A required data span to detect sea level rise

T. Niedzielski
Space Research Centre, Polish Academy of Sciences, Warsaw,
Poland Institute of Geography and Regional Development, University of Wrocław, Wrocław, Poland

W. Kosek
Space Research Centre, Polish Academy of Sciences, Warsaw, Poland

ABSTRACT: Altimetric measurements indicate that the global sea level rises about 3 mm/year, however, in various papers different data spans are adopted to estimate this value. The minimum time span of TOPEX/Poseidon (T/P) and Jason-1 (J-1) global sea level anomalies (SLA) data required to detect a statistically significant trend in sea level change was estimated. Seeking the trend in the global SLA data was performed by means of the Cox-Stuart statistical test. This test was supported by the stepwise procedure to make the results independent of the starting data epoch. The probabilities of detecting a statistically significant trend within SLA data were computed in the relation with data spans and significance levels of the above-mentioned test. It is shown that for the standard significance level of 0.05 approximately 5.5 years of the SLA data are required to detect a trend with the probability close to 1. If the seasonal oscillations are removed from the combined T/P and J-1 SLA data, 4.3 years are required to detect a statistically significant trend with a probability close to 1. The estimated minimum time spans required to detect a trend in sea level rise are addressed to the problem of SLA data predictions. In what follows, the above-mentioned estimate is assumed to be minimum data span to compute the representative sample of SLA data predictions. The forecasts of global mean SLA data are shown and their mean prediction errors are discussed.

1 INTRODUCTION

Climate change studies are usually associated with seeking variation rates of various elements of the environment. Among others, the sea level rise reflects current global climatic changes as it is caused by complex interactions between the solid Earth, atmosphere, oceans, hydrosphere and cryosphere. Thus, the rate of sea level rise may be used as an indicator of the global environmental changes and can be extrapolated in order to build future scenarios.

Sea level rises as a result of several natural processes acting in the global environment, which can be classified into three main groups: geological, eustatic and steric effects (e.g. Dobrovolski 2000). The first group concerns the processes, which make the ocean basins and the coasts change their parameters. The main changes of this type are associated with orogenic movements, spreading, sedimentation, tectonics, subsidence of the sea floor and the post-glacial rebound. The second group of processes is connected to the climate itself as being forced by the increase of water mass of the oceans. In fact, the eustatic changes are mainly derived from melting the ice-sheets and glaciers. Finally, the steric effect is connected to the increase in the water volume without the change in its mass. This is largely caused by the thermal expansion of the water in the oceans as a result of the increase in the global sea surface temperature.

The considerable number studies focus on the determination of the rate of sea level change and these estimates differ due to the wide spectrum of methods and data sets applied. The trends are being usually fitted by the least-squares or the robust techniques. The data on the sea surface topography may be measured both relatively to the Earth crust (tide gauges) or absolutely (satellite altimetry). For instance, Douglas (1991) analysed the tide gauge data and found that a good approximation of the rate in question is of 1.8 ± 0.38 mm/year. In contrast, Leuliette et al. (2004) used the recent and precise sea level anomaly (SLA) time series obtained from the satellite altimetry TOPEX/Poseidon (T/P) and Jason-1 (J-1) and argued that the rate of sea level change was of 2.8 ± 0.4 mm/year. If no Jason-1 date is considered, the discussed trend computed by a robust procedure is equal to 1.46 mm/year (Kosek 2001).

Various data sets on the sea level variation are usually of dissimilar lengths and hence may be sparse. The vast majority of the estimates in question is based upon fitting a trend without much concern whether it is statistically significant. Hence, there is a need to reverse the problem and not to estimate the trend itself but, in turn, to estimate the data span which is required to detect a statistically significant trend. The practical usage of such estimates follows from the SLA prediction studies. In what follows, in order to construct the representative sample of SLA forecasts, one needs to fix arbitrarily the first starting prediction epoch. If one knows the minimum time span of the SLA data to detect a statistically significant trend in them (which is the main and the most straightforward component for extrapolation), it is assumed to be the first starting prediction epoch.

The method for seeking the above-mentioned estimates was proposed by Niedzielski & Kosek (2006) and presented first at the General Assembly of the European Geosciences Union in Vienna in April 2006. The results gained using this simulation-based statistical technique (Niedzielski & Kosek, submitted) are applied in this article to support the evaluation of the prediction results obtained by different forecasting techniques. Thus, this paper aims to combine the SLA predictions with the detailed analysis of the rate of sea level rise.

2 METHODS

2.1 *Estimation of minimum data span for prediction*

According to Niedzielski & Kosek (submitted), the minimum time span of SLA data required to detect a statistically significant trend in sea level rise can be estimated using the statistical simulation based upon the Cox-Stuart test (McCuen 2003). This statistical test is designed to test for the existence of an upward and/or downward trend within the time series. For the analysis of sea level change it is straightforward to focus only on upward trends. If the latter applies, the null hypothesis assumes that there does not exist a trend in the time series, whereas the alternative hypothesis assumes an upward trend in the underlying data. In general, the idea behind the Cox-Stuart technique is simple. It is based on subdividing the time series x_t of size n (even number) into two smaller data sets. The first one is comprised of the first $n/2$ data and the second one consists of the remaining $n/2$ elements of the initial time series. If n is an odd number, the middle data point is excluded from the study and hence n should be replaced by n-1. The objective of the subsequent statistical analysis is to compare these two data sets using the 0-1 random variable defined by

$$N_i = \begin{cases} 1 & if \quad a(i) < b(i), \\ 0 & if \quad a(i) > b(i), \end{cases} \tag{1}$$

where $a_t = (x_1, \dots, x_k)$, $bt = (x_m, \dots, x_n)$ and $k = n/2$ and $m = (n/2) + 1$ (if n is an even number); $k = (n-1)/2$ and $m = (n+3)/2$ (if n is an odd number). Hence, the random variable

$$T = \sum N_i \tag{2}$$

counts the number of elements of the second time series b_t being greater than the corresponding elements in the first data set a_t. The probability law of T is binomial $b(l,p)$, where l is a number of a_t (or equivalently b_t) elements. The null hypothesis stated before may be expressed in terms of N_i as the equal amount of zeros and ones. Thus, under the null hypothesis the probability distribution of T is $b(l, 1/2)$. Testing the hypothesis of no trend in sea level rise is based upon T values and hence – as a result of the alternative hypothesis definition (upward trend) – the upper tail of the probability distribution is considered.

In order to make the analysis independent of the specific starting data epoch it is convenient to apply the simulation (Niedzielski & Kosek, submitted). In what follows, one ought to test the above-mentioned hypothesis for various subsets of a given SLA time series. To do this, one fixes the small positive integer t and defines the data block of size t. Subsequently, one moves the block forward and applies the Cox-Stuart test for the new subset of data of size t. The procedure should be repeated $N-t+1$ times. This allows the computation of the probability of detecting the trend after the time t as

$$p^{(s)}(t) = \frac{\#\{j : p_t(j) < s\}}{N - t + 1}, \tag{3}$$

where $p_t(j)$ is a p-value of the Cox-Stuart test for the j-th location of the block of size t within the entire time series and s is a significance level. The subsequent analysis is based on the stepwise algorithm that performs the above-mentioned analysis by increasing t in each step.
Thus, the required time to detect a trend in sea level rise is a minimum t for which the probability given by the equation (3) is close to 1. This time may support forecasting SLA data as t can be assumed to be a minimum number of data points.

2.2 *Prediction techniques*

The variety of forecasting methods is big and hence one may consider both uni- and multivariate prediction techniques of linear and non-linear structures. In this paper we apply the most straightforward time series tools, i.e. fitting and extrapolating the harmonic-polynomial deterministic model (LS), autoregressive stochastic modelling and prediction (AR) and multivariate autoregressive stochastic modelling and prediction (MAR). The LS approach is used as a preprocessing tool and thus it estimates and subsequently extrapolates well-known oscillations and trends. The residuals from the fitted LS models are being modelled and predicted using the above-mentioned stochastic approaches.
The LS model can be denoted as

$$f_t = \sum_{k=1}^{S} A_k \sin(\omega_k t + \varphi_k) + Bt + \gamma \tag{4}$$

where $A_1,\ldots, A_s$, B and $\varphi_1,\ldots, \varphi_s$, γ have to be estimated by least-squares algorithm and $\omega_1,\ldots, \omega_s$ are known frequencies. There are two objectives of LS modelling. First, the LS model can be extrapolated to obtain the deterministic prediction. Second, the LS model may be applied to calculate the residuals, which can be subsequently modelled and predicted by stochastic methods.

The AR technique is a simple stochastic method used to build a model for stationary residuals. The AR approach is based upon the following equation

$$Y_1 = c_1 Y_{t-1} + \ldots + c_p Y_{t-p} + \varepsilon_t, \tag{5}$$

where Y_t is the residual time series obtained as the difference between the data and the LS model; p is the order of the autoregressive process; $c_1,\ldots, c_p$ are the autoregressive coefficients; and ε_t is the white noise (e.g. Brockwell & Davis 1996). The order p is being usually chosen by the Akaike Information Criterion (AIC) and the autoregressive coefficients are estimated using the combination of Yule-Walker and maximum likelihood methods.

The MAR method is a multivariate extension of the AR technique. The MAR process is defined by the following equation

$$\mathbf{Y}_t = \mathbf{M}_1\mathbf{Y}_{t-1} + \ldots + \mathbf{M}_p\mathbf{Y}_{t-p} + \mathbf{E}_t, \tag{6}$$

where $\mathbf{Y}_t$ is a vector of stationary residuals computed at each axis as the difference between the data and the corresponding LS model; p is the order of the multivariate autoregressive process; $\mathbf{M}_1,\ldots\mathbf{M}_p$ are the coefficient matrices for multivariate autoregression; and $\mathbf{E}_t$ is the white noise vector with mean 0 and covariance matrix $\mathbf{C}$ (e.g. Reinsel 1997). The common method for order selection is a Schwarz Bayesian Criterion (SBC). The estimation of the coefficient matrices is performed by the stepwise LS algorithm (Neumaier & Schneider 2001).

For the prediction equations we relate to Brockwell & Davis (1996) and Reinsel (1997) or more specifically in the field of satellite geodesy to Niedzielski & Kosek (2005). Three prediction approaches are used, i.e. LS – extrapolation of the LS deterministic polynomial-harmonic model; LS+AR – combination of the LS extrapolation and the AR prediction of residuals; LS+MAR – combination of the LS extrapolation and the MAR prediction of vector residuals.

The verification of the computed predictions is performed by the analysis of root mean square error (RMSE) defined as

$$RMSE(L) = \left(\frac{1}{d} \sum_{i=1}^{d} \left(PX_{t(i)+L} - X_{t(i)+L} \right)^2 \right)^{1/2}, \tag{7}$$

where $X_{t(i)+L}$ are the data at the time $t(i)+L$, $PX_{t(i)+L}$ is the L-step prediction of the data.

3 DATA

The modelling and prediction is based upon the two data sets. The first one is the global mean SLA data and hence corresponds to the sea level change. The second time series is a mean global sea surface temperature (SST) and corresponds to the physical description of the steric effect.

The SLA data are obtained from T/P and J-1 satellite altimetry. In fact, the SLA itself is the difference between sea surface height (SSH) computed in respect to the reference ellipsoid and the mean sea level computed in respect to the geoid JGM-3. Altimetric measurements are absolute, i.e. the sea level fluctuations are not mixed with vertical land movements. The T/P and J-1 satellites are providing the data on SSH every 1 cycle which is equal to 9.9140625 days. In this study we use the T/P global mean SLA time series measured in the period 01.01.1993 – 01.08.2002, which corresponds to the T/P cycles No 12-364. As regards J-1 global mean SLA data, the period 04.02.2002 – 14.07.2003 is chosen, i.e. the J-1 cycles No 3-56 are considered. In this study, both 10-day and monthly global mean SLA are analysed. As the T/P and J-1 time series overlap in time, both data sets are combined and bias correction between the two is introduced. The combined T/P and J-1 data set exhibits an upward trend. The most energetic oscillations in the data are annual and semi-annual seasonal components.

The SST data are NOAA OI.v2 SST monthly fields derived by the weakly optimum interpolation. These gridded fields are averaged over the entire ocean in order to obtain the global mean SST data. The analysed time period coincides with the time period of the analysed SLA data, i.e. 01.01.1993 – 14.07.2003. The most energetic oscillations in the global mean SST data are – similarly to the global mean SLA data – annual and semi-annual components. However, the strength of the semi-annual oscillation in global mean SST data is relatively greater than the strength of the semi-annual seasonality in the global mean SLA data. Besides, there is no trend within global mean SST time series.

4 RESULTS

4.1 *Estimates of the minimum data time span*

Following Niedzielski & Kosek (submitted), seeking the minimum time span to detect the statistically significant trend in sea level rise may be subdivided into two parts. First, the global mean SLA data are being processed. Second, the non-seasonal global mean SLA data are considered. If the latter applies, the global mean SLA time series should be pre-processed by removing annual and semi-annual components. The removal of these oscillations allows the analysis of the linear trend itself and stochastic fluctuations. Figure 1 shows the probability of detecting the statistically significant trend in both seasonal and non-seasonal global mean SLA data. The probability is dependent on time. Hence, the time for which the probability reaches 1 is assumed to be the minimum required data span. For the standard significance level of 0.05 the estimates are equal to 5.5 years (the analysis for seasonal global mean SLA data) or 4.3 years (for non-seasonal global mean SLA time series). As noted earlier, the trend belongs to the key deterministic components within the studied data and hence – in order to extrapolate it – one needs to know the data span which guarantees the statistical significance of the model. Thus, considering the minimum data time span of 4.39 years we assume the cycles No 162 and No 53 (for SLA data with 1 cycle and 1 month sampling interval, respectively) to be the first starting prediction points.

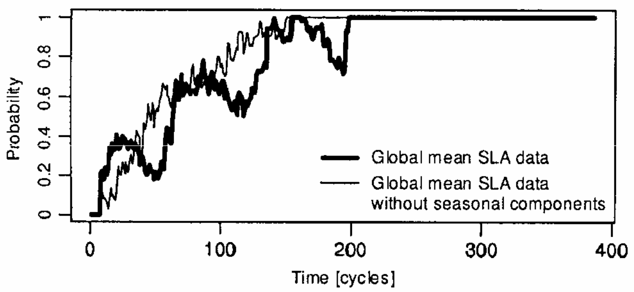

Fig. 1. The probability of detecting the trend in sea level rise as a function of time (T/P and J-1 cycles) for a standard significance level of 0.05

4.2 *Prediction of global mean sea level anomalies*

In accordance with the above-mentioned prediction methods we apply LS, LS+AR and LS+MAR prediction procedures. In the LS and LS+AR cases, the predictions are based on the past of global mean SLA data. However, the MAR approach is applied to combine both global mean SLA data with global mean SST data in order to consider the contribution from the steric effect as the explanatory variable. The deterministic LS modelling of the data is based on the equation (4) in the following way: for the global mean SLA data we model annual, semiannual oscillations and the trend, whereas for the global mean SST data we consider annual and semiannual oscillations. As the SST data are monthly, the LS+MAR analysis is only performed for the data with the sampling interval of 1-month.

Table 1. Basic statistics (in cm) for the LS predictions of global mean SLA data.

Statistics	Length of prediction			
	2-month	6-month	1-year	1.5-year
For data in cycles				
Maximum	1.829	1.752	1.803	2.000
RMSE	0.485	0.503	0.589	0.704
For data in months				
Maximum	0.960	1.179	1.123	1.118
RMSE	0.338	0.403	0.459	0.577

Tables 1-3 present both (1) maximum absolute values of the difference between global mean SLA data and their predictions, (2) values of RMSE. In general, the comparison of the values of these statistics shows that the predictions of monthly SLA data are significantly more accurate than the predictions of the SLA data with the sampling interval of 1 cycle. The interpretation is straightforward and follows form smoothing of time series. Indeed, the monthly data are essentially smoothed relatively to the data in cycles due to the time-averaging procedure. Thus, the extremes which exist within the time series recorded in cycles are eliminated (or lessened) in the process of smoothing. In fact, predicting the extremes is usually difficult and introduces the considerable error. In the case of absence of the extremes, the predictions often work well.

Table 2. Basic statistics (in cm) for the LS+AR predictions of global mean SLA data.

Statistics	Length of prediction			
	2-month	6-month	1-year	1.5-year
For data in cycles				
Maximum	1.953	1.827	1.807	2.002
RMSE	0.496	0.524	0.608	0.707
For data in months				
Maximum	0.901	1.229	1.030	1.121
RMSE	0.340	0.441	0.476	0.580

It is difficult to address the issue of comparison of the calculated predictions. In fact, the analysis of maximum absolute values of the difference between the data and their predictions and RMSE indicate that all selected procedures lead to the forecasts of similar accuracy (Tab. 1-3).

Table 3. Basic statistics (in cm) for the LS+MAR predictions of global mean SLA data.

Statistics	Length of prediction			
	2-month	6-month	1-year	1.5-year
For data in months				
Maximum	1.147	1.122	1.079	1.044
RMSE	0.371	0.411	0.451	0.531

One should suspect that the application of multivariate time series analysis would improve the predictions. This is true only for the long-term (1.5year) forecasts. In this case the improvement is of order 0.5 mm RMSE and hence is rather insignificant.

5 CONCLUSIONS

The required time span of global mean SLA data to detect a statistically significant trend in them is found to be 4.3 years. This estimate is utilized in this paper to find the minimum data span for forecasting these data. The comparison results in the conclusion that the performances of these three approaches are similar.

ACKNOWLEDGEMENTS

This paper was supported by the Polish Ministry of Education and Science under the project No 4 T12E 039 29.The authors thank the Center for Space Research, University of Texas at Austin, USA for providing T/P and J-1 data. The National Centers for Environmental Prediction, USA are also acknowledged for NOAA OI.v2 SST monthly data provided in ASCII format by the Data Support Section of the Scientific Computing Division at the National Center for Atmospheric Research. NCAR is supported by grants from the National Science Foundation. The authors of R 2.1.1., a Language and Environment 2004 (GNU General Public License) and the additional packages are also acknowledged.

REFERENCES

Brockwell, P.J. & Davis, R.A., 1996. *Introduction to time series and forecasting*. New York: Springer.

Dobrovolski, S.G., 2000. *Stochastic Climate Theory: Models and Applications*. Berlin – Heidelberg: Springer.

Douglas, B.C., 1991. Global sea level rise. *Journal of Geophysical Research* 96: 6981-6992.

Kosek, W., 2001. Long-term and short period global sea level changes from TOPEX/Poseidon altimetry. *Artificial Satellites* 36: 71-84.

Leuliette, E.W., Nerem, R.S. & Mitchum, G.T., 2004. Calibration of TOPEX/Poseidon and Jason Altimeter Data to Construct a Continuous Record of Mean Sea Level Change. *Marine Geodesy* 27: 79-94.

McCuen, R.H., 2003. *Modeling Hydrologic Change: Statistical Methods*. Boca Raton, London, New York, Washington, D.C.: Lewis Publishers.

Neumaier, A. & Schneider, T., 2001. Estimation of parameters and eigenmodes of multivariate autoregressive models, *ACM Transactions on Mathematical Software* 27: 27-57.

Niedzielski, T. & Kosek, W., 2005. Multivariate stochastic prediction of the global mean sea level anomalies based on TOPEX/Poseidon satellite altimetry. *Artificial Satellites* 40: 185-198.

Niedzielski, T. & Kosek, W., 2006. A minimum time span of TOPEX/Poseidon and Jason-1 global sea level anomalies data required for trend determination and the multivariate autoregressive forecast of these data. *Geophysical Research Abstracts* 8, 2006, European Geosciences Union, abstract EGU-06-A-04198.

Niedzielski, T. & Kosek, W. Minimum time span of TOPEX/Poseidon and Jason-1 global altimeter data to detect the significant trends in sea level change. Submitted to *Advances in Space Research*, under review.

Reinsel, G.C., 1997. *Elements of multivariate time series analysis*. Berlin, Heidelberg, New York: Springer.