floater with encoder (ENC) is the type of tide gauge where the mechanism that translates the float movement to water level values is an electric current.

In the last 20 years, the mechanical sensors were replaced by electronic sensors mainly of three types (in addition to the ENC sensors mentioned above): pressure sensors, acoustic tide-gauges and radar sensors. Their basic characteristics and principles of operation are described in IOC (2002) and are summarized below.

*Pressure sensors (PRS)* are commonly fixed below the low tide level and measure the hydrostatic pressure of a column of water above the sensor. This pressure is then converted to sea level assuming a value for the water density and atmospheric pressure at the surface of the sea. The problems of this type of tide sensors are the accurate determination of their datum and their drift (gradual change of their zero point, with a rate of 8mm/yr; Miguez et al, 2012) over time, mainly because of the aging of the sensor.
Acoustic sensors (AS) provide sea level values that are based on the time that an acoustic pulse transmitted by a sensor located directly above the sea level, requires to reach the sea surface and return back to the transmitter (time of flight). This signal is then converted to a distance from the sea surface. Acoustic pulses are, in the vast majority of cases, restricted within a vertical tube. This way, the sensor is less affected by changes in the temperature and pressure, factors that need to be taken into account for accurate sea level determination.

Radar sensors (RAD) are fixed above the maximum water level and measure the distance between the fixed radar and the water level below through a radio signal reflected on the water surface. Their advantage over an acoustic sensor is that they are not affected by temperature variations (IOC, 2002).

1.2. PREVIOUS STUDIES ON UNCERTAINTIES OF TIDE-GAUGES

All types of measurements are affected by measurement errors of different types, and tide-gauges are no exception. However, till recently, the understanding of their errors was nearly impossible, because tide-gauges are isolated instruments, there is no possibility to reproduce tidal effects, and no comparisons with other records (independent constrains) were possible. For example, it was found that in certain regions wrong type of lubricants were occasionally used for mechanical floaters, and saline water had the tendency to gradually reduce the viscosity of these lubricates with time. This led to gradually increased damping and attenuation of small-amplitude oscillations, hence biasing tide-gauge records. Some of these records have been discarded, but it is unknown how many others are affected, especially because no detailed metadata are available.

Industrial tests were used to derive the quality of specific types of tide-gauges, but results were limited to specific laboratory conditions (Miguez et al., 2008). In the last decades, however, efforts have been made to estimate their uncertainty in field, operational conditions. Lentz (1993) studied the accuracy of tide-gauge observations for the determination of the sea level using the differences between the values obtained from 3 pairs of tide-gauges within 50 km distances. He used the root mean square of monthly mean differences between observations to establish an upper bound for the error in measurements which was found to be of the order of 1.4 cm. He attributed this to errors in establishing the datum of the tide-gauges. His study also included the comparison of data between tide-
gauges and bottom pressure sensors which resulted in a similar value for the error in observations.

However, differences in the hydraulic energy along coasts are important even in nearby areas, as can be indirectly derived from significant variations of the elevation of the coastal biological zoning (Stiros and Pirazolli, 2004), and for this reason recordings of strictly non-collocated sensors are different, because they are influenced by unknown, different processes reflected in differences of local hydraulic energy of the sea water (mostly wave action).

Exploiting the emerging availability of collocated sensors Woodworth and Smith (2003) used data collected by a radar (RAD) and a bubbler pressure gauge (BPRS) in Liverpool (UK) over a period of one year at a sampling rate of 15 minutes. Linear regression revealed a small scale error between the two sensors, while the root mean square of the differences between corresponding values was of the order of 1.4 – 1.5 cm. The radar sensor seemed to be noisier than the bubbler, but both sensors had the same accuracy of 1 cm. Although the radar sensor was deemed to be biased up to 5 cm during a storm, the authors suggest that it should be considered for future projects/applications due to ease of installation and maintenance. The authors also suggest that a value of the root mean square (RMS) of the differences below 1.4 cm would ensure that the accuracy of the sensors is better than 1 cm. The latter is the accuracy requirement for all Global Sea Level Observing System (GLOSS) sensors, the global monitoring network that provides sea level data for oceanographic and climate research.

Mehra et al. (2009) were also favourable towards the installation of radar sensors. Their study, based at Verem (India), was focused on two types of sensors: a radar (RAD) and a pressure (PRS) gauge. Their work highlighted the effects of atmospheric factors (atmospheric pressure, water density and rainfall) on the pressure gauge and the advantage of radar sensors over other types of sensors.

A comparative study using four different types of sensors (acoustic, pressure, and two types of radar) was conducted by Miguez et al. (2005) at a port in NW Spain. They compared data sets from a 6-month period at sampling rates of 5 minutes, and mean hourly and daily values. They found that all sensors were providing data of acceptable accuracy but there were specific advantages and disadvantages associated with each of them. The radar sensor, for example, was not affected by bad weather conditions, but it could be easily vandalised due to lack of protection, while the acoustic and pressure sensors were biased by variations in the air pressure and salinity, respectively.
The effect of atmospheric variables on pressure gauges was the focus of the study by Mehra et al. (2013) who compared the data (5-minute average values) obtained by a radar (RAD) and a pressure (PRS) sensor deployed at three different locations within an area 10 km x 20 km. Each data set had a duration between 10 and 17 months. Their results revealed that the data from both types of sensors were similar if an atmospheric pressure correction was applied to the pressure sensor time series and they recommended the collection of atmospheric pressure data along with data collected by a pressure gauge for projects related to storm surges and tsunamis.

2. METHODOLOGY

Uncertainty in measurements is defined in terms of accuracy and precision (Mikhail and Ackermann, 1976). Precision is a measure of the consistency of measured (usually repeated) values, and also an indicator of the repeatability of the measurements. Accuracy, on the other hand, defines a quasi-random difference between recorded values and a “true” value (systematic error, bias), but it can rarely be estimated, mostly through comparison with the output of instruments of much higher quality/specifications, or using external constraints. A simple and efficient strategy to estimate both precision and accuracy is to analyse the output of several fully collocated identical sensors recording the same known reference signal (natural effect, for example an oscillation). In this case, the differences between sensors define the precision while the differences between sensors’ output and the input reference signal define accuracy of each sensor (cf. Moschas et al., 2015).

Tide gauge measurements cannot be reproduced to obtain a standard deviation, and in addition, their accuracy cannot usually be estimated for lack of independent evidence of the instantaneous sea-level. If, however, two collocated tide-gauge records A and B are available, the time series of the differences of their instantaneous recordings, corrected for their central tendency value, can identify differences which originate from uncertainties (noise, bias) in one or both instruments. As is schematically shown in Figure 1, if A and B are two tide gauges free of errors (i.e. noise), the time series of their difference would be a straight horizontal line (Figure 1a). If A and B contain low-amplitude white noise, i.e. a signal that contains a wide range of frequencies of
the same low amplitude and has a flat frequency spectrum, their difference will correspond to a nearly horizontal line with limited noise (Figure 1b). More important fluctuations (Figure 1c) are likely to indicate coloured noise, i.e. a signal that contains frequencies of different amplitudes and its frequency spectrum is not flat, but it is not possible to identify which of the two is the noisy instrument.

However, if three collocated instruments A, B and C are available, there can be formed three time series of differences between two instruments, A-B, B-C, A-C. These three time series, corrected for their central tendency value, can be plotted as a function of time. If instrument C is characterized by high coloured noise, this noise will characterise time series B-C, A-C, but not the third one (A-B), as is schematically shown in Fig 1(d-f). Hence, it will be possible to identify an instrument with possible malfunction/high noise, and then analyse its noise from Figure 1e or 1f.

A requirement for such a study is that the instruments are fully collocated, so that they are affected by identical meteorological and the same long-period hydrodynamic effects; this means that they should be located in the same harbour. This is a main difference from most previous studies.

In our study we examined the global record of GLOSS tide-gauges from the Intergovernmental Oceanographic Commission (IOC) website in order to find two or three fully collocated tide gauges covering the same time interval. Then we examined the whole available time records to identify time intervals fully covered with data, in which there is evidence of important differences between them. The diverging parts of time series were then analysed in order to identify their statistical characteristics and significance.

Because the number of harbours with fully collocated tides gauges is small relative to the total number of tide gauges, even a very small percentage of collocated tide gauges characterized by diverging records is very significant, and implies that errors and biases are very frequent, but they were not possible to be recognised so far.
2.1. DATA COLLECTION

For our study we focused on tide-gauges that are part of the Global Sea Level Observing System (GLOSS). The GLOSS station data used were obtained from the IOC website (http://www.ioc-sealevelmonitoring.org). Among the available 300 GLOSS stations worldwide, we identified about a dozen harbours with two or more collocated tide gauges. Ten stations were chosen for this study (Figure 2). The choice of the stations was based on the following criteria which ensure reliable results:

1) complete high-rate records (at least 0.2 measurements per minute) for at least two collocated electronic sensors simultaneously operating and available for at least one year.
2) each sensor record should contain no more than 3.5% missing values in total.
3) only absolute values (raw data) were examined with a common reference time (GMT), excluding records already corrected (adjusted) for errors.
From the study of the available data set, 3 sets containing 3 collocated tide-gauge sensors and 7 sets containing 2 collocated tides-gauge sensors were found to satisfy the specifications of our study and are summarised in Table 1. Sensors are mostly of pressure (PRS) and radar type (RAD), and only in one case, of floater type with encoder (ENC).

A preliminary investigation of the differences between the collocated sensors was made, and this led to search for metadata and to specific investigations (for example, inquiries to tide-gauge or data basis curators) in order to verify and understand the causes of differences between records (possibility of an atypical behaviour to reflect installation of a new instrument at a different level, etc.). No detailed response was possible, but what is clear, is that the policy of IOC is to check data in time due and automatically discard or repair sections of time series, so that final records are free of errors and of significant inconsistencies. This policy, however, reduces the possibility to study uncertainties and biases in raw data. For this reason, our study is based on raw data, i.e. absolute values as recorded at each sensor, without any smoothing or filtering applied. The only exceptions of filtering are the removal of certain high-frequency outliers, as noticed below, and downsampling of certain records so that compatible records for each station are formed (for Zanzibar station only, see Table 1 below).

![Map of GLOSS sea level station locations worldwide. The stations used in this study are marked with black dots. The green and red dots correspond to online and offline stations, respectively (map modified from http://www.ioc-sealevelmonitoring.org/map.php, station status as shown on 18/03/2018 9:45 am GMT).](image)
Table 1: Summary of identified and analysed stations with collocated tide gauges

<table>
<thead>
<tr>
<th>(Station Code) Country - Location</th>
<th>Type of collocated sensors*</th>
<th>Sampling rate mpm (measurements per minute)</th>
<th>Time period covered and missing values (in days and %):</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Acaj) El Salvador - Acajutla</td>
<td>PRS, RAD</td>
<td>1</td>
<td>28/08/2010 – 09/11/2011 (439, 0.03%)</td>
</tr>
<tr>
<td>(Acap) Mexico - Acapulco</td>
<td>PRS, RAD</td>
<td>1</td>
<td>01/10/2009 – 25/04/2011 (572, 1%)</td>
</tr>
<tr>
<td>(Aden) Yemen - Aden</td>
<td>PRS1, PRS2, RAD</td>
<td>1</td>
<td>09/10/2007 – 14/11/2011 (1498, 0.83%)</td>
</tr>
<tr>
<td>(Alex) Egypt - Alexandria</td>
<td>PRS, RAD</td>
<td>1</td>
<td>04/11/2009 – 31/12/2011 (788, 1.38%)</td>
</tr>
<tr>
<td>(Kaum) U.S.A. - Kaumalapau, Hawaii</td>
<td>PRS, RAD</td>
<td>1</td>
<td>27/06/2008 – 06/10/2011 (1197, 0.4%)</td>
</tr>
<tr>
<td>(Manz) Mexico - Manzanillo</td>
<td>PRS, RAD</td>
<td>1</td>
<td>12/06/2008 – 25/11/2011 (1262, 3.4%)</td>
</tr>
<tr>
<td>(Ptnca) Dominican Republic - Punta Cana</td>
<td>PRS, RAD1, RAD2</td>
<td>1</td>
<td>24/06/2010 – 23/11/2011 (518, 0.11%)</td>
</tr>
<tr>
<td>(Setm) Bahamas - Settlement</td>
<td>PRS, RAD</td>
<td>1</td>
<td>23/03/2009 – 30/11/2011 (983, 0.5%)</td>
</tr>
<tr>
<td>(Fren) U.S.A. - Tern, Fr. Frigate</td>
<td>PRS, RAD</td>
<td>1</td>
<td>05/06/2008 – 30/11/2011 (1274, 1.42%)</td>
</tr>
<tr>
<td>(Zanz) Tanzania - Zanzibar</td>
<td>ENC, PRS, RAD</td>
<td>PRS: 1 ENC: 0.2 RAD: 0.3</td>
<td>01/05/2009 – 31/12/2011 (975, 0.07%)</td>
</tr>
</tbody>
</table>

*PRS: pressure sensor, RAD: radar sensor, ENC: floater with encoder. Where more than one sensors of the same type are found at the same location, a number is added next to the sensor type, e.g. RAD1 and RAD2.

2.2. ANALYSIS STEPS

Our methodology to provide datasets that are suitable for further evaluation can be summarised in the following steps:

2.2.1. **Step 1: Homogenisation (Correction for datum differences)**

Different sensors in the same harbour may refer to different datum (reference heights), differing by up to several meters. In order to compare such collocated sensors, their corresponding records should be homogenised into records with a common datum. This is not easy with the available data, because the constituents of the astronomical tide require data covering at least 19 years to be resolved and the “true” zero level to be derived. Therefore, we followed a different approach: Mean values of yearly records were approximated by yearly medians. First, any gross errors were removed, i.e. single, isolated data points significantly different to their neighbouring values, and the median of the available yearly time series was computed. This is because the median is less affected by the
presence of any extreme values, which may represent instrument malfunction or problems in the data logging system, and are in the focus of our analysis.

The following mathematical expression describes the homogenisation of the raw data (collocated time series for each station/harbour) into time series referring to a common datum with an acceptable precision.

\[ X_{i,j,s}^* = X_{i,j,s} - \text{median}_{i,j,s} \]  

(eq.1)

where \( X^* \) represents the homogenised values, \( X_{i,j,s} \) the raw data and \( \text{median}_{i,j,s} \) the median value in year \( i \), from sensor \( s \) at location \( j \), for \( i=1,...,K \) year, \( j=1,...,L \) location, and \( s=1, ...S \) collocated sensors.

2.2.2. **Step 2: Calculation of the difference (bias) between collocated sensors**

With the values of all collocated sensors fluctuating around zero after Step 1, the difference between two collocated sensors \( s_1 \) and \( s_2 \) at time \( t \) can be calculated using eq. 2:

\[ \Delta X_{s_1s_2t} = X_{s_1t}^* - X_{s_2t}^* \]  

(eq. 2)

where \( X_{s_1t}^* \) and \( X_{s_2t}^* \) are the homogenised values at time \( t \) as calculated by eq. 1 for sensor \( s_1 \) and \( s_2 \), respectively. The time refers to the full duration of the data set (one year to four years). This difference represents the bias in the sea level values.

An estimate of the mean difference of the two time series is provided by the root mean square (RMS) of the differences for each pair of collocated sensors at station \( L \)

\[ \text{RMS}_{\Delta X_{s_1s_2t}} = \sqrt{\frac{\Delta X_{s_1s_2t}^2}{n}} \]  

(eq. 3)

with \( \Delta X_{s_1s_2t} \) calculated by eq. 2 and \( n \) the total number of available pairs of observations at a specific location.

2.2.3. **Step 3: Analysis of the calculated differences between collocated sensors**

The differences between collocated sensors were subsequently analysed to determine (1) the range of the bias at each station and (2) the type of this bias regarding its amplitude ( > 15 cm for high amplitude differences) and its pattern, so that it could be categorised into one or more types, e.g. drift at one or more collocated sensors at the same station, sensor or data logging system malfunction, periodic behaviour etc. The possibility of malfunction of a sensor
or the data logging system was investigated, using also the raw recordings for each sensor and location.

Additional analyses were made necessary at a following step in order to evaluate the results obtained.

3. RESULTS

3.1. RANGE OF BIAS

Following the steps described in the previous section, we calculated the differences between pairs of collocated sensors for all available records and stations in Table 1 and the corresponding RMS. These results are summarized in Table 2. According to Woodworth and Smith (2003), an RMS value of 1.4 cm sets the accuracy of each sensor at 1 cm or less. The highlighted row in Table 2 corresponds to the only station and pairs of sensors among the available data which were found to satisfy this condition. The largest ranges and RMS errors correspond to station Zanzibar with the differences having a max value of 4 m and the RMS being approximately 58 cm. The smallest differences are observed at station Punta Cana with values up to 10 cm. This station exhibits the smallest RMS values (< 2 cm).

In summary, over one billion pairs of values between collocated sensors have been used in total for our analysis. The vast majority of the differences, i.e. 80%, fall within ± 15 cm, and approximately 19.6% of the differences are extreme with an absolute value > 30 cm.

Table 2. Range of bias and corresponding root mean square (RMS) values at each station.

<table>
<thead>
<tr>
<th>Station (station code)</th>
<th>Types of sensors</th>
<th>Range of bias (cm)</th>
<th>RMS of bias (cm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acajutla (Acaj)</td>
<td>PRS-RAD</td>
<td>-25 to 50</td>
<td>9.5</td>
</tr>
<tr>
<td>Acapulco (Acap)</td>
<td>PRS-RAD</td>
<td>-15 to 10</td>
<td>3.2</td>
</tr>
<tr>
<td>Aden (Aden)</td>
<td>PRS1-PRS2</td>
<td>-30 to 90</td>
<td>14.7</td>
</tr>
<tr>
<td>Aden (Aden)</td>
<td>PRS1-RAD</td>
<td>-90 to 20</td>
<td>14.4</td>
</tr>
<tr>
<td>Aden (Aden)</td>
<td>PRS2-RAD</td>
<td>-20 to 20</td>
<td>5.8</td>
</tr>
<tr>
<td>Alexandria (Alex)</td>
<td>PRS-RAD</td>
<td>-55 to 55</td>
<td>28.9</td>
</tr>
<tr>
<td>Hawaii (Kyam)</td>
<td>PRS-RAD</td>
<td>-10 to 70</td>
<td>40.5</td>
</tr>
<tr>
<td>Manzanillo (Manz)</td>
<td>PRS-RAD</td>
<td>-100 to 100</td>
<td>30</td>
</tr>
<tr>
<td>Punta Cana (Ptca)</td>
<td>PRS-RAD1</td>
<td>-10 to 10</td>
<td>1.9</td>
</tr>
<tr>
<td>Punta Cana (Ptca)</td>
<td>PRS-RAD2</td>
<td>-5 to 5</td>
<td>1.1</td>
</tr>
<tr>
<td>Punta Cana (Ptca)</td>
<td>RAD2-RAD1</td>
<td>-10 to 10</td>
<td>1.8</td>
</tr>
<tr>
<td>Settlement (Setm)</td>
<td>PRS-RAD</td>
<td>-10 to 10</td>
<td>2.2</td>
</tr>
<tr>
<td>Tern (Fren)</td>
<td>PRS-RAD</td>
<td>-50 to 70</td>
<td>16.4</td>
</tr>
<tr>
<td>Zanzibar (Zanz)</td>
<td>PRS-RAD</td>
<td>-50 to 400</td>
<td>58</td>
</tr>
<tr>
<td>Zanzibar (Zanz)</td>
<td>PRS-ENC</td>
<td>-100 to 400</td>
<td>55</td>
</tr>
</tbody>
</table>
3.2 TYPES OF BIAS

The characteristics of the differences between pairs of collocated sensors at each station, fall into one of the following categories:

3.2.1 Drift of one or more sensors (evidence of a linear trend)

Drift of one or more sensors, as evidenced by the presence of a linear trend (Míguez et al., 2012), has been observed at stations Hawaii and Tern. Both stations had the same type of collocated sensors, i.e. PRS and RAD. Figure 3 shows the differences between pairs of collocated sensors for the two stations above. The scale of the x and y axes (indicating differences between two sensors versus time) is the same for all graphs. Both stations show evidence of a linear trend in the differences for at least part of the time period examined, with the estimated slope value being statistically significant at 95% confidence level.

Table 3 presents the values of slope b of the equation

\[ y = a + b \times x \]  

(eq. 4)

where \( y \) represents the differences between sensors, \( x \) the time, \( a \) and \( b \) variables to be determined with \( b \) representing the linear trend, fitted to the differences of each of the two stations shown in Figure 3. It should be noted that the linear fit was applied to the daily averages of the differences instead of the values at the sampling interval of 1 minute. The fit was based on least squares. This method is sensitive to the presence of extreme values (such as those representing possible sensor malfunction as is shown in 3.2.2). The use of daily averages resulted in a smoothed time series with reduced white noise, a better condition for the estimation of the slope. The presence of an almost one year gap in the time series of Tern, due to missing data, does not affect the fit.
Figure 3. Daily average values of the differences between collocated sensors PRS and RAD for Tern, U.S.A (top) and Kaumaulau, Hawaii (bottom). The highlighted areas indicate the time period and data to which a linear trend was fitted using least-squares.

Table 3. Slope of linear fit to the differences at stations Tern and Hawaii.

<table>
<thead>
<tr>
<th>Station</th>
<th>Time period accounted for in linear regression</th>
<th>Slope (b) (cm/day) ± st.error</th>
<th>(R^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tern</td>
<td>June 2008 – January 2011</td>
<td>-0.022 ± 0.0005</td>
<td>0.78</td>
</tr>
<tr>
<td>Hawaii</td>
<td>July 2008 – April 2010</td>
<td>-0.017 ± 0.0002</td>
<td>0.92</td>
</tr>
</tbody>
</table>

3.2.2. Sensor or data logging system malfunction

Sensors in all locations examined had some anomalies in their records. Such anomalies are generally treated as outliers and are discarded. A closer examination of the raw data revealed that such anomalies rose at instances when a zero value had been logged for one of the sensors. Such instances were observed in the data for both types of sensors, i.e. RRS and RAD, and are more frequent in PRS. For stations Aden and Zanzibar in particular, a larger number of extreme values, i.e. anomalies, were observed at the differences between collocated sensors (Figure 4b and 5b) when compared to the other stations. These anomalies were very frequent during the whole examined time period.
Figure 4. Station Aden. (a) Raw data (absolute values) as recorded at sensors PRS1, PRS2 and RAD. (b) Differences between collocated sensors after the removal of the yearly median. Values mostly fall within ±10 cm. The anomalies (extreme values) observed throughout the examined period are due to either malfunction of the data logging system and/or the fact that
sensor PRS1 could not record data with values below the threshold of 0.902 m (as indicated by the dashed black line).

At Aden, there was clearly a problem with the data logging system. Between February and July 2008 and March to May 2011, a series of zero values were recorded at all sensors (see Figure 4a) mostly at the same time. Where zero values were recorded at one or two sensors only, this appears as an extreme peak in the residuals (Figure 4b). In Figure 4b, positive extreme values are due to sensor PRS2 for the top plot and sensor RAD for the last two plots recording zero values. Negative extreme values are due to sensor PRS1 for the top and bottom plots and PRS2 for the plot in the middle, respectively due to zero values. It becomes evident that the two pressure gauges experience data logging problems more frequently. In addition, sensor PRS1 appears to experience a clipping effect, i.e. did not record any values smaller than 0.902 m (threshold indicated by dashed black line on PRS1 plot in Figure 4a) for the whole time period examined. This appears as an apparent periodicity in the homogenised data (Figure 4b, top and bottom plot).

Extensive data logging problems in the form of zero values in the recordings of all sensors were also observed at Zanzibar station (Figure 5a). This can be clearly seen for RAD and ENC sensors during September - October 2010, March - May 2011 and September – December 2011. On the other hand, during those periods, sensor PRS does not appear to experience the same problem. However, a gradual increase in its sea level recordings can be observed (‘hogging’ effect). A gradual significant increase was also observed at station Tern (see Figure 3) and a gradual decrease (‘sagging’ effect) at station Acajutla.

Homogenised data at station Zanzibar also exhibit a periodic behaviour. Possible periodicities in the differences of the recordings between collocated sensors are discussed in section 3.2.4.
Figure 5. Station Zanzibar. (a) Raw data as recorded at sensors RAD, PRS and ENC without any processing. (b) Differences between collocated sensors after removal of the median. Values fall mainly within ±10 cm up till approximately September 2011 with the exception of the differences where sensor PRS is involved.

3.2.3. Sudden ‘step’ in the recordings
In three cases, a sharp step (offset) appeared in the time series of the differences. This was the case for stations Hawaii, Manzanillo and Alexandria. The range in the amplitude of such steps was 50 cm for Hawaii (between April and May 2010, Figure 6a), 6 m for Manzanillo (September 2008 and September 2009, Figure 6b) and 40 cm for Alexandria (July 2010, Figure 6c). The lasting duration of this anomaly was 6 hours for Hawaii, 6.5 months and 1 year for Alexandria and Manzanillo, respectively. 

For Hawaii, the study of the raw sea level records (at 1 minute sampling interval) indicated that this step is due to the RAD sensor. There was a gap of 6 hours in the data for both sensors, then the datum for the RAD sensor appears to have changed by approximately 0.5 m (Figure 7). Apart from this, no other changes are observed, i.e. same range of amplitude, same waveform characteristics.

![Figure 6. Sudden 'step' in the differences between collocated sensors PRS and RAD at stations (a) Hawaii, (b) Manzanillo and (c) Alexandria.](image)

In contrast to Hawaii, at Alexandria station the jump happened suddenly, between consecutive 1 minute observations. The anomaly in the differences seen in Figure 6c is due to sensor PRS as shown in more detail in Figure 8. PRS station exhibits an unstable behaviour with significant fluctuations over the whole time period examined and a number of zero readings indicating malfunction. Three jumps can be seen in the recordings during a 5-day period. The first, on 13\(^{th}\) June, was approximately 0.4 m, the second, on 16\(^{th}\) June was 0.2 m and the third, on 17\(^{th}\) June, was 1 m. In the time period between these changes, the range of
the waveform amplitudes also changed slightly. No differences in the frequency characteristics were found.

At station Manzanillo the jump in the differences is also due to sensor PRS. Figure 9 shows the start of the jump in the absolute sea level recordings (sampling interval = 1 minute) for PRS. It can be seen that not only there is an abrupt datum change of approximately 5 m but also the waveform is very different before and after the jump, both in the frequency and time domain.

Figure 7. Raw sea level data as recorded by sensor RAD at Hawaii (top) and zoom (bottom graph) on the sharp change (highlighted area in the top graph) in the recordings. In April 2010, what most probably looks like a datum change for sensor RAD, probably due to maintenance, can be clearly seen.
Figure 8. Raw sea level data as recorded by sensor PRS at Alexandria station (top) and zoom (bottom graph) on the sharp change (highlighted area in the top graph) in the recordings. During a five-day period, three sharp changes, each occurring between consecutive samples, are visible.

Figure 9. Raw sea level data as recorded by sensor PRS at Manzanillo (top) and zoom (bottom graph) at the sharp change (highlighted area in the top graph) in the recordings. After the jump, the recorded waveform is very different compared with the time period before the jump both in amplitude and frequency.

3.2.4 Evidence of periodicity
Spectral analysis of the differences calculated at Step 2 was performed for those stations for which the residuals were visually deemed to exhibit periodic behaviour. The aim of this type of analysis was to determine whether there were any periodical effects in the residuals, and if so, what their values were. Periodical effects could represent systematic errors in the recordings of the tide-gauges.

Due to the large number of consecutive missing values in some of the available data sets, the commonly used spectral analysis method of Fourier Transforms was not applicable (Press et al., 2007). In order to overcome problems with gaps in the data, and to avoid using any interpolations that could introduce further bias, the Lomb normalised periodogram was used instead. A detailed description of the method is provided by Press et al. (2007). The analysis was performed using the Normperiod code (Pytharouli and Stiros, 2008).
Among the ten locations examined, stations Settlement and Zanzibar showed evidence of a periodic behaviour in the differences in the sea level between the pairs of collocated sensors. This appears as periodic fluctuations in the time series of the differences, as shown in Figure 5b RAD-ENC and PRS-ENC plots during 2009 for example. For both locations, the periodicity is observed in all pairs of collocated sensors and in their differences. A spectral analysis of the residuals at both stations revealed no statistically significant periods other to those corresponding to some of the main tidal constituents. It should be noted here that the gradient evident in both spectra of Figure 10 does not represent coloured noise, but is the result of the logarithmic scale used for both x and y axes for illustration purposes only. A linear scale in the x axis would have resulted in a nearly horizontal spectrum but for the dominant peaks.

Figure 10. Lomb normalised spectra for (a) Zanzibar and (b) Settlement stations. The period resolution of the two spectra is different due to the different length of the data sets used and their sampling rates. Some of the main tidal constituents are visible in both spectra. The scale in both x and y axes is logarithmic. The dashed line indicates the threshold above which any peaks are statistically significant at 99.9% confidence level.

4. DISCUSSION

In this study we analysed collocated tide gauge IOC records from 10 locations (stations) with data satisfying certain strict criteria specified in section 2.1. Each station had at least two different types of sensors available, the most common pair being a pressure (PRS) and a radar (RAD) sensor. Our analysis did not take into account the particular model of sensor, but focused instead on their general type (i.e. PRS, RAD, ENC). One of the aims was to determine
whether specific types of sensors are more prone to bias. To our knowledge, this is the first time that a systematic comparison between different types of tide gauges was made at global scale. This was made possible by comparing the data from sensors that are fully collocated, usually within a few metres from each other, and hence they record the same process without local perturbation. This section discusses our results, their significance and the implications for coastal engineering projects and earth processes studies.

4.1 AMPLITUDE OF BIAS

The differences between collocated sensors were found as high as 4 m, with the majority of the differences found to be within 15 cm, while only one pair of sensors had an RMS below 1.4 cm. This last value is suggested as the threshold below which it is ensured that each of the sensors has an accuracy better than 1 cm (Woodworth and Smith, 2003). Amongst the ten stations examined, this value was only reached at one station: Punta Cana and only two of the three sensors at this station satisfied this criterion (see Table 2).

4.2 VAN DE CASTEELE TEST (OR TEST FOR THE PRESENCE OF SYSTEMATIC ERRORS)

The differences presented did not account for potential scale errors or clock drifts between the pairs. The latter was rather unlikely, as the available data were synchronised. Nevertheless, we performed a Van de Casteele test (IOC, 1985; Miguez et al., 2008) for the pairs of collocated sensors at stations that exhibited extreme bias and RMS values (> 30cm). These stations were Hawaii, Manzanillo and Zanzibar (see also Table 2). This test permits to visualise the differences of recorded tidal levels (ΔH) between a sensor and another selected as reference. The test requires data that span a period of at least as long as a full tide cycle. Following results of previous studies (e.g. Miguez et al. 2012, Pérez et al. 2014) which have suggested that the radar sensors appear to have a more stable behaviour over the pressure gauges, we used the radar sensors (RAD) as the reference gauge and the pressure gauge (PRS) or floater with encoder (ENC) as the sensor under testing. The duration of the data used for the test was approximately one month. This month (Zanzibar: 15/5/2009 – 14/5/2009; Settlement: 23/3/2009 – 22/4/2009; Hawaii: 27/01/2009 – 26/2/2009; Manzanillo: 19/7/2008 – 18/7/2008) was chosen within a time period free of any extreme values. The Van de Casteele plots permit to understand whether the difference between the two sensors
indicates white noise (compare with Figure 1a), time shift errors, instrument/system malfunction.
This test was applied to the pairs of collocated sensors at Hawaii, Manzanillo and Zanzibar, as well as Settlement station. The latter was used as a station with just white noise. Representative results from Zanzibar and Settlement stations are summarized in Figure 11. The test revealed that the differences of all examined locations are not due to time shift errors. This was the case at all but one locations (Zanzibar). Figure 11 shows the Van de Casteele plots for the differences of three pairs of collocated sensors (Figure 11a, b and c) at Zanzibar and one pair (Figure 11d) at Settlement station. The latter was used as an example of a station free of time shift errors. The resulting shapes of the Van de Casteele plots are characteristic of specific conditions. An ellipse such as that in Figure 11a, b and c is typical of a time shift between the sensors under comparison (Miguez et al., 2008). On the other hand, a straight column centred around zero as that in Figure 11d represents only instrumental noise (Miguez et al., 2008). In Figure 11b there is evidence of one more source of errors. The ellipses are not smooth but appear to have clear periodic oscillations which are representative of malfunctioning on one or both sensors (Miguez et al., 2008). The results of the Van de Casteele test for Hawaii and Manzanilo indicated different fault mechanisms. For Hawaii, the test indicated instrument malfunction while for Manzanillo a scale error (linear slope in the diagram).

Figure 11. Van de Casteele tests (i.e. plots of the differences of recorded sea levels (ΔH) between a sensor under testing and another sensor acting as the reference sensor versus the
recorded sea levels $H$ from the sensor under testing) for selected stations: Zanzibar (a-c) and Settlement (d) stations. (a) sensor under testing: PRS, reference station: RAD. (b) sensor under testing: ENC, reference station: PRS. (c) sensor under testing: ENC, reference station: RAD. (d) sensor under testing: PRS, reference station: RAD.

4.3 SCALE ERRORS

Scale errors can be determined by linear regression analysis between the recordings of pairs of collocated sensors. One type of sensors is used as the dependent variable and the other as the independent. In this case, the slope $b$ of the fitted line (eq. 4) can be used for the calculation of the scale error $\varepsilon$, as (Pérez et al., 2014):

$$\varepsilon = (b-1)x100$$  \hspace{1cm} (eq. 5)

Because our aim was to examine which type of sensors was more prone to any type of bias, no prior filtering was applied to the data, apart from removal of certain high-frequency outliers, and downsampling of the Zanzibar station records so that compatible records were formed. As a result, some of the data sets did not allow for linear regression analysis based on least-squares. The analysis was performed for all pairs of PRS and RAD sensors at six stations out of the ten available. Four stations: Tern, Alexandria, Manzanillo, Zanzibar, had a large number of extreme values, hence, they were not included in this analysis. Table 4 shows the results.

**Table 4.** Slope $b$ of linear regression between pairs of PRS and RAD sensors (based on data sets with sampling interval of 1 minute) and calculated scale error. PRS was taken as independent variable and RAD as the dependent.

<table>
<thead>
<tr>
<th>Station</th>
<th>X/Y</th>
<th>Slope, $b$</th>
<th>Scale error, $\varepsilon$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acajulta</td>
<td>PRS/RAD</td>
<td>$0.9735 \pm 0.0002$</td>
<td>-2.65</td>
</tr>
<tr>
<td>Acapulco</td>
<td>PRS/RAD</td>
<td>$1.0165 \pm 0.0002$</td>
<td>1.65</td>
</tr>
<tr>
<td>Aden</td>
<td>PRS2/RAD</td>
<td>$0.9725 \pm 0.0001$</td>
<td>-2.75</td>
</tr>
<tr>
<td>Hawaii</td>
<td>PRS/RAD</td>
<td>$0.9949 \pm 0.0003$</td>
<td>-0.51</td>
</tr>
<tr>
<td>Punta Cana</td>
<td>PRS/RAD1</td>
<td>$0.9840 \pm 0.0001$</td>
<td>-1.6</td>
</tr>
<tr>
<td>Punta Cana</td>
<td>PRS/RAD2</td>
<td>$0.9714 \pm 0.0002$</td>
<td>-2.86</td>
</tr>
<tr>
<td>Settlement</td>
<td>PRS/RAD</td>
<td>$0.9932 \pm 0.0001$</td>
<td>0.68</td>
</tr>
</tbody>
</table>
Table 4 reveals systematic scale differences (errors) between the PRS and RAD sensors, with the later mostly recording lower tidal ranges. In the past, scale errors of the same order of magnitude have been reported (e.g. Pérez et al., 2014), while in another study pressure gauges have been associated with lower values, probably because of bias due salinity etc effects (Miguez et al., 2005).

4.4 DRIFT OR DATUM SHIFT

For two stations, Tern and Hawaii (see Figure 3), a linear trend was observed in the differences between collocated sensors. In the international literature, such linear trends are interpreted as changes in the reference datum or a drift of one of the sensors, e.g. Míguez et al. (2012), Pérez et al. (2014). The rate of change was estimated (absolute values) between 0.017 cm/day (62 mm/yr) and 0.022 cm/day (80 mm/yr) as shown in Table 3. All slope values computed in the analysis were significantly different from zero. Such a rate of drift is small, but when considered over a period of a year, it becomes significant. If uncorrected, it could bias the long-term trends in climate change and geological processes studies (Pérez et al., 2014).

4.5 JUMP IN THE RECORDINGS

As can be derived from Figure 6, in three out of ten stations, a sharp change in the differences between collocated sensors was observed on both types of sensors, i.e. PRS and RAD, although the PRS sensors were those affected more (in two out of the three locations). Such sharp changes had a different effect for each sensor. For the radar sensor in this study the jump is more likely associated with a change in the datum (Figure 7), most likely a result of instrument maintenance/calibration as it appeared after a gap of 6 hours in the recordings. For the pressure gauges, the jump could also indicate instrument malfunction (Figure 8 and 9).

4.5 SAGGING OR HOGGING

When a gradual but significant change was observed at three stations, our analysis showed that in all three cases, it was the pressure gauges that had caused this change. Such gradual
fluctuations might indicate changes in atmospheric pressure and temperature, factors that may cause significant bias in pressure gauges, if not taken into account in the transformation of readings to sea level values (e.g. Mehra et al., 2009). This could be the case for the three stations in this study, since we examined raw, uncorrected data.

4.6 PERIODIC BEHAVIOUR

Our analysis also revealed periodic components at two stations, Zanzibar and Settlement (Figure 10). Spectral analysis of the differences between collocated sensors at each location did not reveal periodicities other than those corresponding to some main tidal constituents. The absence of additional data sets e.g. meteorological data, makes it very difficult to suggest with confidence what could be affecting the pressure gauges or if there is a correlation between the ‘biased’ values and sea surges. However, a number of studies have highlighted the fact that pressure gauges are significantly affected by weather conditions (e.g. Woodworth and Smith, 2003; IOC, 2016; Mehra et al., 2009) and this could well be the case for this study.

4.7 SUMMARY OF FINDINGS

Table 5 summarises the findings of the present analysis. The bias does not include effects such as instrumental jolts that might appear at very short time intervals (shorter than a couple of hours). Entry ‘1 or 2’ in the column ‘Number of sensors affected’ in the line ‘Datum change’ indicates that it was not possible to identify which sensor (PRS or RAD) within each pair was drifting at the two locations involved (Tern and Hawaii). Therefore, at both stations, both sensors PRS and RAD have the same likelihood to be the sensor that is drifting relative to the other. The ENC sensor is not included in Table 5 because we only had one record available and any generalisations of the results would not be precise.

Table 5. Summary of bias characteristics from this study.

<table>
<thead>
<tr>
<th>Type of bias</th>
<th>Evidenced by</th>
<th>Amplitude of bias</th>
<th>Number of sensors affected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>PRS (out of 11)</td>
</tr>
<tr>
<td>sharp offsets</td>
<td>Up to 4 m</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Instrument/logging system malfunction</td>
<td>frequency (waveform) changes</td>
<td>-</td>
<td>1</td>
</tr>
<tr>
<td>--------------------------------------</td>
<td>-----------------------------</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>Datum change/drift</td>
<td>Linear trend</td>
<td>Up to 80 mm/yr</td>
<td>1 or 2</td>
</tr>
<tr>
<td>Response to meteorological factors</td>
<td>Gradual change over a time period – ‘hogging’ or ‘sagging’</td>
<td>Up to 1.2 m</td>
<td>3</td>
</tr>
<tr>
<td>Other</td>
<td>Periodicity</td>
<td>-</td>
<td>2</td>
</tr>
</tbody>
</table>

### 4.8 OVERPASSING THE NOISE PROBLEM

In the last years the number of harbours with two or more collocated tide-gauges, especially sensors with on-line data communication system, is increasing, because of the reduction of both the sensor and the maintenance cost and because of the new data communication facilities. This trend is expected to continue, and in the near future it is expected that in most harbours/locations there will be available several collocated tidal sensors. Hence, it is expected that it will be possible to compare and identify differences in sensor recordings, or even to identify faulty sensors easily.

However, care should be taken so that all sensors are fully collocated or exposed to similar environmental conditions, especially concerning the hydraulic load. This is because in main, composite harbours confluence of waves may lead for example to amplification or attenuation of the amplitude of the oscillations of the water level and to “true” tidal local differences (semi-static tidal level) which may not be representative of the “true” characteristics of the tide.

In addition, in order to control and assess in a statistical way the sensors’ output, at least in some selected sites, there should be installed and analysed sets of identical collocated sensors with a common sampling and data logging system to avoid additional noise. This will permit to compare the output. This last task, however, is not easy, because errors usually found in raw tidal data are removed and not displayed in data bases. Such modification tends to shadow important metrological information. On the contrary, raw and corrected data sets should be fully shown, as is the case in disciplines, for example in data bases used in Climatology, both raw and transformed data are both displayed and permit to understand errors and correct errors (“homogenize” data) using different approaches; for example in Global Historical Climatology Network – Monthly Version 3 (GHCN – M), available at [ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/v3/](ftp://ftp.ncdc.noaa.gov/pub/data/ghcn/v3/)
5. CONCLUSIONS
An overall conclusion is that errors and bias were found in a small portion of tide-gauges among those existing in a global scale. The global implications of these results are high, because noise was found in nearly all cases in which data from collocated sensors permitting a systematic and reliable analysis exist. This is likely to imply that tide-gauge errors are more frequent and important than was believed before, but they cannot be identified in isolated sensors. This result may have important implications because tide gauges are used in various studies, from climatic to geotechnical engineering.

However, the trend to increase the number of collocated sensors in the last years is expected to continue and to offer redundant data to assess data quality and significance, especially through systematic study of collocated sensors operating under similar environmental conditions in selected sites. Furthermore, data bases containing both raw and corrected tidal records (and not only corrected records), as is the case in other disciplines, are necessary.

6. FUNDING SOURCES
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

7. ACKNOWLEDGEMENTS
This paper has benefited from the comments and suggestions of two anonymous reviewers.

8. REFERENCES


