Ocean Dynamics (2015) 65:509–521
DOI 10.1007/s10236-015-0824-z

A new hybrid model for filling gaps and forecast in sea level: application to the eastern English Channel and the North Atlantic Sea (western France)

Imen Turki · Benoit Laignel · Nabil Kakeh · Laetitia Chevalier · Stephane Costa

Received: 2 July 2014 / Accepted: 24 February 2015 / Published online: 12 March 2015

Abstract This research is carried out in the framework of the program Surface Water and Ocean Topography (SWOT) which is a partnership between NASA and CNES. Here, a new hybrid model is implemented for filling gaps and forecasting the hourly sea level variability by combining classical harmonic analyses to high statistical methods to reproduce the deterministic and stochastic processes, respectively. After simulating the mean trend sea level and astronomical tides, the nontidal residual surges are investigated using an autoregressive moving average (ARMA) methods by two ways: (1) applying a purely statistical approach and (2) introducing the SLP in ARMA as a main physical process driving the residual sea level. The new hybrid model is applied to the western Atlantic sea and the eastern English Channel. Using ARMA model and considering the SLP, results show that the hourly sea level observations of gauges with are well reproduced with a root mean square error (RMSE) ranging between 4.5 and 7 cm for 1 to 30 days of gaps and an explained variance more than 80 %. For larger gaps of months, the RMSE reaches 9 cm. The negative and the positive extreme values of sea levels are also well reproduced with a mean explained variance between 70 and 85 %. The statistical behavior of 1-year modeled residual components shows good agreements with observations. The frequency analysis using the discrete wavelet transform illustrate strong correlations between observed and modeled energy spectrum and the bands of variability. Accordingly, the proposed model presents a coherent, simple, and easy tool to estimate the total sea level at timescales from days to months. The ARMA model seems to be more promising for filling gaps and estimating the sea level at larger scales of years by introducing more physical processes driving its stochastic variability.

Keywords Sea level forecast · Astronomical tides · Nontidal residual surges · ARMA · Sea level pressure

Responsible Editor: Birgit Andrea Klein

I. Turki (✉) · B. Laignel · L. Chevalier
UMR CNRS 6143 Continental and Coastal Morphodynamics 'M2C'
University of Rouen, 76821 Mont-Saint-Aignan Cedex, France
e-mail: imen.turki.cnrs@gmail.com

N. Kakeh
Department of Applied Physics, Universitat Politècnica de Catalunya-Barcelona Tech, Barcelona, Spain

S. Costa
University of Caen Low Normandy, Geophen UMR-CNRS LETG, 6554 Normandy, France

1 Introduction

Present-day sea level change is of considerable interest because of its potential impact on human populations living in coastal regions and on islands. The long-term changes in sea level are mainly explained by the global warming (Church et al. 2001). The increase of the global average sea levels was between 1 and 2 mm/year during the twentieth century (Church et al. 2001). Cabanes et al. (2001) have recently shown an acceleration of the long-term trend with a rate of 2.5 mm/year between 1993 and 2000. In the future, global average sea level is expected to increase more rapidly as a result of anthropogenic climate change. The fifth Assessment Report of the Intergovernmental Panel on Climate Change (Church et al. 2013; Stocker et al. 2013) has concluded that the global rise is of 52–98 cm by the year 2100 which would threaten the survival of coastal cities and entire island nations. Processes responsible for sea level changes are complex. To help simplify the matter, it is useful to consider separately the components related to the astronomical oscillations, surges, and eustatic processes (Pugh 1987). Progress in the study of climate and global change depends heavily on the creative use of the sea level data sets. Their measurements rely on tide gauges which are considered one of the most reliable sources. However, one common problem in the sea level records is the presence of gaps (a sequence of missing values or omitted observations) due to the malfunction of gauges under instrumental and/or meteorological conditions. Such problems disrupt and make impossible the use of sea level records for research and practical purposes. Moreover, the forecast of sea level scenarios at different timescales of days to years is needed for coastal engineering studies, operational systems, and practical applications.

According to the previous studies, there are three methods for estimating the sea level: harmonic analyses, dynamical, and statistical. The classical harmonic analyses are specialized techniques able to take advantage of the "deterministic" nature of tidal processes, the dominant component of the total sea level in determined beaches, which is modeled as the sum of a finite set of sinusoids at specific frequencies. For dynamical methods, physical models are used to derive fine scale winds and atmospheric surface pressures from the large-scale climate simulated by the coarse global climate model. For the alternative statistical methods, relationships between meteorology and sea level heights are developed from observations or simulations of the recent past and present day. These relationships must be capable of explaining the sea level variability, and for predictive purposes, the climate changes projections (Von Storch and Reichardt 1997). The classical harmonic analyses can well predict the deterministic process of astronomical tide; however, they are not enough to reproduce the total sea level since surge components are not taken into consideration. Regarding the dynamical methods, simulations require calibrations, large data sets of bathymetry, and boundary conditions. These methods take significant logistic efforts and time for their high computational costs (Flather et al. 1998; Flather and Williams 2000; Lowe et al. 2001). Such requirements are less restrictive, with respect to available acknowledgments, for statistical methods, whose progress are significantly considered, nowadays, for the sea level forecast. Here, commonly used methods for reconstruction of the sea level are based on single linear, multiple linear, and nonlinear regressions. The samplers were also applied (Dergachev et al. 2001; Kondrashov and Ghil 2006; Moffat et al. 2007; Musial et al. 2011). Hilmi et al. (1997) was developed a stochastic model of short-term variations of sea level in the St. Lawrence estuary (Canada) using autoregressive moving average (ARMA) model. An innovative approach of artificial neural networks (ANN) was used by Pashova et al. (2013) as a nonparametric modeling framework for the nonlinear sea level forecasting and filling the missing values in the daily sea level series. However, statistical methods are limited to mathematical concepts and cannot be accurate for the physical forecast of the sea level needs for further analyses. The intent of the present research is to propose a new hybrid model for filling gaps of missing data and forecasting the sea level by combining the classical harmonic analyses with the statistical methods. Here, an autoregressive moving average technique is implemented and applied to the sea level with the aim of filling gaps in hourly time series and forecasting the sea level variability for days to months.

The remainder of this paper is organized as follows. A new hybrid model is developed in section 2 for estimating the total sea level in oceans and coastal areas. The Atlantic sea and the eastern English Channel (western France) are selected in section 3 as a case of study for accurate applications. Modeled results are also presented and validated with observed data. In the last section, the implications of obtained results are discussed. Concluding remarks and further works are finally suggested.

2 Methodology

Coastal water levels are influenced by a variety of astronomical, meteorological/oceanographical, and tectonic factors, the most readily apparent being the tides. At times, these factors interact in a complex way to elevate water levels significantly above normal tide level. Storms, which develop low atmospheric pressure, are the most common cause of elevated water levels. Strong winds and large waves contribute also in the sea surface perturbations in oceans and coastal zones. Therefore, the total sea level (SL) is composed of different components: (1) the mean sea level (MSL), (2) the vertical local movement (VLM) described by the vertical reference datum for water level and related to the subsidence and tectonic deformations of land, (3) the astronomical tide (AT), and (4) the residual surges (S):

$$\mathrm{SL} = \mathrm{VLM} + \mathrm{MSL} + \mathrm{AT} + S \tag{1}$$

The MSL is influenced by longer-term climate fluctuations related to global warming. The VLM results from the glacial isotactic adjustment, tectonics, subsidence, and sedimentation; it changes on decadal and longer timescales and affects the global mean sea level. Both components (MSL and VLM) are gathered to a new component $\mathrm{MSL_V}$ ($\mathrm{MSL_V}$=MSL+VLM) since they are characterized by linear evolutions, specifically at the timescales analyzed in the research. So, Eq. 1 can be expressed as the following:

$$\mathrm{SL} = \mathrm{MSL_V} + \mathrm{AT} + S \tag{2}$$

The AT are directly forced by the gravitational effect of the moon, and to a lesser extent, the sun, and other planets on the water mass of the oceans. Tides, considered as a large signal in the ocean, are characterized by a broad hump with a low-frequency maximum and a decline at higher frequencies. Superimposed components are a number of sharp tidal peaks near diurnal and semidiurnal frequencies, and sometimes a broader peak associated with Coriolis or inertial effects. They move from the deep ocean to coastal shallow water as a wave or a combination of waves. The residual surges *S* are defined as the difference between the relative observed water level and the AT. The dynamical analysis of the sea level requires the separation of the deterministic signal of the AT from the stochastic behavior of the residual surges *S* (Tawn and Vassie 1989). The main purpose of the decomposition of the tide gauge records is the timescales of variability. In this way, the specific circumstances of each component fluctuation can be clearly observed, in order to provide the conditions for the in-depth study of its variation.

The methodology developed in this research aims to forecast the hourly SL during days, months, and years by combining different types of models: (1) a linear regression for the long-term component of MSL_V, (2) a deterministic model of harmonics for AT, and (3) a stochastic model of ARMA for the residual surges *S*. An overview of the proposed methodology is displayed in Fig. 1.

2.1 The sea level trend

The MSL_V trend was estimated by a linear regression of the mean annual values (Eq. 3), calculated from the hourly time series of sea level. These mean annual values were filtered

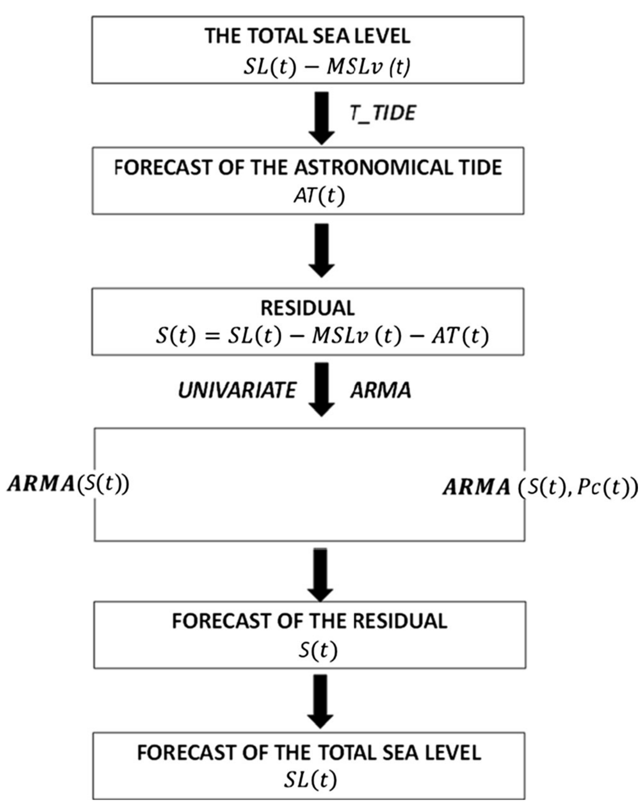

Fig. 1 An overview the proposed methodology for the sea level forecast

from years with less than 80 % of data and the outliers out of the 95th percentile.

$$MSL_V(t) = a + b.t \tag{3}$$

where *a* and *b* are constants.

2.2 The astronomical tides

The difference between the total SL and MSL_V, previously evaluated, represent both harmonic (AT) and residual (S) components (Eq. 2). The separation between tidal and nontidal energy is an important task in any analysis of oceanic time series. In classical harmonic analysis, the tidal forcing is modeled as a set of spectral lines, i.e., the sum of a finite set of sinusoids at specific frequencies. These frequencies are a combination of six fundamental frequencies arising from planetary motions (Godin 1972). Many of the more important frequencies, known also as harmonic components, have names such as "M2," "K1", etc. Classical harmonic analysis were used for the prediction of astronomical tides by the determination of the resulting phase/amplitude of each given sinusoid in a determinist way. This analysis is based on a series of algorithms and *FOTRAN* codes (Godin 1972; Foreman 1977, 1978) through a *MATLAB* package known as *T-TIDE* (Pawlowicz et al. 2002). The tidal response was modeled as:

$$AT(t) = b_0 + b_1 t + \sum_{k=1}^{N} a_k e^{i\sigma_k t} + a_{-k} e^{i\sigma_k t} \tag{4}$$

where *N* constituents are used. Each constituent has a frequency σ_k which is known from the development of the potential, and a complex amplitude a_k which is not known, although if AT(*t*) is a real time series a_k and a_{-k} are complex conjugates. The model estimates Eq. 4 by using a least-square method and applying phase and the nodal corrections (Pawlowicz et al. 2002). In this work, the longest continuous time series (without gaps) of the component $SL(t)-MSL_V(t)$, was selected to determine the harmonic coefficients and predict the astronomical tides.

2.3 The residual surges

The stochastic component of the sea level is represented by the residual surges *S*. Its prediction needs high statistical models to be investigated. In this research, an ARMA model was used. This method seems to be the most adequate and suitable technique for the forecast of sea level since it assumes a temporal dependence of the random variables.

ARMA model is a kind of commonly used stochastic time series model, popularized by Box et al. (1994), although the AR (autoregressive) and MA (moving average) models have been previously known and used as an adequate approach of the AR and MA terms modeling in models. The basic idea of this method consists of the following: some time series is a set of random variables depends on the time t, although the single sequence values that constitute the time series is uncertain; however, the changes of the entire sequence has a pattern to follow. By analyzing it through the establishment of the corresponding mathematical model, we can get a better understanding of the structure and the characteristics of time series and achieve the optimized prediction results within the minimum variance (Galman and Disney 2006; Yi 2008).

The ARMA models assume that data are stationary, meaning that their mean and variance do not change with time. However, these models require the marginal distribution associated with the stochastic process under study to be normally distributed, which may not be the case in our case. Therefore, with the aim of preserving the original marginal distribution, the stochastic process of our variable S was transformed into a normalized Gaussian to obtain a new stochastic process with random variables S_n by the use of the following transformation (Liu and Der Kiureghian 1986).

$$S_n = \Phi^{-1}[F(S)] \tag{5}$$

where F is the cumulative distribution function (CDF) of the marginal distribution associated with the original stochastic process S and $\Phi(.)$ is the CDF of the standard normal random variable. Once transformed, the stochastic temporal dependence of the random variables of the process S_n is reproduced by ARMA model with the incorporation of the temporal structure.

The univariate ARMA (p,q) process S_n is mathematically expressed as:

$$S_n(t_i) = \sum_{j=1}^{P} \Phi_j S_n(t_{i-j}) + \varepsilon(t_i) - \sum_{j=1}^{q} \theta_j \varepsilon(t_{i-j}) \tag{6}$$

with p autoregressive parameters $\Phi_1, \Phi_2, \ldots, \Phi p$, and q moving average parameters $\theta_1, \theta_2, \ldots, \theta_q$. The term ε in Eq. 6 stands for an uncorrelated normal stochastic process with mean zero and variance σ_2. Stochastic process ε is also referred to as white noise, innovation term, or error term. As illustrated in Eq. 6, S_n illustrates a linear combination of white noises, and as such, the marginal distribution associated with its stochastic process is necessarily normal to characterize the autocorrelation structure. Note that the extreme data are by definition independent, and contain no information about autocorrelations.

Then, the $p+q$ model parameters should be estimated. Two current methods can be used for this problem: least-squares and the maximum likelihood estimations. Since there are various reasons to keep the model order as low as possible, information criteria may be introduced to combine the need for a good fit with the principle of parsimony. The criterion used in this research is Akaike's Information Criterion (AIC) which joins the residual variance on the one hand and the method orders on the other. The analyst's aim is then to minimize such a criterion. In order to find the number of AR and MA parameters, autocorrelation functions (ACF) and partial autocorrelation functions (PACF) should be calculated. Once ARMA parameters are estimated, the autocorrelation structure is incorporated into the output time series, reproduced by the ARMA model. This method was applied to estimate the residual surges during a selected gap (g) of missing values elapsed from $t_{n\text{-}g}$ and t_n.

The continuous time series of the residual S previous to the gap g ($S(t_1)$, $S(t_2)$, …, $S(t_{n\text{-}g})$) should be used to fit the model parameters (Eq. 6). Then, the missing values of the gap ($S(t_{n\text{-}g})$, $S(t_{n\text{-}g+1})$, $S(t_{n\text{-}g+2})$, …, $S(t_n)$) were simulated hourly. The length of gapes considered in the present analysis varies from days (3, 6, 12, 18, 24, and 30 days) to months (6 to 12 months).

a. Autoregressive ARMA model using the single variability of the residual surge

In this part, a simple application of ARMA model was carried out with only the use of a continuous series $S(t_1)$, $S(t_2)$, …, $S(t_{n\text{-}g})$. The detailed simulation of the model can be summarized in the following sequential steps:

- Step 1: Conversion of variable S to a normal distributed variable S_n, according to Eq. 5, for the continuous domain $[t_1, t_{n\text{-}g}]$. The performance of the model increases with the length of this domain.
- Step 2: Fit S_n to Eq. 6, obtaining the ARMA parameters (Φ and θ) and errors ε.
- Step 3: Simulation of the missing values of the gap g having an independent normal errors ε_g with the same variance of ε obtained from step 2.
- Step 4: Calculation of S_n during the interval $[t_{n\text{-}g}, t_n]$ using Eq. 5, the ARMA parameters (from step 2) and ε_g (from step 3).
- Step 5: Conversion of $S_n(t_{n\text{-}g})$, $S_n(t_{n\text{-}g+1})$, …, $S_n(t_n)$ to $S(t_{n\text{-}g})$, $S(t_{n\text{-}g+1})$, …, $S(t_n)$ using the following expression related to Eq. 5:

$$F^{-1}[\Phi(Sn)] \tag{7}$$

b. Autoregressive ARMA model using the residual surges and a new climate parameter Pc

The residual surge S was investigated together with a climate parameter Pc being linked to the meteorological conditions and sea level variability. In this case, the implementation of ARMA model was carried out taking into consideration

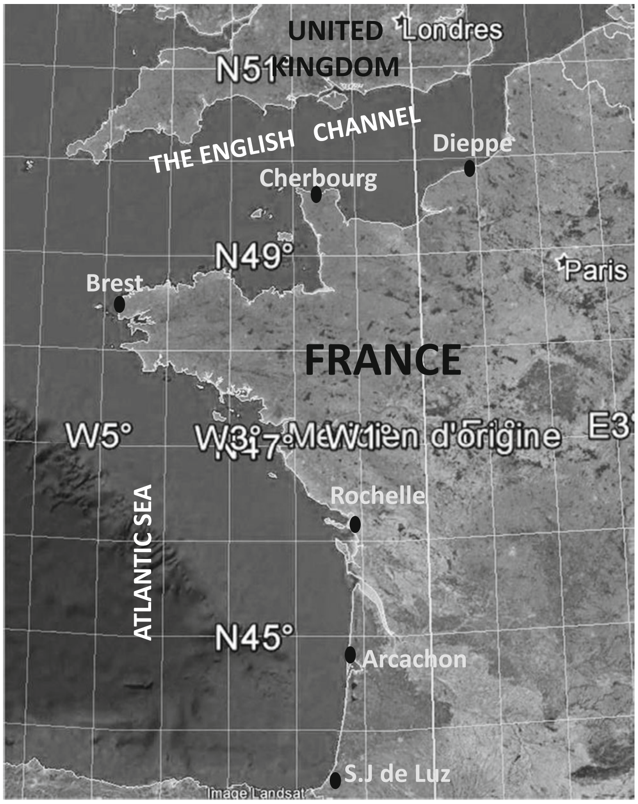

Fig. 2 Study area and localization of tide gauges: the eastern English Channel (Dieppe, Cherbourg) and the North Atlantic Sea of western France (Brest, Rochelle, Arcachon, and S.J. de Luz)

both variables S and Pc. During the time period of the gap g, the local statistical distribution of the missed residual values is assumed to be directly related to the distribution of the climate parameter Pc. According to this hypothesis, we can ameliorate our estimation of the CDF of the residual within the gap $F(S_g)$ using the values of S associated to time periods during which Pc has a statistical behavior similar to Pc_g (Pc within the gap) . In order to do that, we look for the subset of S according to the following expression:

$$F\left(\widetilde{S_g}\right) = S\left(\tilde{t}\right),\ \forall\ \tilde{t} \in [t_1, t_{n-g}] : Pc\left(\tilde{t}\right) < \max\left(Pc_g\right)\ and\ Pc\left(\tilde{t}\right) > \min\left(Pc_g\right) \tag{8}$$

$$\widetilde{F} = F\left(\widetilde{S_g}\right) \approx F\left(S_g\right) \tag{9}$$

Here, the steps 1 and 5 of the previous detailed simulations are transformed to steps 1' and 5' as follows:

- Step 1': In order to convert S to $S_{n,}$ we use $\widetilde{F}$ instead of F (Eqs. 8 and 9).
- Step 5': We use $\widetilde{F}$ in the inverse transformation from the normalized fitted values to the real ones.

3 Case study: application to the eastern English Channel and the North Atlantic Sea (western France)

3.1 Data

The hourly sea level records were obtained from tide gauges of six stations located in the western side of France deployed

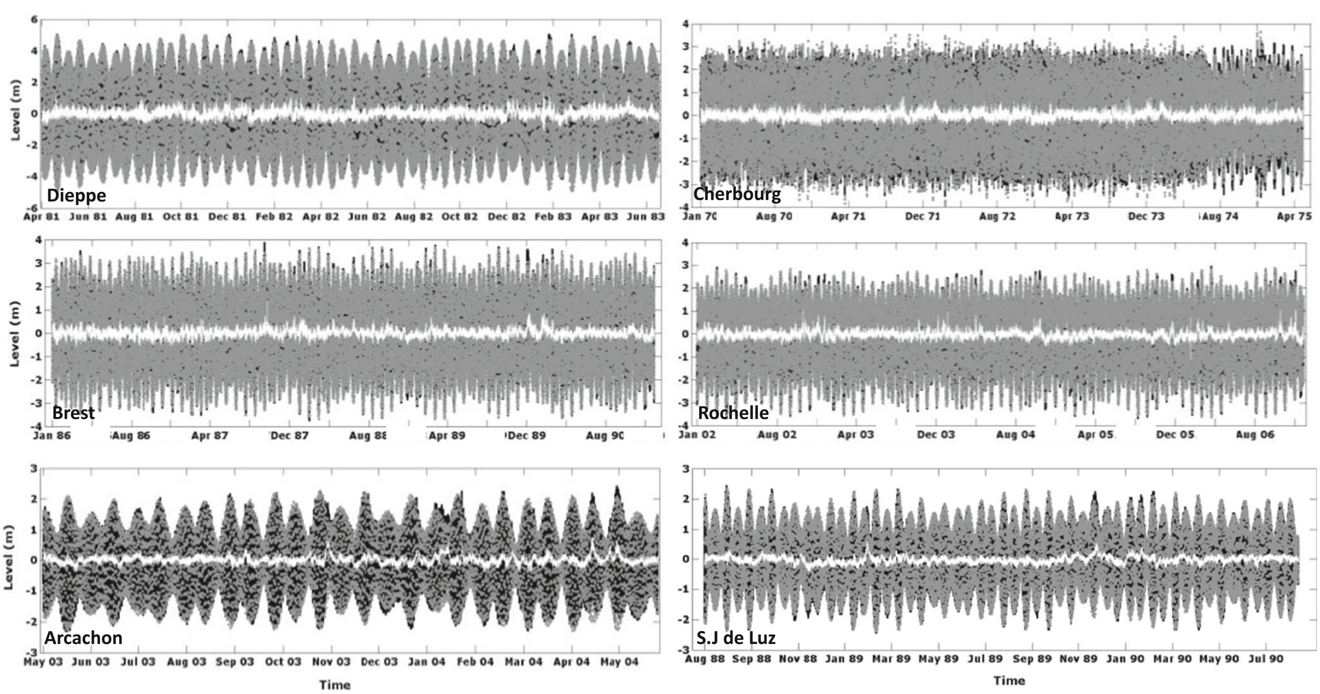

Fig. 3 Sea level components at studied stations: the astronomical tides AT(t) (*grey points*), the residual surges $S(t)$ (*white line*), and the component SL(t) $-$MSL$_V$ (*black line*)

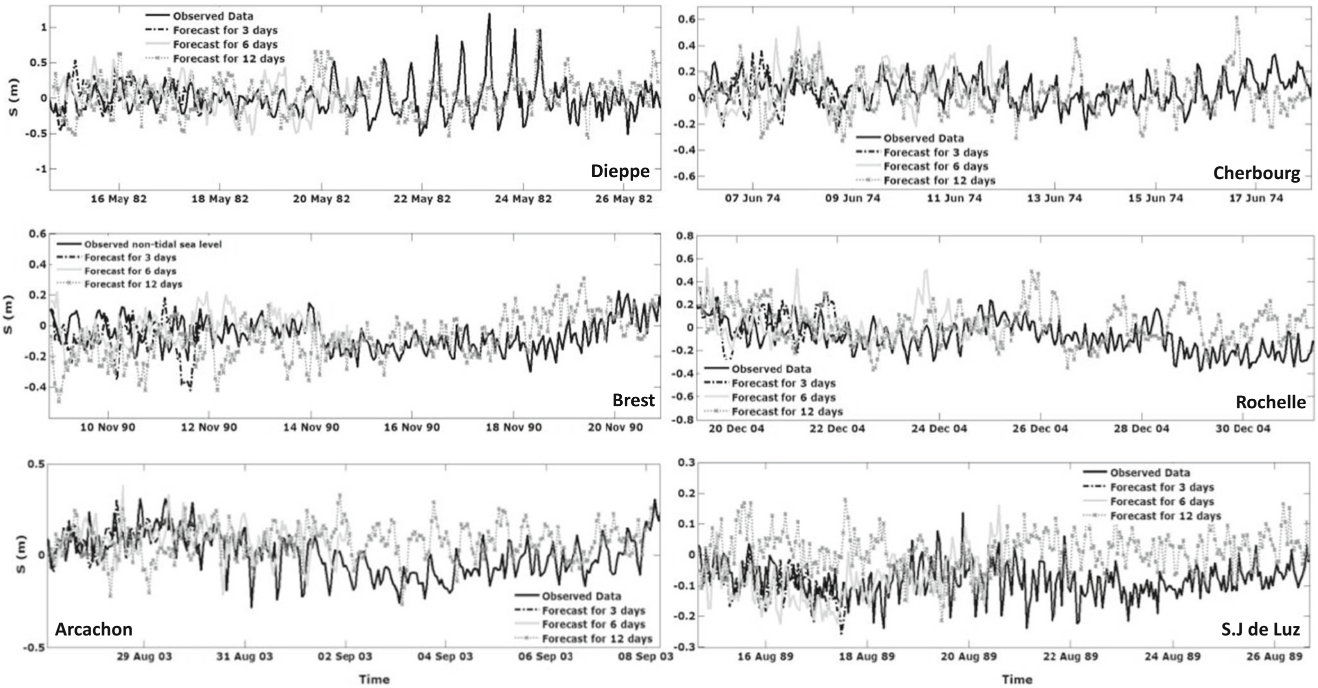

Fig. 4 Forecast of the residual surge $S(t)$. Application of the univariate ARMA model for gaps of 3, 6, and 12 days

in the English Channel and the North Atlantic Sea. These stations (Fig. 2) are operated and maintained by the national French center of oceanographic data REFMAR (http://refmar.shom.fr/fr/maregraphes_french-tide-gauges-data). Observations are referenced to tide gauge zero which corresponds to hydrographic zero level. The used tide gauge records cover a period of 42 years from 1970 to 2012.

Here, the sea level pressure (SPL) was selected as the climatic parameter *Pc* since it is strongly linked with different physical processes (waves, winds, etc.) and the sea level height, as demonstrated by previous studies (Moron and Ullmann 2005). The SPL records at studied stations, extracted from the NOAA website (http://www.ncdc.noaa.gov/oa/climate/research/slp/#desc), were considered as a forcing to predict the stochastic variable *S* using the ARMA model.

3.2 Sea level components

The sea level signal, including the tidal and nontidal components ($SL(t)-MSL_V(t)$), was calculated by the algorithm T-TIDE to extract the astronomical tides during the longest continuous segment of time, selected for each station (section 2). The tidal component $AT(t)$, estimated from the phases and the amplitudes of the different harmonics, is displayed in Fig. 3. Tides are between 6 and −6 m in Dieppe with the most important height of sea level compared to the other stations. They decrease between −4 and 4 m at Cherbourg, Brest, and Rochelle; they vary from −3 to 3 m in Arcachon and S.J. de Luz. Simulated harmonic variability of studied stations differs from each other in the frequency of amplitude of the oscillations. The root mean square error (RMSE) would to range between 20 and 50 cm if we approximate the astronomical tides to the total component $SL(t)-MSL_V(t)$ without considering the residual surges *S*. This latter was obtained from the difference between $SL(t)-MSL_V(t)$ and $AT(t)$ as shown in Fig. 3. This component shows high frequency oscillations with small amplitudes less than 1 m and short wavelengths.

3.3 Estimation of the residual surges

a) Estimation using the single variability of residual surges

The forecast of the hourly residual *S* was performed by the univariate autoregressive model ARMA for short gaps of 3, 6, and 12 days. Results are displayed in Fig. 4 where the simulated and observed data are plotted. As seen, the variability is well reproduced by the ARMA model. However, the magnitudes of the level are not fully enclosed. A better forecast of

Table 1 RMSE between observed and modeled residual component *S*. Error values are calculated in centimeters

	3 days	6 days	12 days
Dieppe	9	16	20
Cherbourg	8.8	13.6	14.6
Brest	8.3	12.1	14.6
Rochelle	7.8	13	15
Arcachon	7.6	13	15
S.J. de Luz	7.1	10.8	11.4

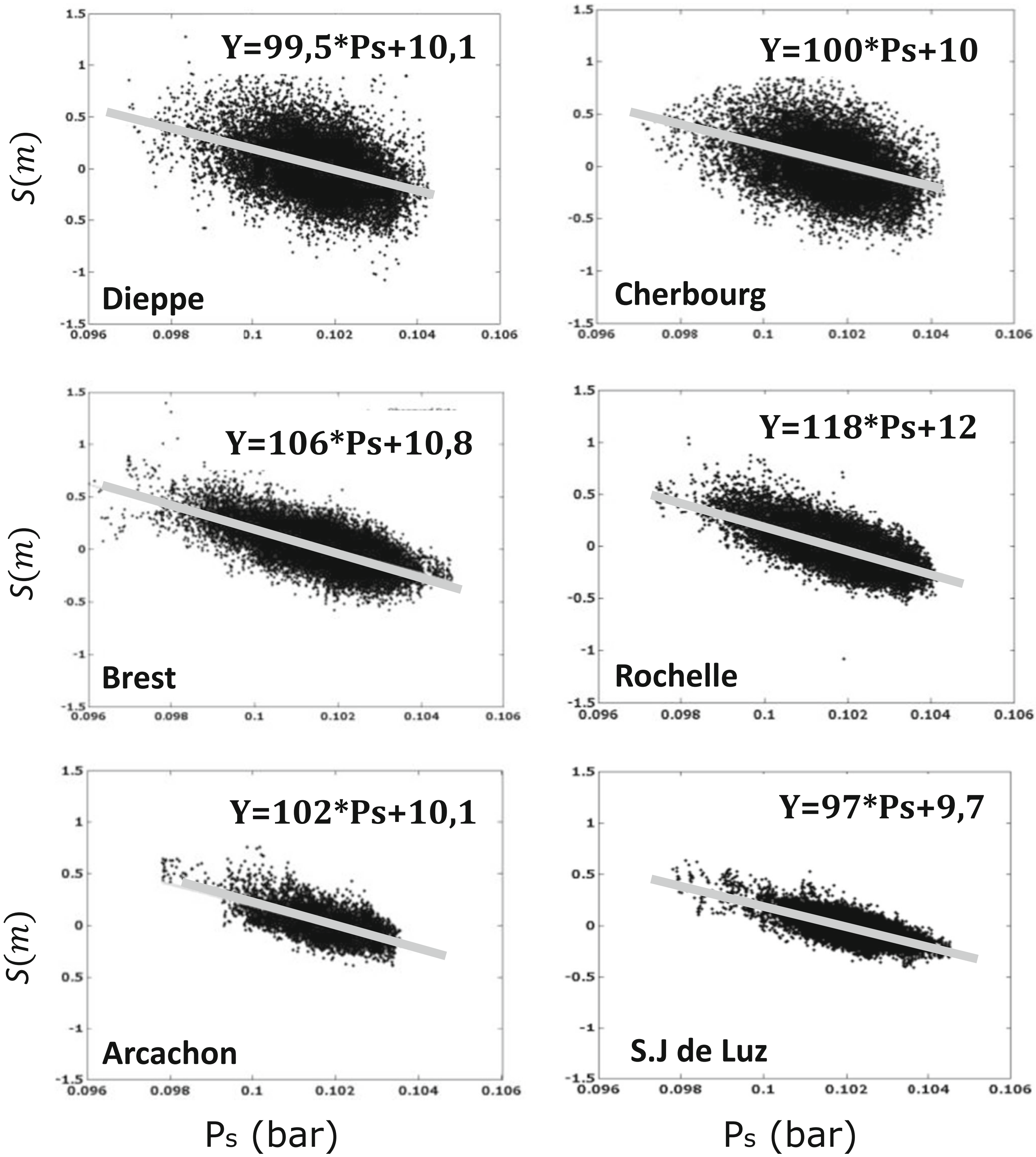

Fig. 5 Scatter plot of the residual surges *S*(*t*) versus the pressure SLP (*bar*). The linear fitting function is also shown

the real data can be observed for 3 days where both trends and magnitudes are reproduced. The RMSE between modeled and observed measurements varies between 10 and 20 cm; it increases with the length of gaps and shows low values in S.J. de Luz and high ones in Dieppe (Table 1).

- Estimation using the variability of the residual surges and the SLP

In this part, the forecast of the residual surges was performed using the sea level pressure. Firstly, the relation between both parameters was investigated as seen in Fig. 5

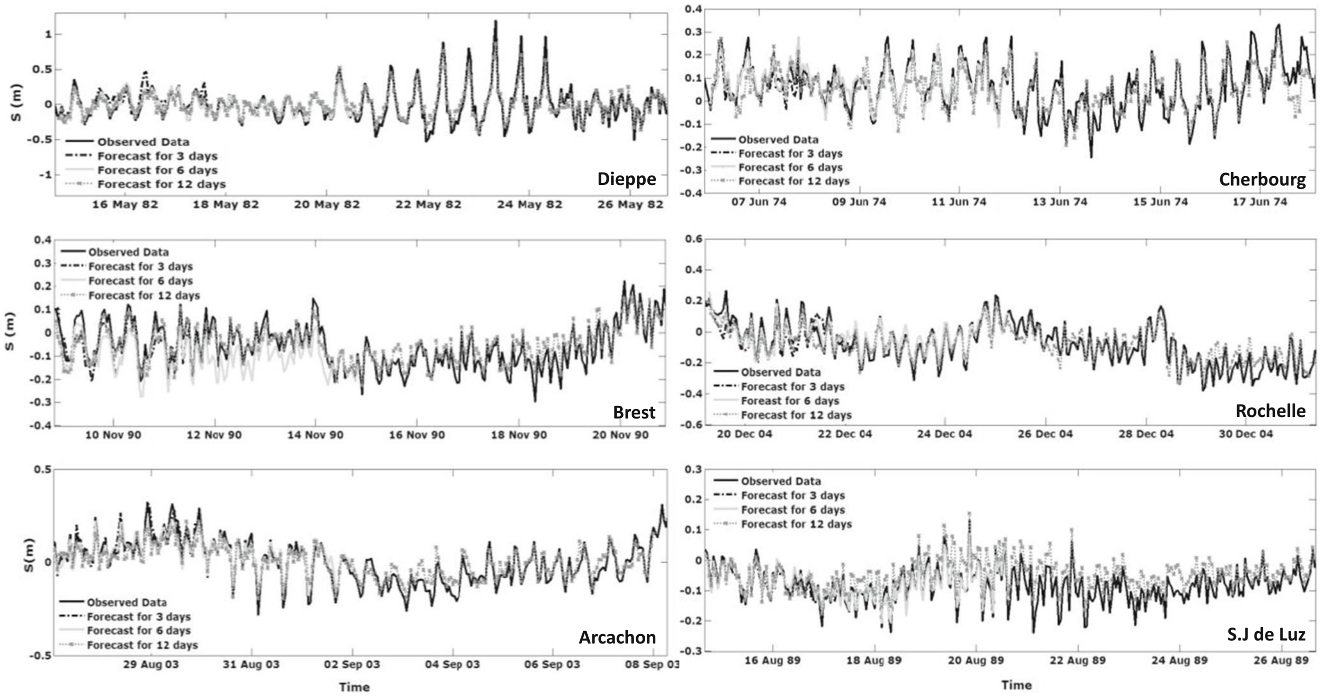

Fig. 6 Forecast of the residual surges $S(t)$. Application of the univariate ARMA model, using SLP distribution, for gaps of 3, 6, and 12 days

where a scatter plot of S versus SLP is presented. The linear regressions, observed between the pressure and the nontidal component of the sea level, seem to be similar at all studied stations. It can be approximated to a linear function $a.x+b$ where a is of 97.9–110 and b is of 9.9–12. However, data differ in their distribution between stations. They display more dispersion in Dieppe and Cherbourg compared to Arcachon and S.J. Luz where the most of data are strongly gathered.

There, linear approximations are not suitable to resolve the complexity of the physical relations between the residual surges and the sea pressures. High accuracy regression methods are required to explain most of the variability between the water elevations and the pressure in the ocean and coastal zones.

In the next part, the SLP data were used to simulate the hourly residual surges by ARMA for gaps 3, 6, and 12 days. Observed and modeled residuals are compared in Fig. 6. Here, the model encompasses the trend and the magnitude of the residual oscillations with an RMSE less than 5 cm (Table 2). The RMSE increases slightly at Dieppe and Cherbourg; it decreases at Arcachon and S.J. de Luz. The high values of surges observed along the gap are well reproduced by the model with a mean explained variance of 75 %. For example, the 74 % of the high values of S, observed between 22 and 25 May 1982 at Dieppe, are strongly correlated with the modeled data; while 82 % of the negative variability produced between 18 and 22 August 1989 is modeled at S.J. de Luz.

Furthermore, the goodness of the forecast was checked for periods more than 12 days. The first time, model simulations were performed for gaps of 18, 24, and 30 days. Three examples in Dieppe, Brest, and Arcachon are presented in Fig. 7 where modeled and observed data are shown. The RMSE ranges generally between 6 and 8 cm. The peaks of high variability, produced during stormy events, are well simulated and seem to be better enclosed by the model for shorter gaps. The mean explained variance of the peak correlations with observed data is of 70 % at Dieppe, 73 % at Brest, and 78 % at Arcachon.

For scales larger than days, a gap of 6 months was used to forecast the residual component at studied stations. An example of this simulation at Brest, between 11 June and 11 December 1990, is shown in Fig. 8a. The highest RMSE reaches an order of 9.2 cm. The observed high values of the residual show good agreements more than 70 % with simulated results.

The statistical behavior of the 1-year modeled data was compared to observations as seen in Fig. 8b where the

Table 2 RMSE between observed and modeled residual component S. Error values are calculated in centimeters

	3 days	6 days	12 days
Dieppe	4	4.3	4.5
Cherbourg	3.8	4	4.2
Brest	3.3	3.8	4
Rochelle	3.5	3.8	4.2
Arcachon	3.2	3.7	4.1
S.J. de Luz	2.8	3.4	3.8

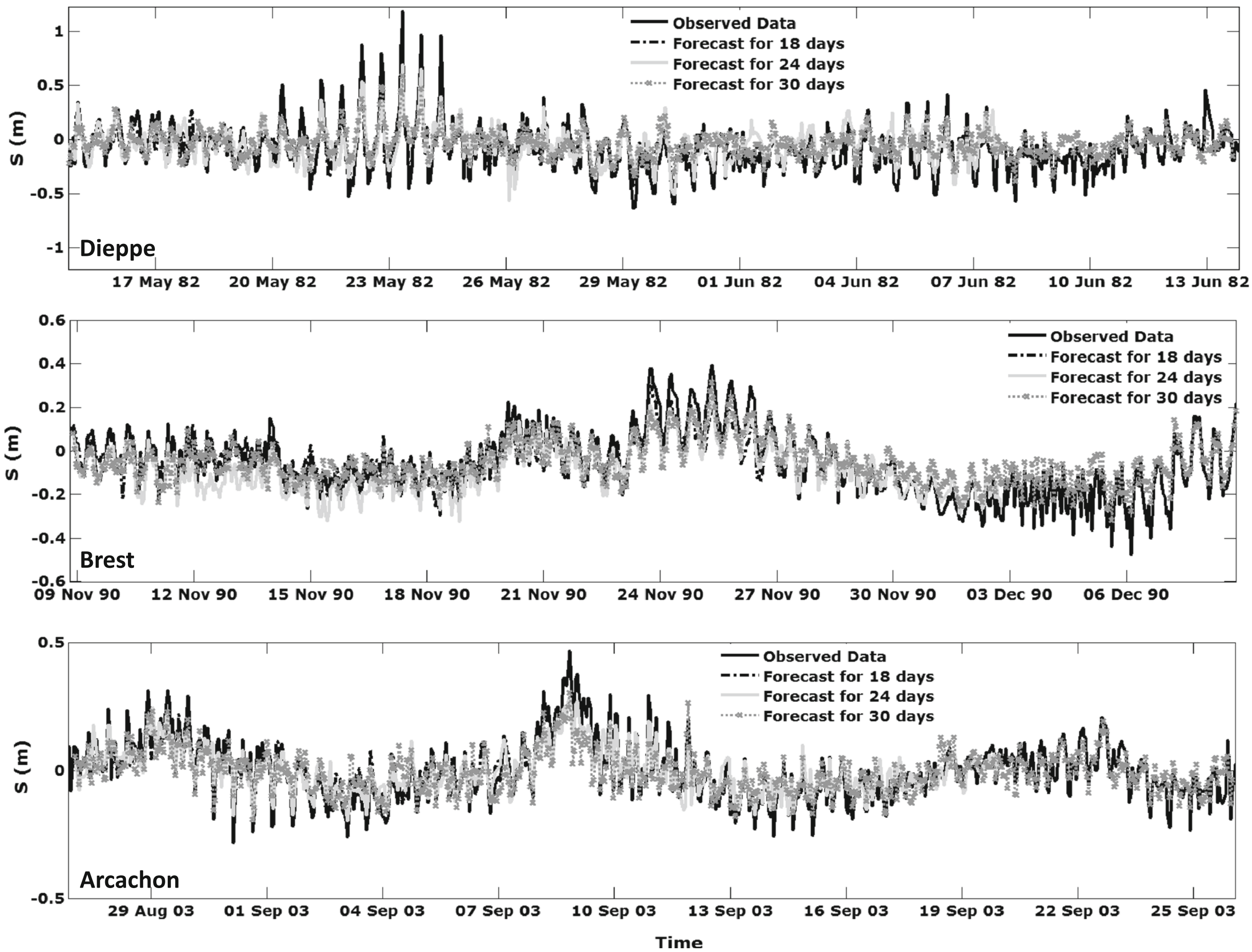

Fig. 7 Forecast of the residual surges $S(t)$. Application of ARMA model, using SLP distribution, for gaps of 18, 24, and 30 days

probability distribution density function (PDF) and the QQplot diagram are displayed. The Gaussian behavior of the observed nontidal component is clearly emphasized by simulated data. Both graphs illustrate high statistical similarities between the observed and modeled distributions. Few differences can be observed for the lowest and the highest values where the model underestimates the extremes with a mean explained variance between 18–30 %.

In the frequency domain, the comparison between 1-year observed and modeled residual components was also performed using discrete wavelet analysis (Fig. 9). The energy spectrum of the simulated residual shows good agreements with observations. Four frequency bands were extracted from the full spectrum: 350, 200, 105, and 50 days (Fig. 9a). Reconstructed signal demonstrates that the variability of the different bands seems to be enclosed by the modeled frequency showing high similarities with observations. This similarity increases for high frequency bands (50 and 105 days) with an explained variance of 86 %. Only, a mean of 72 % of the low frequency bands (200 and 350 days) is reproduced.

- Estimation of the total sea level

Using the previous results, the total sea level SL can be calculated by the sum of the astronomical tide AT, the residual surges S, and the component MSL_V. Examples with different length of gaps (from days to months) were simulated at studied stations. The modeled total sea level was simulated using (1) the deterministic method of harmonic analysis and (2) the new hybrid model combining the linear regression, the harmonic analysis, and the ARMA method using the physical forcing SLP to forecast the three main components MSLv, AT, and S, respectively. An example of these simulations can be illustrated in Fig. 10. The use of the single analysis harmonics (grey circles in Fig. 10a) could not approximate the variability of the total sea level (black line in Fig. 10a) with a RMSE more than 50 cm. The new hybrid model simulates the total sea level with a mean RMSE of 15 cm. The grey bands shown in Fig. 10 in 15–19 and 25–29 October 1989 illustrate the ameliorated results from the first to the second forecast. The grey circles of modeled records, using the harmonic

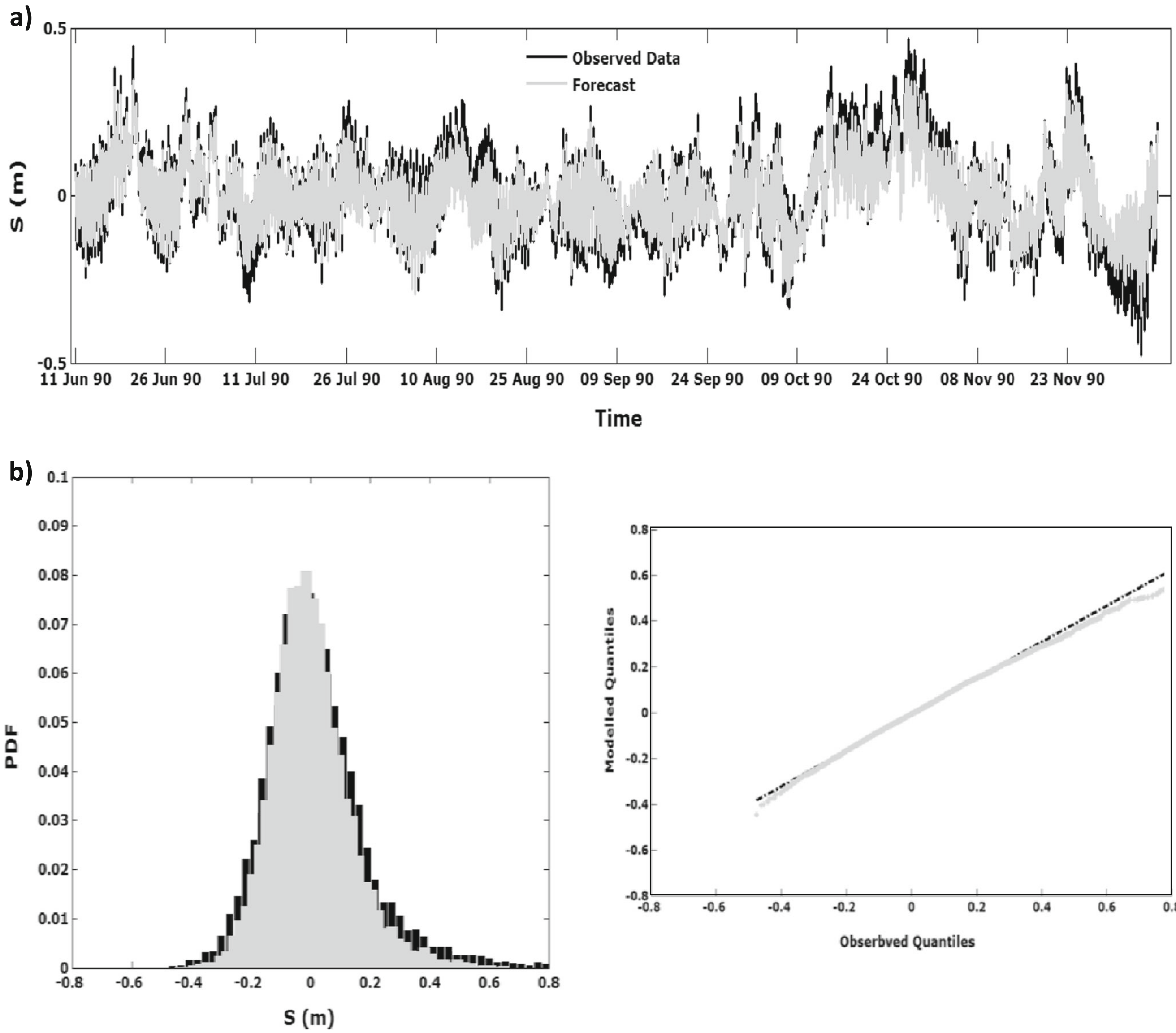

Fig. 8 **a** Modeled (*grey line*) and observed (*black line*) nontidal component for a gap of 6 months at Brest. **b** PDF of observed (*black line*) and modeled (*grey line*) 1-year residual component *S*(*t*) (*left side*); diagram of QQplot between 1-year observed and modeled data

analysis, underestimate the amplitude of the sea level oscillations, especially during neap tides (Fig. 10a). This error is recovered the forecast is clearly improved by the use of the hybrid model (Fig. 10b).

4 Discussions and conclusions

A methodology was developed to estimate the short to medium-term variations of hourly sea level using an autoregressive moving average (ARMA) model. Modeling both astronomical tides and residual surges at timescales of days to months is required to a better understanding of changes in climate events, which is currently on the main concerns relating the climate change. The methodology was implemented with the aim of filling the missing values in the time series of hourly sea level and forecasting its variability.

The astronomical components were calculated using the T-TIDE algorithm and the tide level was simulated. The residual surges, obtained by the difference between $SL(t)-MSL_V$ and $AT(t)$ was estimated using univariate ARMA model. Two methods were proposed in this research: (1) ARMA using only the time series of the residual surges and (2) ARMA using a coupling between the residual surges and an exterior climate parameter having significant correlations with sea level changes. Here, the sea level pressure was used as the climate parameter since it experiences strong linkages with the residual surges. This methodology was applied to a case study in the eastern English Channel and the North Atlantic Sea of western France using six tide gauge records of sea level.

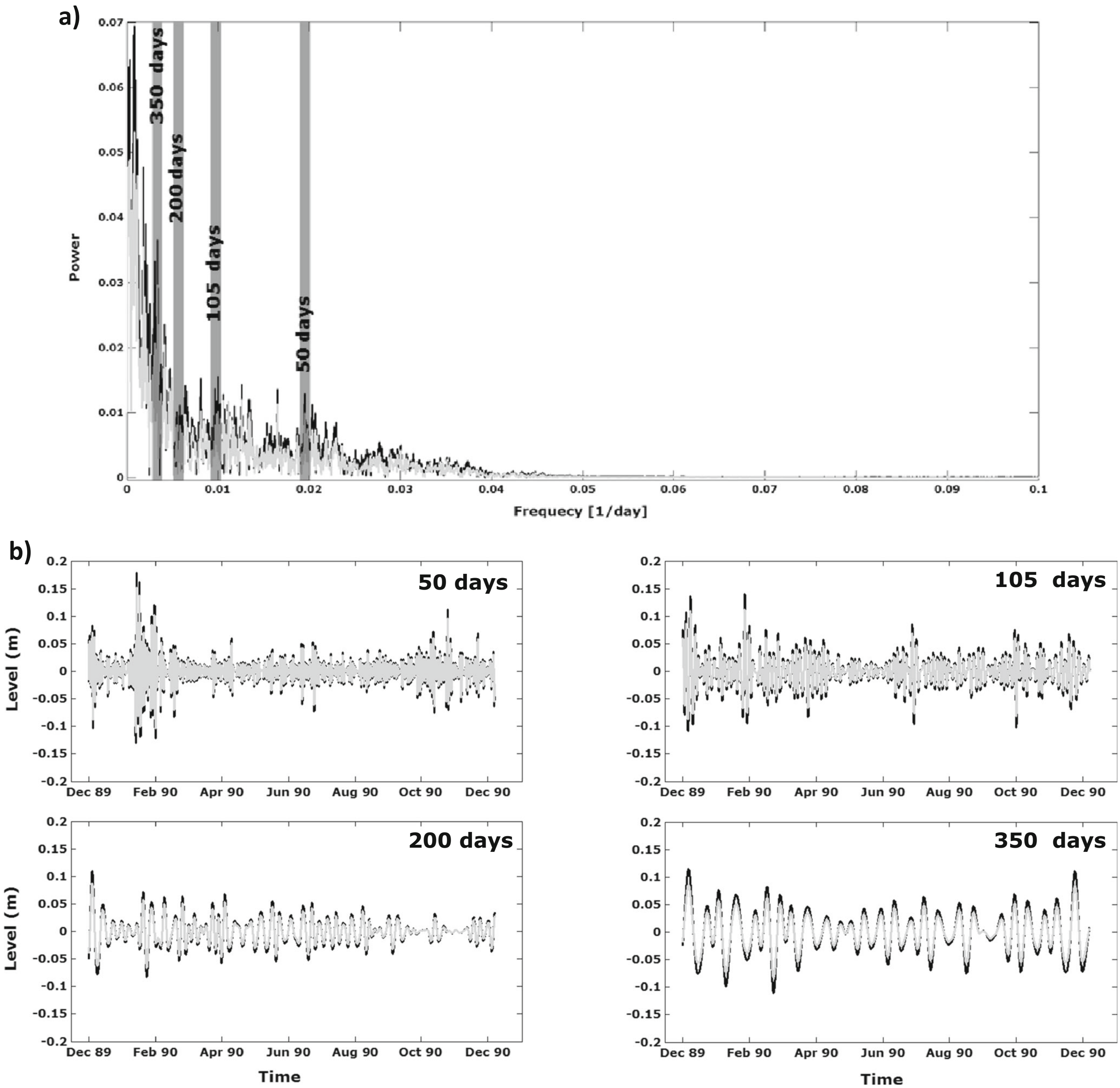

Fig. 9 Frequency analysis of the 1-year residual component: observations (*black line*) and simulations (*grey line*). **a** The discrete wavelet spectrum of data showing four main bands: 350, 200, 105, and 50 days. **b** Reconstructed signal for the different energy bands

Multiple gaps of short (days) to medium (months) timescales were studied. Results have shown that the forecast of hourly residual component is better for short gaps. Considering the linkage between surges and SLP, the application of ARMA has shown significant similarities between modeled and observed data with a mean RMSE of 5.2 cm for a forecast of short periods (3–30 days). For gaps of months, the use of ARMA has emphasized an RMSE of 9.2 cm. For short gaps, the distribution of the SLP is more confined which provides stronger correlation between the residual surges and the SLP. However, the distribution of the SLP and then the residual surges is more dispersed for larger gaps and the relation between both data seems to be less clearly explored. One-year gaps were also used for simulating residual surges with ARMA. Modeled results demonstrate a good statistical representation of the observed data with a strong correlation between the frequency bands. For similar timescales, Minguez et al. (2012) have estimated a stochastic Lagrangian trajectory for drifting objects in the ocean using the univariate ARMA models. They have used this estimation to simulate different trajectories for obtaining location probability density functions at different times.

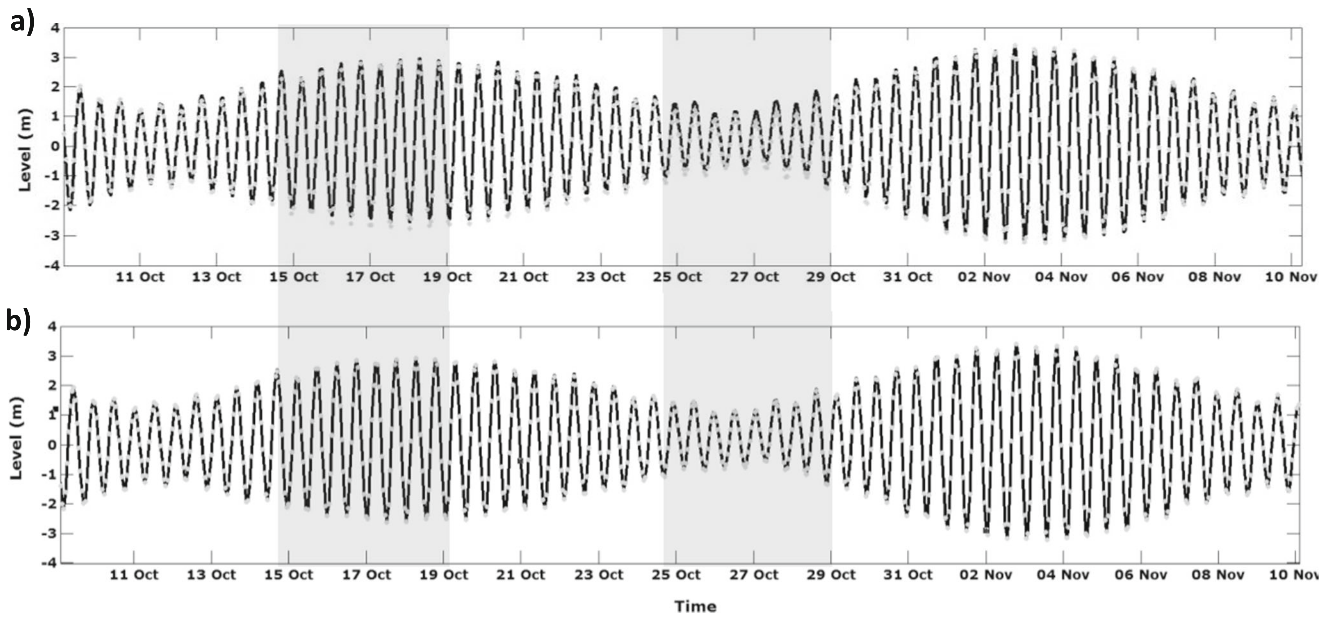

Fig. 10 Forecast of the total sea level using **a** the single harmonic analyses and **b** the new hybrid model. The agreements between observed (*black line*) modeled (*grey circles*) levels can be clearly illustrated within the grey bands

Accordingly, the goodness of filling missing values and predicting sea level data seems to be strongly dependent on the length of gaps and the previous continuous time series. In fact, the predicted missing value is a linear combination of several independent variables—the hourly sea level and several hourly values before it. The coefficients p and q of the univariate ARMA model were determined initially using all available values for the hourly sea levels for a set time series. The estimation of these coefficients is controlled by the length of gaps as well as the available time series previous to the gap. Several simulations of ARMA have shown that a best estimation of ARMA coefficients requires a length previous time series of 700 data for a gap of 140 missing values (20 %). This finding seems to be interesting and agrees with the outcome of previous studies of Pashova et al. (2013) where two artificial neural network architectures for filling the missing values were applied to daily mean sea level data derived from the records of the tide gauge Burgas (located on the western Black Sea coast). They have analyzed 5 years with gaps from 3–4 days to 1–3 weeks and have shown that the required numbers of the previous values is totally 1826 data, 151 of which are missing (8.3 %). Then, the technique of ARMA seems to be more promising for filling gaps and forecasting the sea level at large timescales.

The prediction of the sea level depends also on the period of time showing missing data. In fact, this prediction is more complex for periods of high variability and extreme values caused by stormy events. The present model of ARMA coupled to the SLP has shown its performance to reproduce the peaks of surges with a mean RMSE of 10 cm for all stations.

The proposed methodology has shown good results in the English Channel and the North Atlantic Sea where the level of astronomical tides is dominant and the surge events are not significant with variability of less than 1 m. Then, the relative error respect to the total sea level is small. However, this error should increase for cases of study where the surge components are relatively dominant compared to the astronomical ones.

Concluding, the proposed methodology of the sea level forecast presents a coherent, simple, and easy to estimate the deterministic nature of tidal processes and the stochastic framework of surge events. Its applicability was further reinforced by the results obtained from the case study in the eastern English Channel and the North Atlantic Sea (western France). As future works, we proposed the application of the present methodology to another in the Pacific Sea and the Mediterranean Basin with high residual surges. Here, the SLP seems to be a key physical parameter which can describe the most of the sea level changes since that more than 80 % of the total variability can be reproduced by the use of the ARMA model coupled to the physical process of SLP. Other climate parameters should be taken into consideration in many zones such estuaries where surges are controlled by a series of meteorological indices. Finally, the performance of the model needs to be validated by the comparison of results to dynamical models.

Acknowledgments The first thanks are granted to the CNES and TOSCA program of SWOT for financial support. The authors would like to thank the anonymous reviewers for their careful reading and their valuable comments that helped to improve this work. Many thanks also to the Spanish Government-EU FEDER for partially funding this research (research project CTM2012-35398).

References

Box G, Jenkins G, Reinsel GC (1994) Time series analysis: forecasting and control. Prentice-Hall International, Englewood Cliffs

Church JA, Gregory JM, Huybrechts P, Kuhn M, Lambeck K, Nhuan MT, Qin D (2001) Changes in sea level. Chapter 11 of the IPCC Third Assessment Report, pp. 639–694. Cambridge University Press, Cambridge, UK

Church JA, Clark P, Cazenave A, Gregory J, Jevrejeva S, Levermann A, Merrifield M, Milne G, RS Nerem, Nunn P, Payne A, Pfeffer W, Stammer D, Unnikrishnan A (2013) Sea level change

Cabanes C, Cazenave A, Le Provost C (2001) Sea level rise during past 40 years determined from satellite and in situ observations. Science 294:840–842

Dergachev VA, Makarenko NG, Karimova LN, Danilkina EB (2001) Nonlinear methods of analysis of data with gaps. J Methods Appl Absolute Chronol 20:45–50

Flather RA, Williams JA (2000) Climate change effects on storm surges: methodologies and results. In: Beersma J, Agnew M, Viner D, Hulme M (eds) Climate scenarios for water-related and coastal impact. ECLAT-2 Workshop Report No. 3, pp. 66–78. KNMI, The Netherlands

Flather RA, Smith JA, Richards JD, Bell C, Blackman DL (1998) Direct estimates of extreme storm surge elevations from a 40 year numerical model simulation and from observations. Global Atmos Ocean Syst 6:165–176

Foreman MGG (1977) Manual for tidal height analysis and prediction. Pacific Marine Science Report 77-10. Institute of Ocean Sciences, Patricia Bay, Sidney, BC, 97pp

Foreman MGG (1978) Manual for tidal currents analysis and prediction. Pacific Marine Science Report 78-6, Institute of Ocean Sciences, Patricia Bay, Sidney, BC, 57pp

Galman G, Disney SM (2006) State Space Investigation of Bullwhip with ARMA (1,1) Demand Processes. Int J Prod Econ 35:322–345

Godin G (1972) The analysis of tides. University of Toronto Press, Toronto, p 264

Hilmi K, Chanut JP, EL-Sabh M (1997) Modélisation stochastique des variations à court terme du niveau d'eau Niveau d'cau marin Météorologie Hydrologie Modèle ARMA Estuaire du Saint-Laurent dans l'estuaire du Saint-Laurent, vol 20, N°2. Oceanologica Acta, Canada

Kondrashov D, Ghil M (2006) Spatio-temporal filling of missing points in geophysical data sets. Nonlinear Process Geophys 13:151–159

Lowe JA, Gregory JM, Flather RA (2001) Changes in the occurrence of storm surges around the United Kingdom under a future climate scenario using a dynamic storm surge model driven by the Hadley Centre climate models. Clim Dyn 18:179–188

Liu PL, Der Kiureghian A (1986) Multivariate distribution models with prescribed marginals and covariances. Probabilistic Eng Mech 1(2): 105–112

Minguez R, Abascal AJ, Castanedo S, Medina R (2012) Stochastic Lagrangian trajectory model for drifting objects in the ocean. Stoch Env Res Risk A 26:1081–1093. doi:10.1007/s00477-011-0548-7

Moffat AM, Papale D, Reichstein M, Hollinger DY, Richardson AD, Barr AG, Beckstein C, Braswell BH, Churkina G, Desai AR, Falge E, Gove JH, Heimann M, Hui D, Jarvis AJ, Kattge J, Noormets A, Stauch VJ (2007) Comprehensive comparison of gap filling techniques for eddy covariance net carbon fluxes. Agric For Meteorol 147:209–232

Moron V, Ullmann A (2005) Relationship between sea-level pressure and sea level height in the Camargue (French Mediterranean coast). Int J Climatol 25:1531–1540. doi:10.1002/joc.1200

Musial JP, Verstraete MM, Gobron N (2011) Comparing the effectiveness of recent algorithms to fill and smooth incomplete and noisy time series. Atmos Chem Phys Discuss 11:14259–14308

Pashova L, Koprinkova-Hristova P, Popova S (2013) Gap Filling of Daily Sea Levels by Artificial Neural. Int J Mar Navig Saf Sea Transp. Volume 7, Number 2. DOI: 10.12716/1001.07.02.10

Pawlowicz R, Beardsley R, Lentz S (2002) Classical tidal harmonic analysis including error estimates in MATLAB using T TIDE. Comput Geosci 28:929–937

Pugh DT (1987) Tides, surges and mean sea level: a handbook for engineers and scientists. Wiley, Chichester, p 472

Stocker TF, Qin D, Plattner GK, Tignor M, Allen S, Boschung J, Nauels A, Xia Y, Bex V, Midgley P (eds) (2013) Climate change 2013: the physical science basis, Cambridge University Press, Cambridge, UK and New York, NY, USA

Tawn JA, Vassie JM (1989) Extreme sea levels: the joint probability method revisited and revised. Proc Inst Civ Eng 2 87: 429–442

Von Storch H, Reichardt H (1997) A scenario of storm surge statistics for the German Bight at the expected time of doubled atmospheric carbon dioxide concentration. J Climatol 10:2653–2662

Yi DH (2008) Data analysis and Eviews application. China Renmin University Press, Beijing