AGU PUBLICATIONS

JGR

Journal of Geophysical Research: Oceans

RESEARCH ARTICLE

10.1002/2014JC010360

Key Points:

- Differences between model and observations vary for extremes of similar height
- Unlike return levels, the number of exceedances is not well captured
- Sea level extremes induced by explosive cyclones are better captured

Correspondence to:
F. M. Calafat,
francisco.calafat@noc.ac.uk

Citation:
Calafat, F. M., E. Avgoustoglou, G. Jordà, H. Flocas, G. Zodiatis, M. N. Tsimplis, and J. Kouroutzoglou (2014), The ability of a barotropic model to simulate sea level extremes of meteorological origin in the Mediterranean Sea, including those caused by explosive cyclones, *J. Geophys. Res. Oceans*, *119*, 7840–7853, doi:10.1002/2014JC010360.

Received 2 AUG 2014
Accepted 27 OCT 2014
Accepted article online 31 OCT 2014
Published online 21 NOV 2014

The ability of a barotropic model to simulate sea level extremes of meteorological origin in the Mediterranean Sea, including those caused by explosive cyclones

F. M. Calafat[1], E. Avgoustoglou[2], G. Jordà[1,3], H. Flocas[4], G. Zodiatis[5], M. N. Tsimplis[1], and J. Kouroutzoglou[2,4]

[1]National Oceanography Centre, Southampton, UK, [2]Hellenic National Meteorological Service, Athens, Greece, [3]Institut Mediterrani d'Estudis Avançats (UIB-CSIC), Esporles, Spain, [4]Faculty of Physics, Department of Atmospheric Physics-Meteorology, National and Kapodistrian University of Athens, Athens, Greece, [5]Oceanography Centre, University of Cyprus, Nicosia, Cyprus

Abstract Storm surges are responsible for great damage to coastal property and loss of life every year. Coastal management and adaptation practices are essential to reduce such damage. Numerical models provide a useful tool for informing these practices as they simulate sea level with high spatial resolution. Here we investigate the ability of a barotropic version of the HAMSOM model to simulate sea level extremes of meteorological origin in the Mediterranean Sea, including those caused by explosive cyclones. For this purpose, the output of the model is compared to hourly sea level observations from six tide gauge records (Valencia, Barcelona, Marseille, Civitavecchia, Trieste, and Antalya). It is found that the model underestimates the positive extremes significantly at all stations, in some cases by up to 65%. At Trieste, the model can also sometimes overestimate the extremes significantly. The differences between the model and the residuals are not constant for extremes of a given height, which limits the applicability of the numerical model for storm surge forecasting because calibration is difficult. The 50 and 10 year return levels are reasonably well captured by the model at all stations except Barcelona and Marseille, where they are underestimated by over 30%. The number of exceedances of the 99.9th and 99.95% percentiles over a period of 25 years is severely underestimated by the model at all stations. The skill of the model for predicting the timing and value of the storm surges seems to be higher for the events associated with explosive cyclones at all stations.

1. Introduction

The damage resulting from the combined effect of rising sea levels and storm surges is one of the most visible and costly impacts of climate change. Some of the effects are felt immediately in the form of increased flooding and saltwater intrusion into surface waters, whereas other effects, such as coastal erosion and saltwater intrusion into ground waters, may be only noticeable after sometime [*Nicholls*, 2010]. The damage caused by these effects is potentially very large because, although coastal zones represent just a small fraction of the Earth's total land area, they support significant economic activity and concentrate numerous coastal assets including ports, roads, rail, and residential properties. It has been estimated that average global flood losses will increase from $6 billion per year in 2005 to $52 billion per year by 2050 due to projected socioeconomic change alone, without consideration of future changes in the climate [*Hallegatte et al.*, 2013]. In addition, coastal zones are also home to rich marine ecosystems that are vulnerable to both sea level rise and changes in water mass properties [*Nicholls*, 2010]. Clearly, adaptation and enhanced protection are required with a view to reducing those impacts. For adaption policy to be effective, however, the introduction of adaptation measures must be informed by robust understanding of past sea level changes and expected future changes. In particular, a clear knowledge of sea level extremes, such as storm surges, is especially important for coastal planning. In this sense, numerical models provide a useful tool for better understanding sea level extremes as well as for projecting future changes. Proper validation of these models is, however, essential before they can be confidently used for such purposes. In this study, we test the ability of one of the most widely used numerical models in the Mediterranean Sea to simulate sea level extremes of meteorological origin.

In climate studies of sea level extremes, it is common practice to model the tail of the distribution of sea level observations by means of another probability distribution (e.g., the generalized Pareto distribution), which then is used to derive climatological quantities of interest such as quantiles, return levels, or number of exceedances of a relevant threshold [e.g., *Marcos et al*., 2009; *Menéndez and Woodworth*, 2010; *Feng and Tsimplis*, 2014]. Such quantities are commonly used by engineers and policy makers for coastal planning and management as they provide a good indication of the intensity and frequency of sea level extremes. All these quantities can be directly derived from sea level observations from coastal tide gauge records, provided that such records cover a long enough period. However, tide gauges are sparsely distributed in space, which considerably limits the locations where practical extreme value analysis based on observations is possible. Numerical models are a useful tool to overcome this limitation as they simulate sea level with high spatial resolution, and thus permit the study of the statistical properties of sea level extremes at any coastal grid point within the model domain. In addition, numerical models, forced with atmospheric forcing under scenarios of increased greenhouse gases, also allow us to make projections about future changes in sea level extremes due to climate change. As a consequence, many of the existing studies of sea level extremes use barotropic ocean models driven by atmospheric pressure and wind to investigate the characteristics of sea level extremes and possible future changes [*Woth et al*., 2006; *Woodworth et al*., 2007; *Ratsimandresey et al*., 2008; *Sterl et al*., 2009; *Tsimplis and Shaw*, 2010; *Jordà et al*., 2012a; *Marcos et al*., 2009, 2011, 2012; *Conte and Lionello*, 2013; *Haigh et al*., 2014] (A. Arns et al., Determining return water levels at ungauged coastal sites: A case study for northern Germany, submitted to *Ocean Dynamics*, 2013).

The use of modeled data, however, is not without its limitations. Uncertainty due to incomplete knowledge of boundary conditions (lateral and at surface), bathymetric errors as well as parametric uncertainty due to inadequate model physics, all can affect the model skill and thus also the quality of the modeled data. The ability of a model to simulate sea level extremes is usually assessed by determining the extent to which the model describes the main statistical properties of the observed sea level extremes [e.g., *Marcos et al*., 2009; *Jordà et al*., 2012b; *Conte and Lionello*, 2013], which is in essence equivalent to determining whether the modeled and observed data have the same distribution. However, this is sometimes done by comparing modeled data and observations for a single percentile only [e.g., *Jordà et al*., 2012b] or for the maximum sea level value and the 50 year return level only [e.g., *Marcos et al*., 2009], which provides only a partial picture of the capability of the model to describe the distribution of extremes. Furthermore, it is important to note that good performance in simulating the observed distribution of extremes does not necessarily translate into good predictive capability. In other words, it is possible that the model simulates the main statistical properties of the extremes correctly but fails to predict their timing and magnitude. This may suffice if one is interested only in the climatological aspect of the extremes, but may not be adequate for storm surge forecasting.

Our aim here is to test the ability of a barotropic model to simulate both the observed distribution of sea level extremes as well as their timing and magnitude in the Mediterranean Sea. To address this goal, we compare the output of the model with the sea level observations from which the tidal signal and the seasonal signal have been subtracted (hereinafter referred to as residuals) from six tide gauge records in the Mediterranean Sea. In particular, we quantify the differences between the model and the residuals as a function of the value of the sea level extremes, use quantile-quantile (q-q) plots to identify similarities and differences between the modeled and observed probability distributions of sea level, and assess the capability of the model to capture important climatological quantities such as return levels and number of exceedances of a threshold over a certain period of time.

In addition, we investigate whether the largest sea level extremes are associated with the occurrence of explosive cyclones over the Mediterranean Sea. Explosive cyclones are a special type of midlatitude cyclone that occur in maritime environments being characterized by a rapid deepening of sea level pressure. The effect of such rapid deepening on the model performance is also explored here. Explosive cyclones can cause extreme phenomena such as significant wave heights that severely affect coastal areas, ports and shipping with serious social and economic impacts, and even loss of lives [e.g., *Lionello et al*., 2006]. In the Mediterranean, explosive cyclones, despite their rarity, appear as a high impact phenomenon during the cold period of the year [*Kouroutzoglou et al*., 2011a, 2014]. The model that we use is a barotropic version of the Hamburg Shelf-Ocean Model (HAMSOM), which is widely used for the analysis of sea level extremes in the Mediterranean Sea. Section 2 describes the data used in the analysis, their processing, and the methodology used to derive return levels and number of exceedances of a threshold. In section 3, we present the results of the analysis. Finally, section 4 summarizes and discusses the main results of the paper.

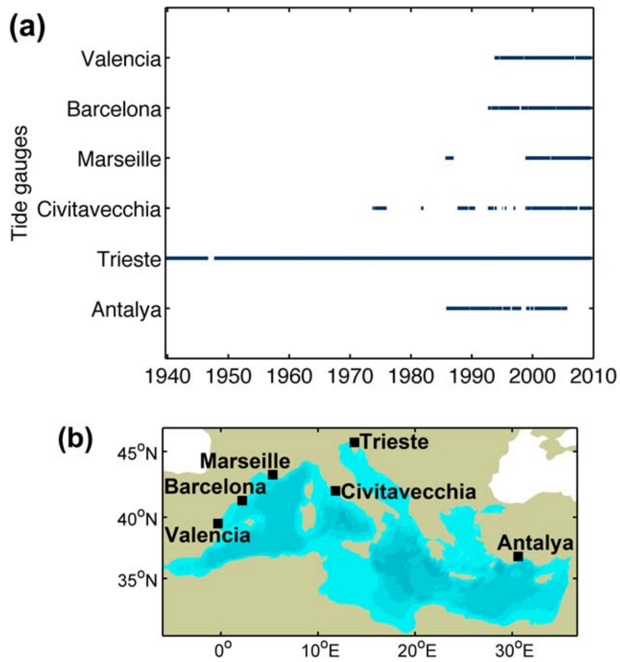

Figure 1. (a) Hourly sea level observations available at the tide gauges. (b) Location of the tide gauges.

2. Data and Methods

2.1. Ocean Model

The model used here to simulate sea level extremes is a barotropic version of the HAMSOM forced by atmospheric pressure and wind. It is also the model used by the Spanish Harbours authority (Puertos del Estado) for operational forecasting of storm surges. The model configuration is the same as that used to generate the Hindcast of Dynamic Processes of the Ocean and Coastal Areas of Europe (HIPOCAS) sea level residual data set [*Ratsimandresy et al*., 2008] except that the forcing is provided by the Agencia Estatal de Meteorologia (AEMET) atmospheric hindcast for the period 1990–2009. The model outputs are hourly sea level fields at $1/6^\circ \times 1/4^\circ$ spatial resolution covering the Mediterranean Sea and the Atlantic coast of the Iberian Peninsula. In the model integration, the pressure gradient and the vertical diffusivity terms are integrated using a semiimplicit scheme while the momentum advection and the horizontal diffusion terms use an explicit one. The bottom friction coefficient is constant and set to 0.0025. Sea level along the western and northern boundaries is prescribed using the inverted barometer effect. No tidal forcing has been included in this simulation. The atmospheric forcing provided by AEMET is a dynamical downscaling of ERA-INTERIM reanalysis [*Dee et al*., 2011] with a spatial resolution of 25 km. Hourly fields of atmospheric pressure at the sea level and 10 m winds are used to force the model. The annual sea level cycle was removed from each model grid point by means of a harmonic analysis. The modeled sea level at the specific locations of the tide gauges is taken from the closest model grid to each tide gauge.

In addition to the sea level simulation based on the AEMET atmospheric hindcast, we generate an additional set of 20 short simulations centered about the time of maximum deepening of 20 explosive cyclones (see subsection 2.3) by forcing the model with wind and atmospheric pressure from the COSMO.GR model [*Avgoustoglou and Tzeferi*, 2013]. The COSMO.GR local nonhydrostatic numerical weather prediction provided atmospheric data with a spatial resolution of 0.0625° (approximately 7 km) and 1 h temporal resolution. It is used operationally by the Hellenic National Meteorological Service and its integration domain covers the whole Mediterranean area (www.cosmo-model.org/content/tasks/operational/default.htm). The COSMO.GR model runs were driven by analysis and forecast boundary conditions of the global model of the European Centre for Medium-Range Forecasts (ECMWF). For the cases, when the explosive cyclogenesis occurred around 00 and 12 UTC, the simulation time range was 72 h (36 h before and after the event), whereas for those when the explosive cyclogenesis occurred around 06 and 18 UTC, the simulation time range was 60 h (30 h before and after the event). The reason for that was to ensure that all simulations started at 00 and 12 UTC given that the simulations with forecast fields from ECMWF are available with starting time only at 00 and 12 UTC.

2.2. Tide Gauge Data

Hourly values from six tide gauge records (Valencia, Barcelona, Marseille, Civitavecchia, Trieste, and Antalya) in the Mediterranean Sea have been used in the analysis (Figure 1). The time series have been collected from different databases, including the Spanish Harbours Tide Gauges Network (REDMAR), the Réseaux de

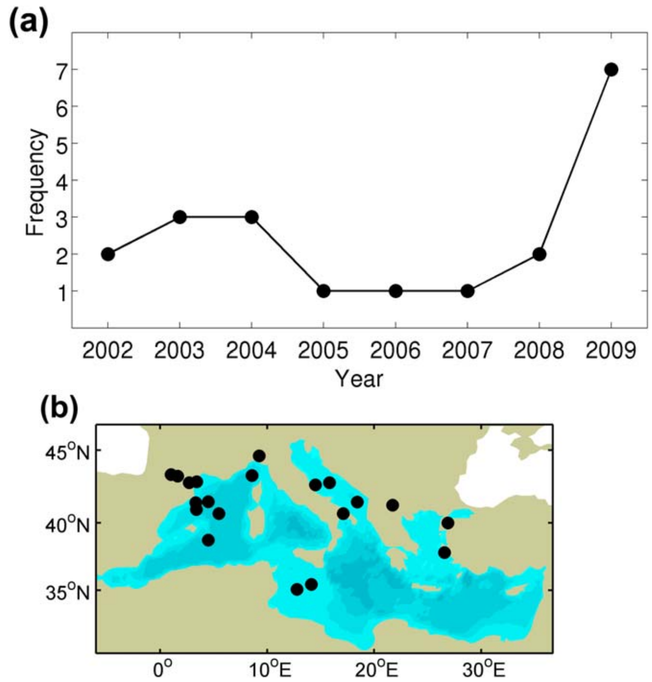

Figure 2. (a) Number of explosive cyclones per year in the Mediterranean Sea. (b) Location of the explosive cyclones at the time of maximum intensity.

référence des observations marégraphiques (REFMAR), and the Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA—IDROMARE). The tide gauge records span different time periods (Figure 1a) from 10 to 70 years; however, in this study, we will restrict our analysis to the period covered by the model run (1990–2009). The quality checks performed on the sea level records are the same as used by *Marcos et al*. [2009] to which the reader should refer for details.

Because the barotropic model is forced by wind and atmospheric pressure, it does not simulate tides or the thermal part of the seasonal cycle associated with the expansion and contraction of the upper layers induced by higher summer and lower winter temperatures. These two components of sea level were removed from the tide gauge records before comparison. First, for each year, the amplitudes and phases of the tidal constituents were estimated by harmonic analysis using the program *t-tide* [*Pawlowicz et al*., 2002]. Only constituents with a signal-to-noise ratio equal or larger than three were used to reconstruct the tidal signal. The annual cycle was not removed during this first step. In a second step, and as for the modeled data, the time-mean annual sea level cycle was removed from the tide gauge records by means of a harmonic analysis performed over the entire period covered by the data (1990–2009).

2.3. Explosive Cyclones Data

Extratropical cyclones are defined as explosive when their deepening rate is maintained at a rate of at least 1 hPa/h for 24 h or more [*Kouroutzoglou et al*., 2011a]. The Mediterranean explosive cyclones were identified with the aid of ERA-Interim data sets with a resolution of $0.5^\circ \times 0.5^\circ$ latitude-longitude grid for the period 2002–2009 employing the Melbourne University tracking algorithm. More details on the application of this algorithm in the Mediterranean can be found in *Kouroutzoglou et al*. [2011a, 2011b]. The average life span of explosive cyclones over the Mediterranean Sea is about 4 days and their average radius is about 300 km [*Kouroutzoglou et al*., 2011a]. Hence, here we will identify a sea level extreme as being caused by an explosive cyclone whenever an explosive cyclone occurred within 3 days of the extreme and within a distance of 250 km from the tide gauge station.

The total number of explosive cyclones in the Mediterranean Sea for the period 2002–2009 is 20. Nevertheless, there is significant interannual variability in their frequency (Figure 2a), with some years showing only one event and one of the years (2009) showing as much as seven events. The spatial distribution of explosive cyclones exhibits also a high degree of nonuniformity (Figure 2b). Explosive cyclones prefer to generate in the maritime environment of the strongly baroclinic northern Mediterranean coast. The most favorable region for maximum explosive cyclone deepening is the Gulf of Lions. It is also worth noting that although explosive cyclones tend to form in the Western Mediterranean, their scale and depth is greater in the Eastern Mediterranean [*Kouroutzoglou et al*., 2011a].

2.4. Return Levels and Exceedances of a Threshold

In this study, we test the ability of the numerical model described in subsection 2.1 to simulate the 50 and 10 year return levels and the number of exceedances of the 99.9th and 99.95% percentiles over a period of

25 years. We use the generalized Pareto distribution (GPD) to model the tail and estimate the climatological quantities of interest. The GDP is used to model the excesses over a high threshold, ϑ, and its cumulative distribution function is defined by [*Pickands*, 1975]:

$$F(x;\xi,\sigma,\vartheta)=1-\left(1+\frac{\xi(x-\vartheta)}{\sigma}\right)^{-1/\xi} \tag{1}$$

where x ($x>\vartheta$) contains the sea level observations exceeding ϑ, σ ($\sigma>0$) is the scale parameter, and ξ is the shape parameter. The scale and shape parameters are estimated by maximum likelihood. Here we choose an average (over the whole period covered by the data) of 10 extremes per year to be fitted by the GPD. *Tsimplis and Blackman* [1997] showed that choosing the number of extremes per year between 5 and 10 led to consistent return levels.

The N year return level (RL_N) is the level that is expected to be crossed on average once every N years and can be estimated by:

$$RL_N=F^{-1}\left(1-\frac{1}{\lambda N};\xi,\sigma,\vartheta\right) \tag{2}$$

where λ is the mean number of exceedances of the threshold ϑ per year.

Finally, the number of exceedances, n_e, of the threshold ϑ_e over a period of N years is given by:

$$n_e=[1-F(\vartheta_e;\xi,\sigma,\vartheta)]\lambda N \tag{3}$$

3. Results

3.1. Time Series Comparison

In order to gain some qualitative insight on the skill of the model, we begin by visually comparing hourly time series of sea level from the model and the residuals at Marseille and Trieste (Figure 3). Overall the sea level variability is well captured by the model at both Marseille and Trieste, which is reflected in the significant correlation between the time series (0.81 and 0.67, respectively). However, the good correlation between the time series does not necessarily imply good simulation of the extreme values. At Marseille (Figure 3a), for instance, the model clearly underestimates both the positive and negative sea level extremes, in some cases by as much as 50 cm. The magnitude of underestimation varies greatly from one extreme to another, even for extremes of a similar height, reflecting the fact that the underestimation is not a linear or simple function of the extreme value. The largest sea level extremes in the model do not necessarily coincide in time with those in the residuals.

The comparison for Trieste (Figure 3b) shares some features in common with that for Marseille, such as the general underestimation of negative extremes by the model and the fact that the differences between the residuals and the model vary rather randomly with the value of the extreme events. However, it also shows some distinctive features, especially in the simulation of positive extremes. In particular, the model does not always underestimate the positive extremes but sometimes it overestimates them, by as much as 40 cm. In fact, the model sometimes shows sea level extremes when none is observed in the residuals.

3.2. Differences as a Function of the Extreme Value

In order to quantify the differences between model and residuals the data are processed as follows. First, all independent extremes (the largest 200 sea level values separated by at least 3 days) in the residuals are sorted in descending order and stored in a vector. It is worth noting that tests using 1 day instead of 3 days for the selection of independent extremes led to practically the same results and conclusions. Second, for each of the independent extremes found in the residuals, we select and store in a vector the largest sea level value from the model within an interval of 6 days centered about the time at which the extreme value is observed in the residuals. Finally, we subtract the two vectors calculated in the previous steps to obtain a vector of the differences between the residuals and the modeled sea level. These differences provide an estimate of the skill of the model for storm surge forecasting. Because extreme events of a similar height can occur several times over the period covered by the residuals, differences are sorted into bins evenly spaced by 5 cm and, for each bin, the mean difference and its standard deviation are calculated. Hence, a

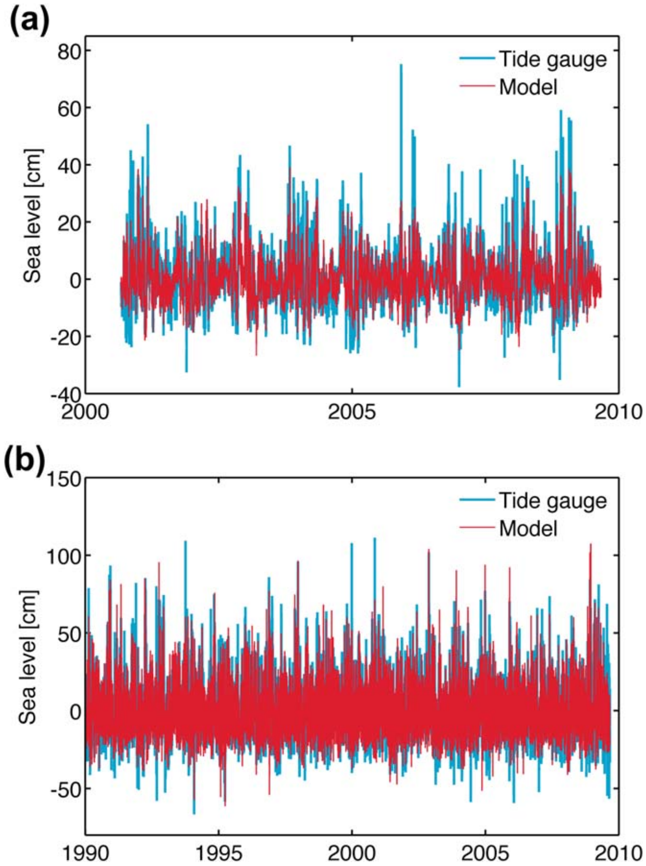

Figure 3. Comparison of the sea level from the model (red) and the deseasoned and detided sea level from the tide gauge records (blue) at (a) Marseille and (b) Trieste.

bin centered at 30 cm will reflect the mean difference and its standard deviation for extreme events in the range [27.5, 32.5]. Changing the bin width does not change the results appreciably. The standard deviation provides a measure of how much the differences vary for extremes of similar magnitude.

Figure 4 shows the average differences for positive extremes as a function of the extreme value in the residuals. The standard deviation (error bars) of the average difference is also shown, but only when three or more values are used in the average. We note that, at Valencia, Barcelona, and Marseille, average differences tend to increase with the value of the extreme residual. This is not the case for the other three stations, where the average differences fluctuate in a random manner over the whole range of extreme values. Focusing on the largest extreme value at each station, we note that average differences are larger than 30 cm at all stations except at Civitavecchia and Antalya, where they are smaller than 15 cm. A maximum average difference of 48 cm is found at both Marseille and Trieste for extreme values of 75 and 110 cm, respectively, which corresponds to an underestimation of the extreme of 64% and 44%. The error bars associated with the average differences convey also important information as they provide a measure of how spread out the differences are for extreme events of a given height. The large error bars found at all stations indicate that the differences between the residuals and the model for extreme events of similar magnitude vary over a wide range of values. This substantial spread in the differences makes it very difficult, if not impossible, to calibrate the model for storm surge forecasting because the model bias varies between stations and in time. To illustrate this, let us focus on the differences for extreme events of 75 cm at Trieste. On average, the model underestimates the extremes of this height by about 20 cm (as denoted by the blue square); however, the error bar indicates that in some cases the model underestimates the extremes by $>$45 cm and in other cases it overestimates the extremes by $>$5 cm. Broadly similar results are found at all other stations, reflecting the low predictive ability of the model with respect to extremes.

The differences for negative extremes as a function of the extreme value in the residuals are shown in Figure 5. Overall the performance of the model is not better than for positive extremes, and the model tends to underestimate the value of the negative extremes at all stations. The largest differences are found at Trieste, where a maximum average difference of 30 cm is found for extreme events of -55 cm. Differences at all other stations are, in general, smaller than 15 cm, and they show a tendency to increase with the magnitude of the extreme value. Another interesting feature is that, unlike for positive extremes, no overestimation is observed in any case but the model always underestimates the magnitude of the negative extremes at all stations. Finally, it is important to note that the error bars are again large, reflecting a notable spread in the differences for extreme events of similar height.

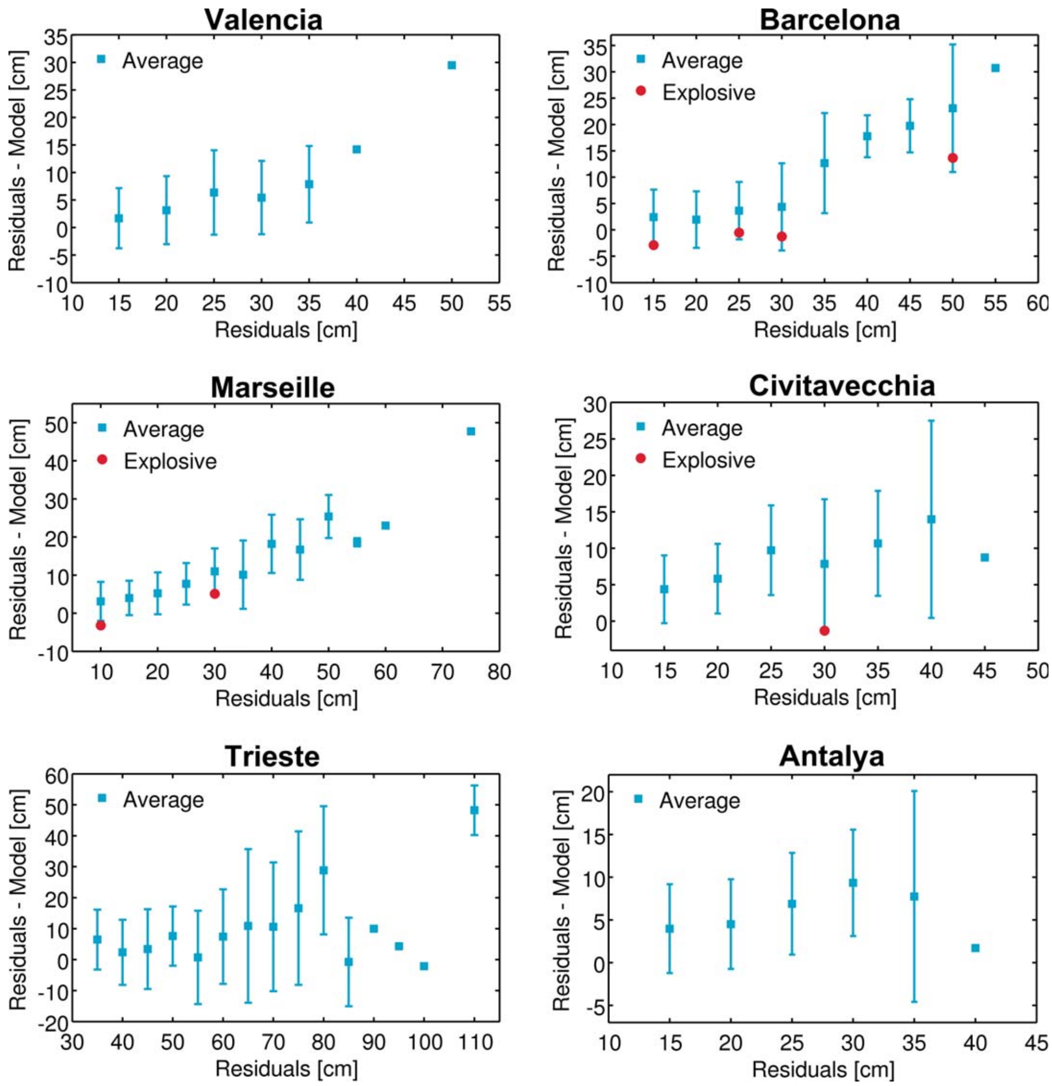

Figure 4. Mean differences in independent positive extremes (blue squares) between the residuals and the sea level from the model as a function of the extreme value in the residuals at Valencia, Barcelona, Marseille, Civitavecchia, Trieste, and Antalya. The error bars represent the standard deviation of the mean and are only shown when there are at least three values. The red dots represent the differences in extremes caused by explosive cyclones.

Finally, it is interesting to address the question of whether the differences between the residuals and the model may be in part the result of nonlinear tide-surge interactions. With the exception of the Gulf of Gabes and the North Adriatic Sea, the tidal range in the Mediterranean Sea is very small and thus, in principle, one would not expect a significant nonlinear tide-surge interaction. Indeed, *Marcos et al*. [2009] investigated the existence of such interaction at several tide gauge stations in the Mediterranean Sea and found no significant interaction at any of the stations investigated. In order to further confirm that there is not significant interaction, we have applied the same methodology as *Marcos et al*. [2009] to the six stations considered in our study. In particular, we calculated at each station the distribution of extreme events in the residuals with the phase of the dominant tidal component (taken here as the phase of M2). Then we used a χ^2 test to test the null hypothesis that such distribution is not significantly different from a flat distribution with the same number of extreme events for each phase bin. In agreement with the results of *Marcos et al*. [2009], we have found that tide-surge interaction is not significant (at the 95% confidence level) at any of the six stations considered in this study.

3.3. Return Levels and Number of Exceedances of a Threshold

In subsection 3.2, we have found that the model fails to simulate the timing and the magnitude of the sea level extremes at all tide gauge stations. Although for operational purposes having confident predictive

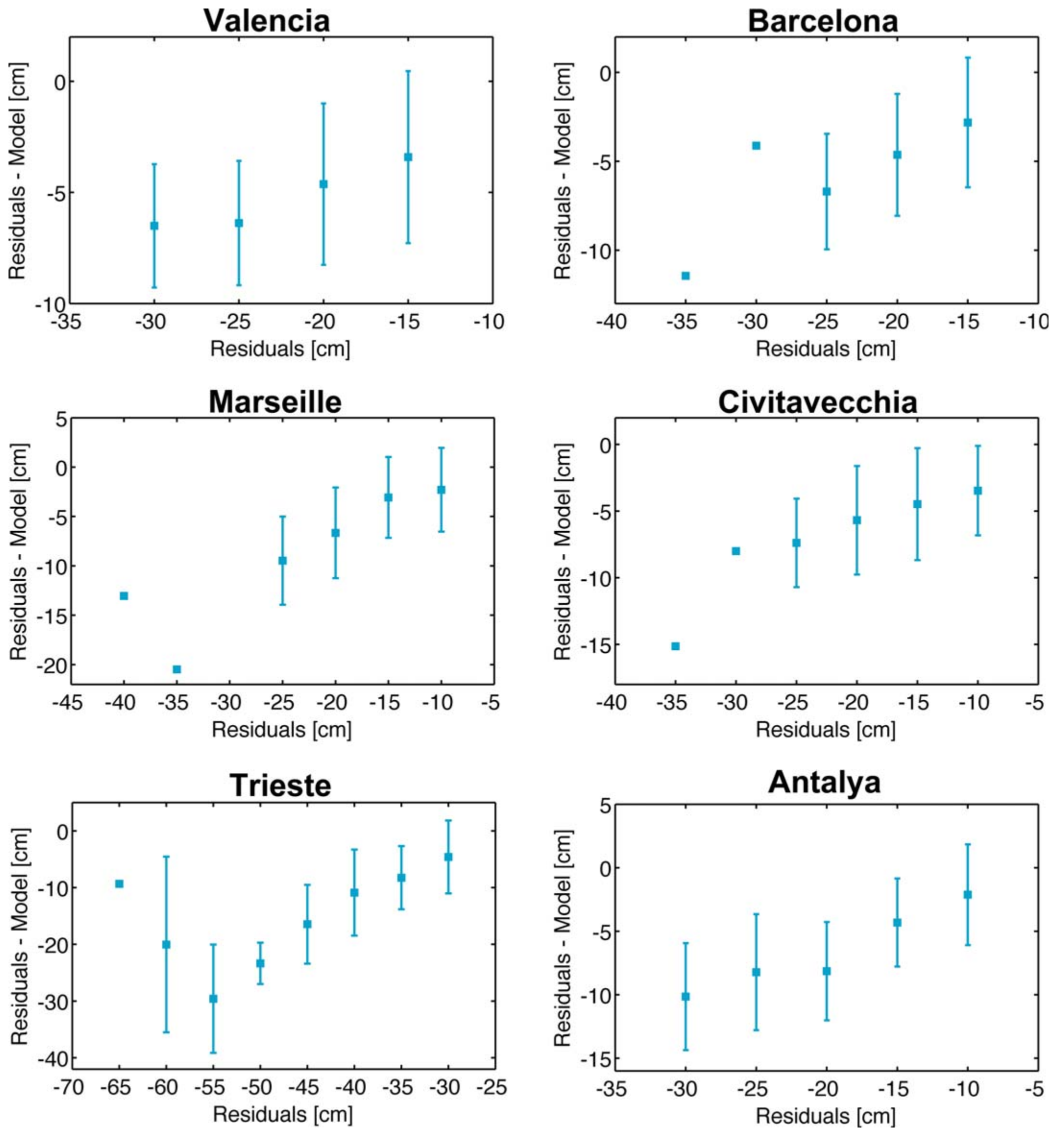

Figure 5. Same as Figure 4, but for negative sea level extremes.

capability is crucial, in climate studies it may suffice if the model correctly simulates the observed probability distribution of sea level values since we are primarily interested in climatological quantities, such as return levels and number of exceedances of a threshold over a certain period of time. Note that, as we will show later, it is possible that the model simulates the main statistical properties of the extremes correctly but fails to predict their timing and magnitude. In this section, we investigate the ability of the model to describe the main statistical properties of the observed sea level extremes. Q-q plots [*Wilk and Gnanadesikan*, 1968] of the modeled sea level versus the residuals are a good starting point for the problem at hand as they provide a convenient graphical device to assess whether or not the two data sets have the same parent distribution. In a q-q plot [*Wilk and Gnanadesikan*, 1968], if the two sets of quantiles being compared come from the same exact distribution, the points in the plot lie along the line $y = x$. If the points fall along a straight line that does not coincide with the line $y = x$ (i.e., it has different slope and/or intercept), then the distributions of the two data sets have the same shape (i.e., the same parent distribution) but they differ in scale and/or location. In this case, the model can, in principle, still be used for extreme value analysis because there is the possibility of the model being calibrated by comparison with the residuals. Finally, if the points in the q-q plot do not fall along a straight line, the two data sets come from different distributions (i.e., the distributions differ in shape) and calibration is difficult because we cannot predict how future

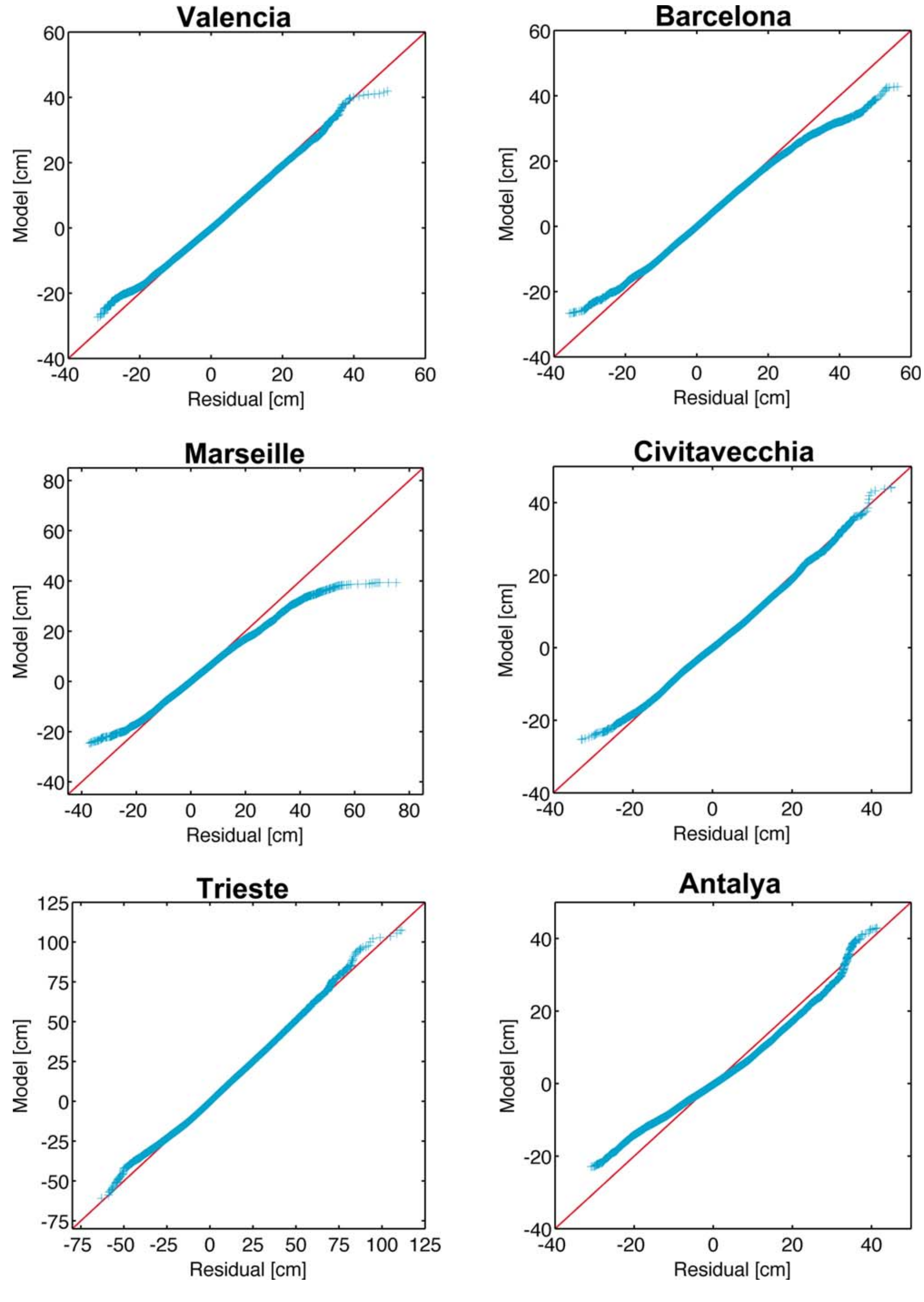

Figure 6. Q-q plots of the modeled sea level versus the residuals at Valencia, Barcelona, Marseille, Civitavecchia, Trieste, and Antalya.

changes in the observed distribution of extremes will affect the differences between quantiles, which diminishes the value of the model for extreme value analysis.

Figure 6 shows the q-q plots at all tide gauge stations. The q-q plots at Valencia, Barcelona, and Marseille are S-shaped, indicating that the distribution of modeled data at these three stations has lighter tails than that of the residuals. In other words, there is a lack of extreme values in the model relative to the residuals, which appears to be more marked for positive extremes. Note also that, at Marseille, there is a leveling off in the q-q curve for extreme values larger than 40 cm, reflecting a possible saturation of the model at this value. Because the modeled and observed distributions are different in shape (i.e., they have different

Table 1. Fifty and Ten Year Return Levels as Derived From the Residuals and the Model Together With the Number of Exceedances of the 99.9% and 99.95% Observed Percentiles Over a Period of 25 Years as Derived From the Residuals and the Modeled Sea Level[a]

	$RL_{50\ year}$		$RL_{10\ year}$		n_e of 99.9%		n_e of 99.95%	
Station	Residual	Model	Residual	Model	Residual	Model	Residual	Model
Valencia	46.8	41.8	41.9	37.1	22.7	10.0 (13.8)	17.6	7.0 (8.3)
Barcelona	66.4	43.1	54.6	38.6	14.5	1.1 (14.1)	9.1	0.2 (8.5)
Marseille	77.2	46.7	66.9	41.6	28.1	0.7 (11.8)	18.8	0.2 (8.3)
Civitave.	44.7	45.2	42.0	38.9	43.3	11.7 (13.4)	27.9	7.2 (7.48)
Trieste	117.8	126.1	102.3	103.2	52.1	39.3 (34.8)	33.8	25.3 (20.5)
Antalya	41.5	42.7	38.0	36.5	17.0	8.1 (14.2)	9.4	4.7 (4.0)

[a]The number of exceedances of the 99.9% and 99.95% modeled percentiles is also shown for the model (numbers inside the parentheses).

distributions), calibration of the model is difficult. One can, of course, use a nonlinear function to model and correct the departures from the residuals for the given q-q plot, however, there is no way to know if the departures will still follow such function if there is, for instance, a change in the scale of the distribution of the residuals (since the observed and modeled distributions are different). This should be considered when using the model at these three stations to derive climatological quantities associated with sea level extremes. The shape of the q-q plots at Civitavecchia, Trieste, and Antalya is rather different from that at the other three stations. It shows that, overall, the modeled distribution is slightly skewed to the right at these three stations, indicating a mild overestimation of the positive extreme values as well as a moderate underestimation of the negative extreme values. A closer look, however, reveals that, at Trieste, although the model indeed tends to overestimate the positive extreme values, for the largest extremes (>99.995th percentile) the model actually shows a slight underestimation. Nevertheless, the deviation of the modeled distribution from the observed distribution in the right tail is weak at the three stations, suggesting that the model can, in principle, be used for analysis of positive extremes without incurring significant error. Note, however, that the deviation is significantly larger in the left tail at Civitavecchia and Antalya, which casts doubt upon the reliability of the model to simulate the negative extremes at these two stations. Finally, it is worth noting that, except at the extremes, the modeled quantiles match the residual quantiles reasonably well at all stations, which explains why it is usually found that the correlation between modeled sea level and residuals is high, for hourly and monthly time series.

One of the uses of such a barotropic model is the projection of future changes in important climatological quantities associated with sea level extremes, such as return levels and number of exceedances of a critical threshold over a particular period of time. The success of the model in capturing these quantities is dependent upon its ability to simulate the observed probability distribution of sea level, and hence differences between distributions will inevitably cause the modeled data to yield different climatological quantities. The analysis of the q-q plots has shown significant differences between the observed and modeled distributions, especially at Valencia, Barcelona, and Marseille. The extent to which these differences are reflected in differences in return levels and number of exceedances determines the suitability of the model for sea level extreme analyses.

Table 1 shows a comparison of the 50 and 10 year return levels as well as the number of exceedances of the 99.9th and 99.95% percentiles over a period of 25 years as derived from the modeled data and the residuals at all tide gauge stations. The return levels in the residuals differ significantly from one station to another (Figure 7a). The highest 50 year return levels in the residuals are found at Trieste (117.8 cm) and Marseille (77.2 cm) whereas the lowest ones are found at Antalya (41.5 cm) and Civitavecchia (44.7 cm). As expected from the analysis of the q-q plots, the model simulates the 50 year return level reasonably well only at Valencia, Civitavecchia, Trieste, and Antalya, where differences are not larger than 8.3 cm. At Barcelona, and Marseille, however, the model underestimates the 50 year return level by 23.3, and 30.5 cm, respectively, which corresponds to an underestimation of 35%, and 40%. For the 10 year return level, the model performs again reasonably well at all stations except at Barcelona and Marseille, where the 10 year return level is underestimated by 16 and 25.3 cm, respectively.

In terms of the number of exceedances of the 99.9th and 99.95% percentiles over a period of 25 years, the model performance is worse (Table 1). It is important to note that, in practice, one is interested in knowing

10.1002/2014JC010360

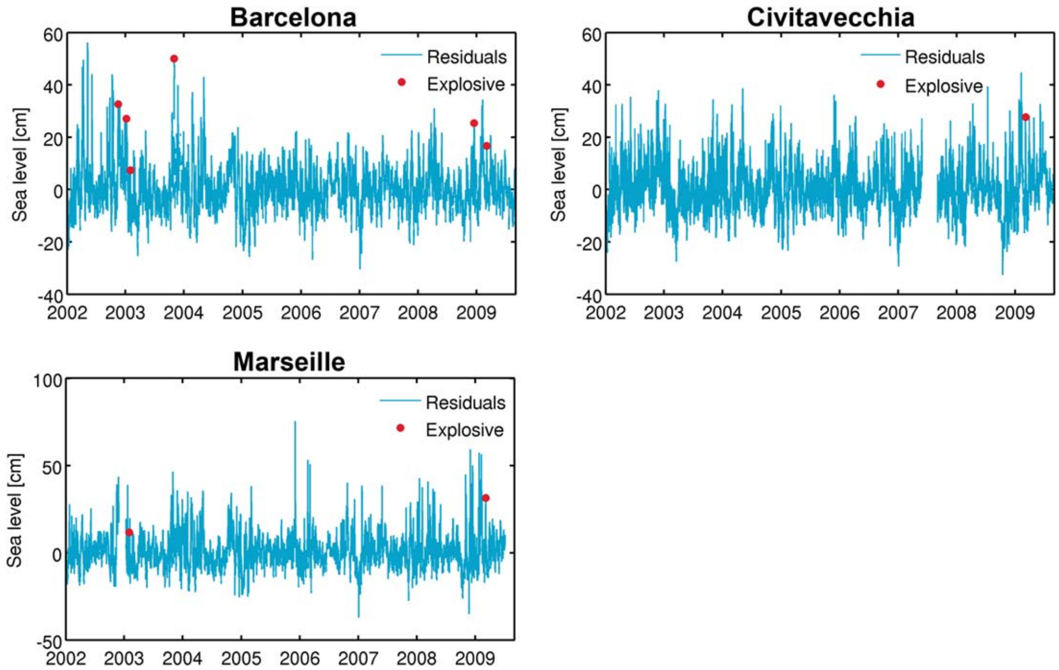

Figure 7. Time series of detided and deseasoned sea level from the tide gauge records at Barcelona, Marseille, and Civitavecchia. The red dots denote the occurrence of an explosive cyclone within a distance of 250 km from the tide gauge.

the number of exceedances of a particular threshold value irrespective of the percentile position of such threshold value. Therefore, to assess the practical value of the model one should use the same threshold for both the residuals and the model. Here we use the 99.9th percentile to select the threshold value just to ensure that the selected value for each station is in fact an extreme value. Note, however, that because the modeled and observed distributions of sea level are different (Figure 6), the modeled and observed 99.9th percentiles are expected to be different as well. If the bias between the modeled and observed percentiles is known, then the model can still be used to derive the number of exceedances provided that the number of exceedances for each percentile is the same in both data sets. Knowing the bias is, however, equivalent to being able to calibrate the model, which is difficult if the observed and modeled data have different parent distributions, as it is the case at most stations (Figure 6).

In order to obtain a full assessment of the model performance, the number of exceedances is calculated for two different cases: (i) using the same threshold value (the 99.9th and 99.95% percentiles from the residuals) for the residuals and the model; and (ii) using the 99.9th and 99.95% percentiles from each data set. For the first case (Table 1), the results from the residuals indicate that the 99.9% observed percentile is exceeded about 23 times at Valencia, 43 times at Civitavecchia, and 17 times at Antalya over a period of 25 years, whereas those from the model suggest that this percentile is exceeded only 10, 12, and 8 times, respectively. Similar large differences are found at Barcelona, Marseille, and Trieste. For the second case, the number of exceedances of the 99.9% modeled percentile as derived from the model increases at some stations, but even in this case the model underestimates such number at all stations except Barcelona, reflecting the poor performance of the model in capturing the number of exceedances. A similar underestimation is found for the number of exceedances of the 99.95% percentile (Table 1).

3.4. Sea Level Extremes Associated With Explosive Cyclones

Here we are interested in two questions. First, whether the sea level extremes associated with explosive cyclones in the Mediterranean Sea differ from those due to nonexplosive cyclones (i.e., whether they are generally larger), and second, whether the skill of the model to simulate the sea level extremes is affected by the rapid deepening rate associated with explosive cyclones. A sea level extreme is considered to be related to an explosive cyclone if the center of the cyclone is within a distance of 250 km from the tide gauge at the time of the extreme event. To address the first question, we plot the time series of the residuals at all tide gauge stations highlighting the times at which an explosive cyclone occurred in the vicinity (distance <250 km) of the station (Figure 7). It is evident from the figure that, although a few extreme values related to explosive cyclones are among the largest ones, they are not, in general, the largest events and do not differ from events due to nonexplosive cyclones. To address the second question, we compare the differences between observed and modeled extreme values for the case of explosive cyclones with average differences (Figure 4). The most interesting feature is that the differences between the residuals and the model in terms of extreme values tend to be much smaller when the extreme event is associated with an explosive cyclone, and this is the case at all stations. In particular, the differences for explosive events fall consistently at the lower end (or even outside) of the standard deviation range (denoted by the error bars in the figure).

In order to shed some light on the reasons why the model performs better for extreme events related to explosive cyclones and also why the model tends, in general, to underestimate the sea level extremes at all stations, we have forced the barotropic model with wind and atmospheric pressure from the COSMO.GR model for a period of 72 h centered in time about each of the 20 events of explosive cyclogenesis. The atmospheric fields from the COSMO.GR model have a higher spatial resolution than the AEMET forcing fields, and thus can be used to assess the impact of the spatial resolution of the atmospheric forcing in the model performance. Overall, we find that the COSMO.GR model shows stronger variability than AEMET (not shown), however this does not translate into a better simulation of the sea level extremes associated with explosive cyclones. In fact, the difference between the time series of sea level derived from the simulations forced by COSMO.GR and AEMET is rather small. This result suggests that the resolution of the atmospheric forcing may not be a leading factor in explaining the differences between the model and the residuals. This does not necessarily mean that the spatial resolution of the forcing is not important to properly simulate the sea level extremes, but rather than the spatial resolution of the numerical model itself as well as that of the bathymetry may play a more important role in explaining the differences. Note, however, that this applies only to the few sea level extremes linked to explosive cyclones but it is not clear if this is also true for nonexplosive extremes. Clearly, further investigation is required to confirm these conclusions, although this is beyond the scope of this study.

4. Conclusions and Discussion

In this study, we have explored the skill of a barotropic model to simulate sea level extremes of meteorological origin in the Mediterranean Sea. In particular, our goal has been to assess how well the model can simulate the timing and the magnitude of sea level extremes as well as the observed distribution of sea level extremes. We have also addressed the question whether explosive cyclones lead to larger sea level extremes and have investigated the effect of the rapid deepening associated with such cyclones on the model performance. Our original expectation was that the high rate of atmospheric pressure change associated with explosive cyclones may degrade the model performance for storm surge modeling due to restrictions on the barotropic flow imposed by the presence of the Gibraltar Strait and other straits in the Mediterranean basin.

We have first quantified the differences between the sea level from the model and the residuals from six tide gauge records as a function of the extreme value. The model underestimates the positive extremes significantly at all stations except at Trieste, in some cases by up to 65%. At Trieste, the model can sometimes underestimate the extremes by up to 45% but it can also sometimes overestimate them and even generate extremes when none is observed in the residuals. The differences between the model and the residuals are not constant for extremes of a given height but they are spread out over a wide range of values. Spreads as large as 65% of the extreme value are found at several stations, and they are larger than 30% at all stations. In addition, the model underestimates the value of the negative extremes at all stations and the differences for extremes of a similar magnitude show also large spreads.

The magnitude of the differences varies rather randomly. This limits the applicability of the numerical model for storm surge forecasting because adjustment of the model is not straightforward. The reasons for these differences are not clear. Two model runs driven by atmospheric data with different resolutions of 25 and 7 km have shown only slight differences in terms of extremes, suggesting that other factors, rather than the resolution of the atmospheric forcing, may be more important in explaining the differences between the model and the residuals. The differences may also be caused by coastal processes which are not resolvable under the model resolution or poor representation of the bathymetry. Further work is needed to resolve the causes and improve the model performance with respect to extremes.

In the second part of the paper, we have explored the ability of the model to capture the observed distribution of sea level extremes. An analysis of q-q plots of modeled versus observed data has shown significant differences in the shape of the distribution between the two data sets at all stations. The differences are larger at Valencia, Barcelona, and Marseille. The fact that the modeled and observed distributions differ in shape at these three stations indicates that such distributions are not linearly related, rendering the calibration of the model complicated. A possibility would be to use a nonlinear function to model and correct the departures from the residuals as shown by the q-q plots, however, because the observed and modeled distributions differ in shape, there is no guarantee that such function will still be valid if sea level extremes change in the future. Finally, the variability of the function between the explored stations indicates that an overall correction for the whole model will be spurious because each location is likely to have a different correction factor.

The 50 and 10 year return level estimates from the model and the residuals have differences smaller than 8.3 cm at Valencia, Civitavecchia, Trieste, and Antalya. At Barcelona, and Marseille, the model underestimates the 50 year return level by 23.3, and 30.5 cm, respectively, and the 10 year return level by 16 and 25.3 cm, reflecting a significantly lower skill of the model at those stations. The number of exceedances of the 99.9th and 99.95% percentiles over a period of 25 years is significantly underestimated by the model at all stations. It is clear from these results that looking at the 50 year return level only, as it is sometimes done, does not provide a good measure of the capability of the model to capture the observed distribution of sea level extremes but it is still the most reliable parameter of the model with respect to extremes.

Finally, in this study, we have also shown that sea level extremes caused by explosive cyclones are not larger than those caused by nonexplosive cyclones. Interestingly, we have also found that the skill of the model for predicting the timing and value of the storm surges is consistently and significantly higher for the events associated with explosive cyclones at all stations. The reasons for this better performance for rapid changes of atmospheric pressure may be linked to the fact that such phenomena, having subbasin spatial scales, depend on redistribution of water within the basin, which are not constrained by the effects of the Gibraltar Strait on water mass exchanges at high frequencies [*Garrett and Majaess*, 1984; *Lascaratos and Gacic*, 1990; *Tsimplis and Vlahakis*, 1994]. Our results show that the skill of the numerical model with respect to extremes is lower than what can be inferred from previous studies. Except perhaps for the 50 year return periods (at some stations), the skill of the model is in bad need of improvement as it significantly underestimates extremes in most places. This study provides evidence for a pressing need for further work to identify the reasons for the departures of the modeled data from observations, and the enhancement of modeling of sea level extremes in the region.

Acknowledgments
F.M. Calafat was supported under a Marie Curie International Outgoing Fellowship (IOF) within the 7th European Community Framework Programme (grant agreement PIOF-GA-2010–275851). G. Jordà acknowledges a postdoctoral grant from the Conselleria d'Educació, Cultura i Universitats del Govern de les Illes Balears, and the European Science Foundation. This research would not have been possible without the tide gauge records from the Puertos del Estado tide gauge network (REDMAR), REFMAR (http://refmar.shom.fr), and IDROMARE (http://www.idromare.it/) databases. The authors also thanks Agencia Española de Meteorología (AEMET) for providing the atmospheric fields used to force the model. The authors wish to thank two anonymous reviewers for their constructive comments on an earlier version of this paper.

References

Avgoustoglou, E., and T. Tzeferi (2013), The effect of a sub-grid statistical cloud-cover scheme applied to the COSMO local numerical weather prediction model over the wider geographical domain of Greece, *Atmos. Res.*, *152*, 69–73, doi:10.1016/j.atmosres.2013.11.003.

Conte, D., and P. Lionello (2013), Characteristics of large positive and negative surges in the Mediterranean Sea and their attenuation in future climate scenarios, *Global Planet. Change*, *111*, 159–173.

Dee, D. P., et al. (2011), The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Q. J. R. Meteorol. Soc.*, *137*, 553–597, doi:10.1002/qj.828.

Feng, X., and M. N. Tsimplis (2014), Sea level extremes at the coasts of China, *J. Geophys. Res. Oceans*, *119*, 1593–1608, doi:10.1002/2013JC009607.

Garrett, C., and F. Majaess (1984), Nonisostatic response of sea level to atmospheric pressure in the Eastern Mediterranean, *J. Phys. Oceanogr.*, *14*, 656–665.

Haigh, I. D., L. R. MacPherson, M. S. Mason, E. M. S. Wijeratne, C. B. Pattiaratchi, R. P. Crompton, and S. George (2014), Estimating present day extreme water level exceedance probabilities around the coastline of Australia: Tropical cyclone-induced storm surges and mean sea level, *Clim. Dyn.*, *42*(1–2), 121–138.

Hallegatte, S., C. Green, R. J. Nicholls, and J. Corfee-Morlot (2013), Future flood losses in major coastal cities, *Nat. Clim. Change*, *3*, 802–806.

Jordà, G., D. Gomis, E. Álvarez-Fanjul, and S. Somot (2012a), Atmospheric contribution to Mediterranean and nearby Atlantic sea level variability under different climate change scenarios, *Global Planet. Change*, *80–81*, 198–214.

Jordà, G., D. Gomis, and E. Álvarez-Fanjul (2012b), The VANI2-ERA hindcast of sea level residuals: Atmospheric forcing of sea level variability in the Mediterranean Sea (1958–2008), *Sci. Mar.*, *76*, 133–146.

Kouroutzoglou, J., H. A. Flocas, K. Keay, and M. D. Hatzaki (2011a), Climatological aspects of explosive cyclones in the Mediterranean, *Int. J. Climatol.*, *31*, 1785–1802, doi:10.1002/joc.2203.

Kouroutzoglou, J., H. A. Flocas, K. Keay, and M. D. Hatzaki (2011b), Assessing the characteristics of explosive cyclones under different resolution, *Theor. Appl. Climatol.*, *105*, 263–275, doi:10.1007/s00704-010-0390-8.

Kouroutzoglou, J., H. A. Flocas, K. Keay, I. Simmonds, and M. D. Hatzaki (2014), A high resolution climatological study on the comparison between surface explosive and ordinary cyclones in the Mediterranean, *Reg. Environ. Change*, *14*(5), 1833–1846, doi:10.1007/s10113-013-0461-3.

Lascaratos, A., and M. Gacic (1990), Low frequency sea level variability in the northeastern Mediterranean, *J. Phys. Oceanogr.*, *204*, 522–533.

Lionello, P., J. Bhend, A. Buzzi, P. M. Della-Marta, S. Krichak, A. Jansá, P. Maheras, A. Sanna, I. F. Trigo, and R. Trigo (2006), Cyclones in the Mediterranean region: Climatology and effects on the environment, in *Mediterranean Climate Variability*, edited by P. Lionello, P. Malanotte-Rizzoli, and R. Boscolo, pp. 324–372, Elsevier, Amsterdam.

Marcos, M., M. N. Tsimplis, and A. G. P. Shaw (2009), Sea level extremes in southern Europe, *J. Geophys. Res.*, *114*, C01007, doi:10.1029/2008JC004912.

Marcos, M., G. Jordà, D. Gomis, and B. Pérez (2011), Changes in storm surges in southern Europe from a regional model under climate change scenarios, *Global Planet. Change*, *77*, 116–128.

Marcos, M., G. Chust, G. Jordà, and A. Caballero (2012), Effect of sea level extremes on the western Basque coast during the 21st century, *Clim. Res.*, *51*, 237–248.

Menéndez, M., and P. L. Woodworth (2010), Changes in extreme high water levels based on a quasi-global tide-gauge data set, *J. Geophys. Res.*, *115*, C10011, doi:10.1029/2009JC005997.

Nicholls, R. J. (2010), Impacts of and responses to sea-level rise, in *Understanding Sea-Level Rise and Variability*, edited by J. A. Church et al., pp. 17–51, Wiley-Blackwell, Chichester, U. K.

Pawlowicz, R., B. Beardsley, and S. Lentz (2002), Classical tidal "Harmonic Analysis Including Error Estimates in MATLAB using T_TIDE," *Comput. Geosci.*, *28*, 929–937.

Pickands, J., III (1975), Statistical inference using extreme order statistics, *Ann. Stat.*, *3*, 119–131, doi:10.1214/aos/1176343003.

Ratsimandresy, A. W., M. G. Sotillo, E. Álvarez-Fanjul, J. C. Carretero Albiach, B. Pérez Gómez, and H. Hajji (2008), A 44-year (1958–2001) sea level residual hindcast over the Mediterranean Basin, *Phys. Chem. Earth*, *33*, 250–259.

Sterl, A., H. van den Brink, H. de Vries, R. Haarsma, and E. van Meijgaard (2009), An ensemble study of extreme storm surge related water levels in the North Sea in a changing climate, *Ocean Sci.*, *5*, 369–378.

Tsimplis, M. N., and D. Blackman (1997), Extreme sea-level distribution and return periods in the Aegean and the Ionian Seas, *Estuarine Coastal Shelf Sci.*, *44*, 79–89, doi:10.1006/ecss.1996.0126.

Tsimplis, M. N., and A. G. P. Shaw (2010), Seasonal sea level extremes in the Mediterranean Sea and at the Atlantic European coasts, *Nat. Hazard Earth Syst.*, *10*(7), 1457–1475.

Tsimplis, M. N., and G. N. Vlahakis (1994), Meteorological forcing and sea level variability in the Aegean Sea, *J. Geophys. Res.*, *99*(C5), 9879–9890, doi:10.1029/94JC00479.

Wilk, M. B., and R. Gnanadesikan (1968), Probability plotting methods for the analysis of data, *Biometrika*, *55*(1), 1–17.

Woodworth, P. L., R. A. Flather, J. A. Williams, S. L. Wakelin, and S. Jevrejeva (2007), The dependence of UK extreme sea levels and storm surges on the North Atlantic Oscillation, *Cont. Shelf Res.*, *27*(7), 935–946.

Woth, K., R. Weisse, and H. von Storch (2006), Climate change and North Sea storm surge extremes: An ensemble study of storm surge extremes expected in a changed climate projected by four different regional climate models, *Ocean Dyn.*, *56*, 3–15.