Seasonal autoregressive modeling of a skew storm surge series

Jérôme Weiss (1,2), Pietro Bernardara (1,2), Marc Andreewsky (1), Michel Benoit (1,2)

(1) EDF R&D Laboratoire National d'Hydraulique et Environnement (LNHE), Chatou
(2) Université Paris-Est, Saint-Venant Laboratory for Hydraulics, ENPC, EDF R&D, CETMEF, Chatou

Colloque MODNAT
10 octobre 2013
Outline

- Context
- Methods
- Data
- Results
- Conclusions
Context

Many applications in ocean and coastal research and engineering require a precise description of several characteristics of the sea level process.

Information on the correlation length.

Need for reliable forecasts:
- Numerical models/meteorological models = surge forecasts up to 48h.
- But they are computationally costly demanding and require a lot of input data.
- Surge forecasts based only on a limited set of past observations could be an interesting alternative.

Simulation: to obtain plausible artificial series.
- Analysis of the simulated extremes.

AR models can produce valuable information on these points.
Sea level and surges

**Sea level**: superposition of two main phenomena
- Astronomic tide (deterministic): gravitational effects of the Moon and the Sun mainly
- Surge (stochastic): influenced by meteorological conditions (wind and atmospheric pressure), and local bathymetry when approaching the shore

Uncertainties are located in the surge process

![Graph showing sea level decomposition from 10/13/1987 to 10/19/1987](image)

Great October storm of 1987

6 days
Skew surges

- Highest sea levels are likely to occur around times of high tide
  - Storm surges are particularly damaging when they occur at the time of a high tide

- **Skew surge**: difference between maximum sea level observations around the time of theoretical high tide and high tide predictions [Simon, 2007]

- Temporal resolution: every 12.4 hours
Methods
Normal Inverse Gaussian (NIG) distribution

\[
X \sim NIG(\alpha, \beta, \mu, \delta) \leftrightarrow f(x, \alpha, \beta, \mu, \delta) = \frac{\alpha \delta}{\pi \sqrt{\delta^2 + (x - \mu)^2}} \exp(\delta \sqrt{\alpha^2 - \beta^2} + \beta (x - \mu)) K_1(\alpha \sqrt{\delta^2 + (x - \mu)^2})
\]

\(\mu\) = location parameter
\(\delta\) = scale parameter
\(\beta\) = symmetry
\(\alpha\) = tail behavior

Density of the NIG distribution \((\mu = 0)\)
AR models with NIG innovations

Let $X = (X_t)_{t \in \mathbb{Z}}$ a weakly stationary process. $X$ is an AR model of order $p$, if $X$ satisfies:

$$X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \varepsilon_t, \text{where } \varepsilon_t \sim NIG(\alpha, \beta, \mu, \delta) \text{ is a white noise process}$$

Model with $p + 4$ parameters: $\theta_p = (\varphi_1, \ldots, \varphi_p, \alpha, \beta, \mu, \delta)$

Estimation of $\theta_p$: maximization of the conditional likelihood

Supposing $p$ is fixed; conditional on the $p$ first observations $(X_1, \ldots, X_p)$, the log-likelihood of the observed path $(X_1, \ldots, X_T)$ is

$$l(\theta_p) = \sum_{t=p+1}^{T} \log(f(X_t - \sum_{i=1}^{p} \varphi_i X_{t-i}, \alpha, \beta, \mu, \delta))$$

How to determine $p$?

- $p$ reflects the surge correlation length
- $p$ selected to minimize the Bayesian Information Criterion (BIC), measuring the relative goodness of fit of a statistical model through a penalized likelihood criterion:

$$BIC(p) = -2l(\hat{\theta}_p) + (p + 4) \times \log(T - p)$$
Forecasting performances

- Forecasts evaluated over the last year of observations, which is excluded from the estimation of the models.

- Given the observations \((X_1, \ldots, X_T)\), the optimal one-step ahead linear predictor of \(X_{T+1}\) is defined as

\[
\hat{X}_{T+1} = \sum_{i=1}^{p} \hat{\phi}_i X_{T+1-i}
\]

- Accuracy evaluated with the Theil’s Inequality Coefficient (TIC):

\[
TIC(\hat{X}) = \sqrt{\frac{1}{n} \sum_{t=T+1}^{T+n} (X_t - \hat{X}_t)^2 \over \sqrt{\frac{1}{n} \sum_{t=T+1}^{T+n} X_t^2 \over \sqrt{\frac{1}{n} \sum_{t=T+1}^{T+n} \hat{X}_t^2}}}
\]

- TIC is ranging from 0 to 1, and forecasts are more accurate for a TIC closer to 0.
Data
Skew surge database

- Temporal series of skew storm surges for 35 sites located along the Spanish, French and UK coasts
  - Temporal resolution ≈ 12.4 hours
  - Mean duration = 34 years
  - Period = from 1846 (Brest) to 2011
Results
Seasonal components of skew surge series (1/2)

- No general trend
- Higher variability during the winter period than the summer period

**Two seasons**: Winter (October – March) & Summer (April – September)
The first two lags of the autocorrelation functions depend on the season.

Seasonal variability of the skew surges + correlation structure varies during the year → seasonal AR models.
Order and coefficients of AR models

- \( p \) range from 1 to 4 for the winter models and from 1 to 6 for the summer models: skew surges become uncorrelated after 48h in winter and 72h in summer.

\[
\begin{align*}
\phi_1 &= 0.66, \\
\phi_2 &= 0.14 \\
\phi_1 &= 0.63, \\
\phi_2 &= 0.27
\end{align*}
\]

BIC evolution for the summer Saint-Nazaire series

Correlation length (days) for the winter season

- The regional averages of the first two estimated AR parameters are \( \phi_1 = 0.66, \phi_2 = 0.14 \) for winter and \( \phi_1 = 0.63, \phi_2 = 0.27 \) for summer.
Parameters of NIG distributions

- Residual distributions are positively skewed and definitively **non-Gaussian**
- $\alpha$ generally low in winter (high variability) and in the North: winter surges follow a more heavy-tailed distribution

\[ \alpha \text{ values (winter season)} \]
Validation: comparison observations/simulations

- For each harbor: Monte Carlo series of the same length as the original samples are generated from the estimated models.

- Quantile comparison: 99% quantile, 99.9% quantile, maximum value.
Validation : comparison observations/simulations

Comparison of simulated and observed autocorrelation functions at Saint-Nazaire
Validation: NIG distribution for the residuals

- Graphical assessment and Kolmogorov-Smirnov test (only rejected for the winter model fitted on the Saint Jean de Luz series)
Validation: forecasting performances

- One-step ahead forecasts (12.4 hour horizon) evaluated over the last year of observations (excluded from the estimation of the models)
- On average, $TIC = 0.35$ in the winter and the summer
- AR models provide quite accurate forecasts
- High $TIC$ values in the North: partially explained by the heaviness of the tail of the NIG probability distribution of the residuals

TIC values for one-step ahead predictions (winter season)
Validation: forecasting performances

Optimal one-step ahead forecasts at Saint-Nazaire over 2010, with 70% confidence bounds from the NIG distribution of the residuals.
Extreme events from the AR models

- 1000 simulations of 1000 years each are performed for each site to estimate the empirical probability of the simulated extreme values
- Comparison between the GEV and AR approaches for the estimations of 1000-year return levels for the 35 series (with 70% confidence intervals)
Xynthia storm of February 2010 at Saint-Nazaire

- Highest surge observed at Saint-Nazaire (1.05 m)

- Let the morning of 28 February 2010 be the time $T+1$ when the highest skew surge was observed in Saint-Nazaire: $X_{T+1} = 1.05$ m.

- Suppose that $(X_{T-1}, X_T)$ are known (winter AR model of order 2), the optimal one-step ahead forecast gives $\hat{X}_{T+1} = 0.42$ m and $P(X_{T+1} > 1.05) = 6.10^{-4}$ from the NIG modeling.

- If a Gaussian distribution was fitted instead: $P(X_{T+1} > 1.05) = 5.10^{-9}$

- NIG distribution made an event like the Xynthia storm $10^5$ more probable (12.4 h before its occurrence) than using a Gaussian distribution.
Extreme surges at Saint-Nazaire

- **From simulations of seasonal AR models:**
  - 1000-year return level estimated at 1.54 m, (70% CI = [1.38 m, 1.68 m])
  - Xynthia surge (1.05 m) exceeded an average of 41.5 times during a 1000-year simulation
  - Estimation of the return period of Xynthia: 24 years (70% CI = [20 years, 29 years])

- **Classical GEV approach from the annual maxima:**
  - 1000-year return level estimated at 1.23 m (70% CI = [0.96 m, 1.51 m])
  - Return period of Xynthia estimated at 60 years
Conclusions

- Seasonal AR models **suitable** for skew surge series modeling

- NIG distribution is applied for the modeling of the residuals, to take into account the **non-symmetrical** and **heavy tailed** features of the surge phenomena

- Surge correlation lengths span up to **3 days**

- Accurate performances in **forecasting surges** at the 12.4 hour horizon:
  - Possibility to complete a classical meteorological surge forecast based on a dynamic model, with less input data

- Simulations of realistic artificial series are possible
  - Provides an alternative approach for the estimation of the probability of occurrence of extreme events

- Limitations: linear dependence between consecutive values
  - Non-linear dependence: cf. Hidden Markov models

- For more information:
  Weiss, J., P. Bernardara, M. Andreewsky and M. Benoit (2012) *Seasonal autoregressive modeling of a skew storm surge series*, Ocean Modelling, 47, 41-54
Thank you for your attention
Questions?

Storm on the Costa Brava, 6 March 2013