How Markov Chains models may contribute to the simulation and modelisation of Sea State Parameters processes

Valérie Monbet
Dept of Applied Statistics, University of South Britanny, Tohannic, F-56000 Vannes
&
Pierre-François Marteau
VALORIA, University of South Britanny, Tohannic, F-56000 Vannes

25th October 2000

ABSTRACT
The purpose of this paper is to explore how a Markov Chain Model of order higher than one may contribute to model and generate time histories of multivariate sea state processes including for instance significant wave height, peak period and wind speed.

KEY WORDS: Markov Model, Sea State, Non Gaussian processes, Multivariate simulation

INTRODUCTION

To study phenomena of erosion or transport of sediment in oceanic environement it is common to use simulations of sea state processes or sea waves as entry of erosion or transport of sediment models. Until recently the chosen entry was only monochromatic sea waves. But some researchers are now investigeting the influence of more realistic sea conditions on erosion and transport of sediments. Such studies will required observed or generated time histories of random sea states processes. For applications like erosion it is important that the duration statistics of the sea states are well restored.

Some methods have been developed to generate time series of uni and multivariate sea state processes like, for instance, the methods based on the simulation of a Gaussian process with given spectrum (Walton'90, Borgman'92, Monbet'00). The main steps of these methods are first to transform an observed process into a "Gaussian" one, then to generate a new Gaussian process with the same statistical properties and finally to apply the inverse transformation to obtain a new time serie of the sea state process. These methods are shown to give good results for bivariate simulation of significant wave height and peak period. Some parameters like the mean direction of propagation of the sea state are more difficult to restored because or their nature (Ailliot'00).
One of the limit of such methods is that they are based on the assumption that the original

process may be transformed into a Gaussian process. And it is easy to show that the process obtained after transformation are generally not Gaussian.

In this paper we propose to explore an alternative method where the multivariate sea state process will be modeled by a Markov model. Some papers have been writen in the field of Markov-based simulations of sea state parmeters such as significant wave height. But most authors consider univariate problems and fit Markov chains of order one.

In the familly of Markov models, there is a lot of possible approaches, for instance:
- Stochastic Differential Equations suitable for continuous time problems,
- Markov chain models used for problems with discrete time and discrete stats spaces and,
- Hidden Markov Models generally applied when the studying system involves an hidden markovian dynamic.

In the first part of the paper, we describe different Markov-based in the context of sea state processes. In the second part some results are presented and analyzed. And in the last part we propose some extensions and modifications suitable to improve our results.

MARKOV MODELS

In this section we will present the theoritical context of three Markov-based Models. Markov models are classically used to describe short memory dynamical phenomena.

Markov Chains

Several studies have been drawn to model the significative wave height by Markov Chains of order one (Haring'75, Vik'81). And Eidsvik (Eidvik'80) has fitted Markov Chains of order three to simulate the wind observed on "Polarfront". Anastasiou'96 uses a markov chain model of order one to estimate duration probability for a given time serie.

Here, we restrict us to the introduction of Markov Chains of order one and the related works for significant wave heightwith discrete time. Indeed recorded time series of sea state parameters are usually highly discrete in time (one measure every 3 or 6 hour).

Let us denote X a Markov Process. X is caracterized by the probability distribution of X_0 and the Markov property:

$$P(X_{t+1} = x_{t+1}|x_0^t) = P(X_{t+1} = x_{t+1}|X_t = x_t) = p_t(x_t, x_{t+1}) \tag{1}$$

where x_t denotes the value of X at current time t and x_1^t denotes the trajectory of X between times 0 to t. The probability function $p_t(.,.)$ is called the transition probability.

Discrete Space - On a discrete space is caracterized by a set S of K classes called states, an initial distribution on S and the transition probability is a matrix P. The states are the different values that the chain can reach. And the transition matrix describes the transition probability between two states in one time step for a Markov chain of order one. The matrix P is such that $\sum_{i=1}^{K} p_{ij} = 1 \ \forall j \in \{1, \cdots, K\}$. In many studies this matrix is chosen to be tridiagonal in order to limite the number of coefficients to be estimated. Furthermore the non zero coefficients of P are completly determined if we impose
(1) the long term distribution Π (i.e. the stationary distribution in Markov chain theory)
(2) the mean duration above each state d.

$$(\Pi, d) \implies P\Pi = \Pi + \text{equations for } d \implies P$$

It leads to a non linear system of equations which can be solved recursively. The advantages of this model are that

- the long term distribution and the mean duration of storms are exactly satisfied in the statistical sense,
- seasonnal variation may be introduced through variations in Π and d,
- the model is simple to use and it is not necessary to have a very large amount of data to perform the estimations.

But there are drawbacks:

- it is difficult to choose the discretization time and space steps because only the transition to neighbour states are allowed,
- the distribution of the durations is not necessarily fitted,
- it is an univariate model and,
- a first order Markov Chain may be too simple to explain the dynamic of sea state parameters.

Bivariate models can be build in the same way. There are two technics. The first one consists in discretizing one of the variable in a few (2 or 3) large classes and to fit one Markov Chain for the other variable in each large class. In the second one we mix both variables to create a new set of states and the Markov Chain is fitted for this new set.

If the amount of data is large enough it is possible to use less spare matrix than tridiagonal ones. But if we want to increase the number of non zero coefficients of the transition matrix and the order of the Markov chain simultaneously, the size of the transition matrix increases with a polynomial rate. And the size of the data set needed to reach a good estimation of the coefficients grows exponentially. As example let us consider a bivariate process (X, Y) both discretized in N classes. If we fit a Markov chain of order one for X, the full transition matrix is of size $N \times N$. If the chain is of order d, the size of the transition matrix becomes $O(N^{d+1})$. Furthermore if the Markov chain is fitted for the bivariate process, the size of the transition matrix is $O(N^{2(d+1)})$. It shows clearly that classical Markov models quickly become intracktable for large d. In this context it may be more convenient to work with Hidden Markov Models.

Continuous Space - One of the main drawback of the Discrete-space Markov Chains for sea state parameters is the artificial splitting of the variable into N classes with quite small N although the parameters are continuous in nature.

Hidden Markov Models

In Hidden Markov Model, we suppose that two groups of variables are of interest S and O. Each of these processes may be multivariate. And we assume that the dynamic of (S, O) is all given by S. The variable S may be hidden and is called "state variable". O is generally called "observed variable". The relations between S and O are formalized as follows:

$$P(o_t|s_1^t) = P(o_t|s_t) \tag{2}$$
$$P(s_{t+1}|s_1^t, o_1^t) = P(s_{t+1}|s_t) \tag{3}$$

where s_t (resp. o_t) denotes the value of S (resp. O) at current time t and s_1^t (resp. o_1^t) denotes the trajectory of S (resp. O) between times 1 to t. Hence if T is the final time, the joint probability of S and O is written

$$P(o_1^T, s_1^T) = P(s_1) \prod_{t=1}^{T-1} P(s_{t+1}|s_t) \prod_{t=1}^{T-1} P(o_t|s_t) \tag{4}$$

The joint distribution and furthermore the Hidden Markov Model is completely specified in terms of:

- initial state probabilies $P(s_1)$,
- transition probabilities $P(s_t|s_{t-1})$ and,
- emission probabilies $P(o_t, s_t)$.

In continuous case, probability density functions are used. And, when the state and the observation variables are discrete or discretized, initial state, transition and emission probabilities are defined by matrix respectively denoted (π, A, B).

In the genral context of Hidden Markov Models there are three basic problems:

1. Given observations $O = \{o_1^T\}$ and model $\lambda = (A, B, \pi)$ efficiently compute $P(O|\lambda)$.
2. Given observations $O = \{o_1^T\}$ and model $\lambda = (A, B, \pi)$, find the optimal state sequence $S = \{s_1^T\}$.
3. Given observations $O = \{o_1^T\}$ estimate model parameters $\lambda = (A, B, \pi)$ that maximize $P(O|\lambda)$.

Algorithms have been developed (ref) to solve these problems:

1. forward-backward algorithm allows to generate all the possible sequences of state associated with the given observations and model,
2. Vitterbi algorithm returns the state sequence which maximizes the log likelihood of the state sequence given the observations and the model.
3. Baum-Welch algorithm estimates the matrix A and B of the model, when the potential states of the model are known.

Algorithms such as Vitterbi may be convenient to predict miising valued in mesured time series of sea state parameters.

Let us now consider the problem of modelization of sea state parameters.

A MARKOV MODEL FOR SEA STATE PARAMETERS

Our problem is, given an observation, to estimate the Markov model and then to use it to generate new sequences of the time series. The obtained simulations are expected to restore statistical properties of the original sequence such as the distribution of the duration statistics and the intantaneous marginal and joint distributions of the sea state parameters.

Phase space reconstruction and Markov property - In practice, dynamical systems cannot be directly handled in their phase space, mainly because only partial knowledge about the dynamics is available to experimentalists. This partial knowledge is obtained through the observation of variables called observable. Nevertheless, under some strong conditions, among which stationarity and ergodicity of the underlying dynamical process are ensured, a phase space reconstruction method could be applied to embed the observable dynamics in a reconstructured phase space topologically equivalent to the original one. This method, originally suggested by Ruelle [] and then mathematically founded by Takens and Mañé [], is based on a time delay embedding techniques which gives $\Omega(k)$ as

$$\Omega(k) = [O(k), O(k-T), ..., O(k-T(d-1)]$$

where d is the dimension of the embedding space and $O(k)$ the observable samples at step k.

This construction means that the one step transition probability $P(\Omega(k)|\Omega(k-T))$ as seen in the d-dimensionnal embedding space is:

$$\begin{aligned} P(\Omega(k)|\Omega(k-T)) &= P(O(k), O(k-T), ..., O(k-T(d-1))|O(k-T), ..., O(k-dT)) \\ &= P(O(k)|O(k-T), ..., O(k-dT)) \end{aligned}$$

The probability $P(\Omega(N), \Omega(N-1), \cdots, \Omega(1))$ factorizes as a markov process of order d.

As a first attempt, we work with discretized versions of the sea state parameters in order to show that a Markov Model is convenient to approximate sea state processes. Each sea state parameter is discretized in K classes chosen to be almost of the same probability.

One of the question raised in this paper is the determination of the minimal dimension of the reconstructed phase space (or equivalently the choice of the order of the Markov model) in which the dynamics of a subset of discretized sea state parameters should be embedded. We propose to tackle this question through an assesment process consisting in comparing by product parameters such as duration statistics evaluated successively on a reference time serie and a synthetized one.

Choice of the model

Discrete Hidden Markov Model - Although, in the context of sea state parameter all the parameters are observed, we can treat the problem through the approach of Hidden Markov Models. For the modelization of sea state dynamic, it is convenient to choose the significant wave height H_s as state variable and the other parameters as observation variable. Indeed the significant wave height is, for most applications, the more determinant variable and it seems to often carry a great part of the dynamic of the sea state. For a Hidden Markov model of order d, three hypothesis are made. Let us note Γ the process of "non H_s" sea state parameters and H_s the process of significative wave height.

1. Γ and H_s are stationary and ergodic processes.
2. Γ_i is independant on Γ_j knowing $h_{s,1}^T$ for all $i, j \in \{1, \cdots, T\}$
3. For all $i \in \{1, \cdots, T\}$, Γ_i depends only on $H_{s,i}$

where $\Gamma^{[d]}$ denotes a vector of d successive values of the "non H_s" sea state parameters. For instance if we consider only H_s and T_p, $\Gamma_i = [T_{p,i}, T_{p,i+1}, \cdots, T_{p,i+d}]$. Similarly, $h_{s,1}^T$ represents a sequence of states including each d successive values of the parameter H_s.

Figure 1: Transition Probability of H_s

Figure 2: Autocorrelation functions for stationarized H_s and T_p

The complexity of the algorithms to be used to estimate the probability matrix A and B of the model involves that the good choice of the dimension parameter d and the number of classes K depends on the amount of available data.

To model sea state parameters, the greater K is the better the model fits because of the continuous nature of the parameters. We have verified on simulations that a dimension equal to 2 or 3 improve the model and that dimension 4 does not give really better results than 3. Furthermore, a simulation of (H_s, T_p) with 51 state and observation classes shows that the transition matrix of the states is highly sparse (1) and that an HMM does not suffice to restore the dynamics (dependance in time) of the observations T_p (2).

Continuous Markov Model - The last remarks drive us to the choice of a Continuous Markov Model without hidden parameter.

RESULTS

The main objectif of the test is to validate the Markov approach for multivariate sea state processes.

Description of the data - The model described above has been tested on real data. The first set contains in wind speed, significant wave height and spectral peak period for 10 years from 1986 to 1995. The data, which represent Northern North Sea, are obtained by pooling data from a number of nearby sites and using hindcast data when no measurements are available. The measured significant wave height H_s and the peak period T_p are computed from 20 minutes time series of sea surface elevation. Between each 20 minutes of time series, we have 3 hours gap. The historiticals of H_s, T_p ans Ws obtained from these measurements provide a time serie of about 26000 points.

Fitting the Markov Model - We have modeled the trivariate processe (H_s, T_p, W_s) by a seasonal markov chain of order 2 with continuous space. The data are not stationarized to perform the fitting of the model. And the transition probability p between 2 states are approximate by non parametrical estimators $\hat{p}$. The estimation is based on multivariate kernel density estimators with a Gaussian kernel K. For a vector of multivariate data $x_{est} = \{x_1, \cdots, x_n\}$

$$\hat{p}(x) = \frac{1}{n\sigma} \sum_{i=1}^{n} K(\frac{x - x_i}{\sigma})$$

The parameter σ is the width of the estimation window and allow to regulate the number of neighbour points included in the estimation and the density estimator may be very sensitive to the choice of σ value. To estimate the transition probability, we have to take into account the non stationarity of the data. Therefore, the vector x_est of the points using in the estimation of $p(x)$ is composed of neighbours of x which belong to an area around x where the stationarity can be assumed. For instance, let us denote $X(i,j) = (H_s(i,j), T_p(i,j), W_s(i,j)$ where i denote the year of the measurement and j the date in the year. Then the estimation of $p(X(i,j), X(i,j+1)$ is performs from $x_est = \{x(i,j-l), \cdots, x(i,j+l); i = 1, \cdots, nyear\}$ with l of order of one mounth.

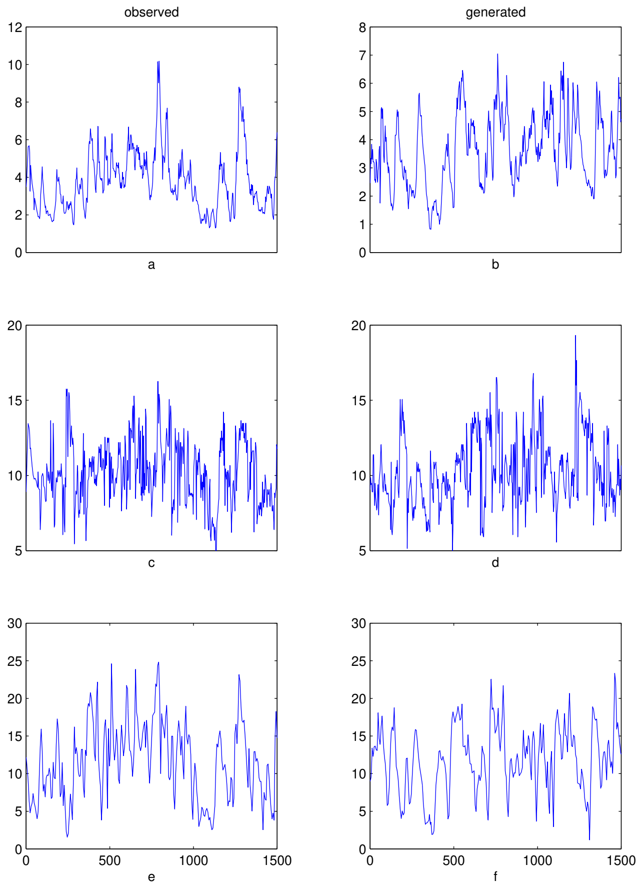

Figure 3: Trajectories H_s, T_p and W_s

It means that a 9-variate density is estimated with about 4800 points. The statistical properties of the estimators will be studied in an other paper.

Results - The chosen model is validate by comparing the statistical properties of observed sea state parameters and the generated ones. A sea state time serie is well characterized by the marginal distribution of its parameters and by the duration probability. notably, the time duration statistics show how the long term dependance are restored in the generated series.

CONCLUSION AND PERSPECTIVES

The main objective of this paper was to show that a markov model is convenient for the simulation of multivariate sea state parameters. We have justified the choice of a Markov Process Model with continuous space and discrete time amoung several other possibility like for instance Hidden Markov Models. The chosen non parametrical Markov model has been fitted for trivariate sea state parmeters including significant wave height, peak period and wind speed. It has been shown through simulations that thi smodel well restores the statistical properties of the parameters such as instantaneous distributions and time duration probability.

There will be several extensions of the study presented here. They will first concern to statistical properties of the estimators of the transition probability in order to check the size of

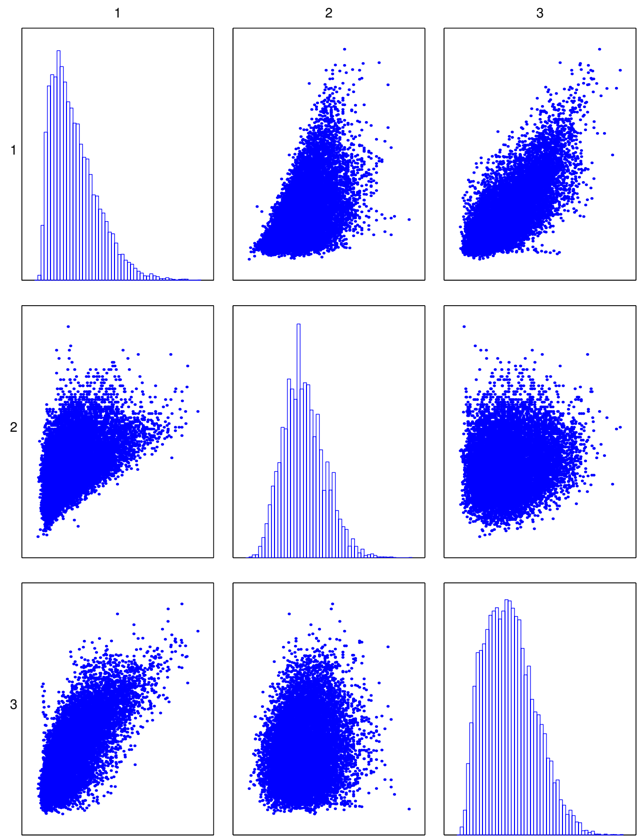

Figure 4: Instantaneous distributions of observed H_s, T_p and W_s

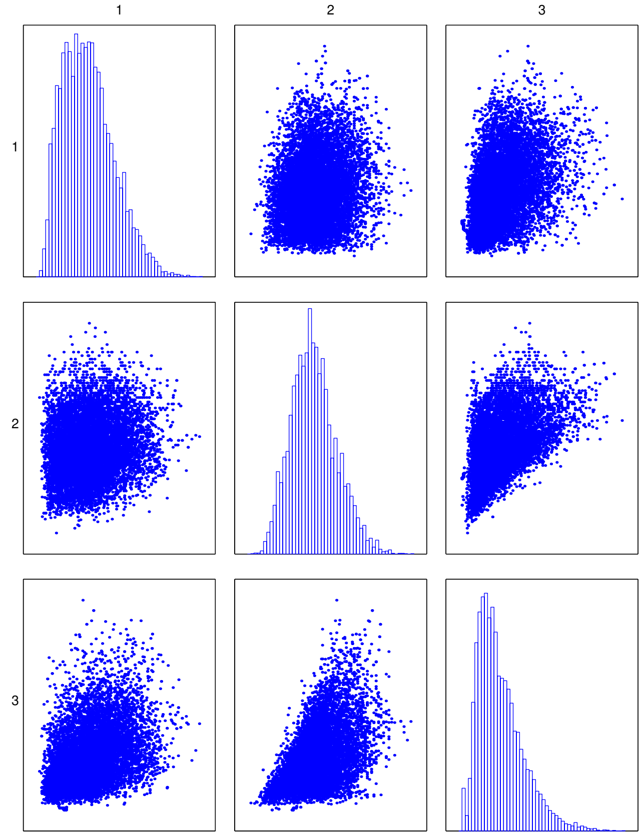

Figure 5: Instantaneous distributions of generated H_s, T_p and W_s

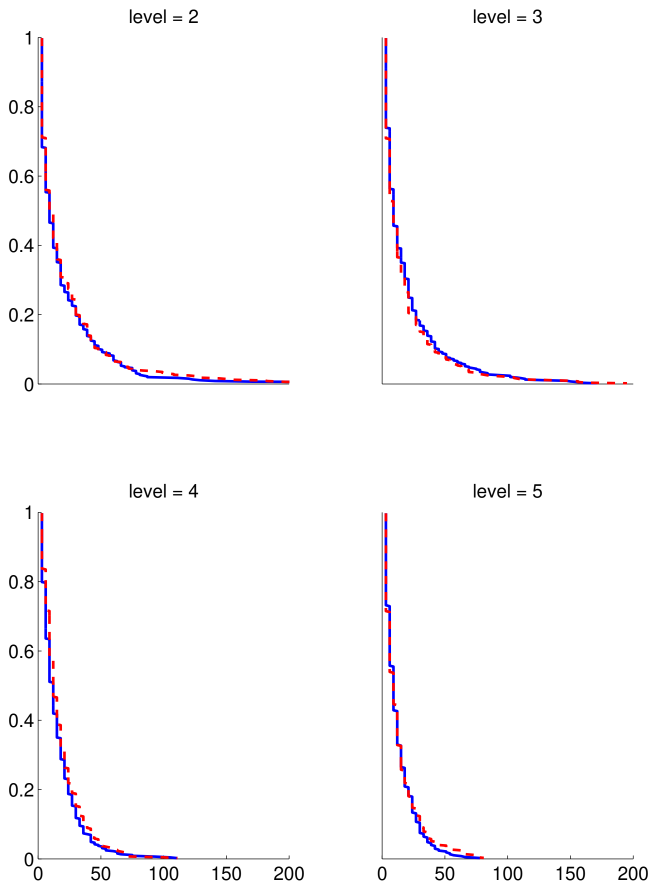

Figure 6: Time duration probabilities for H_s

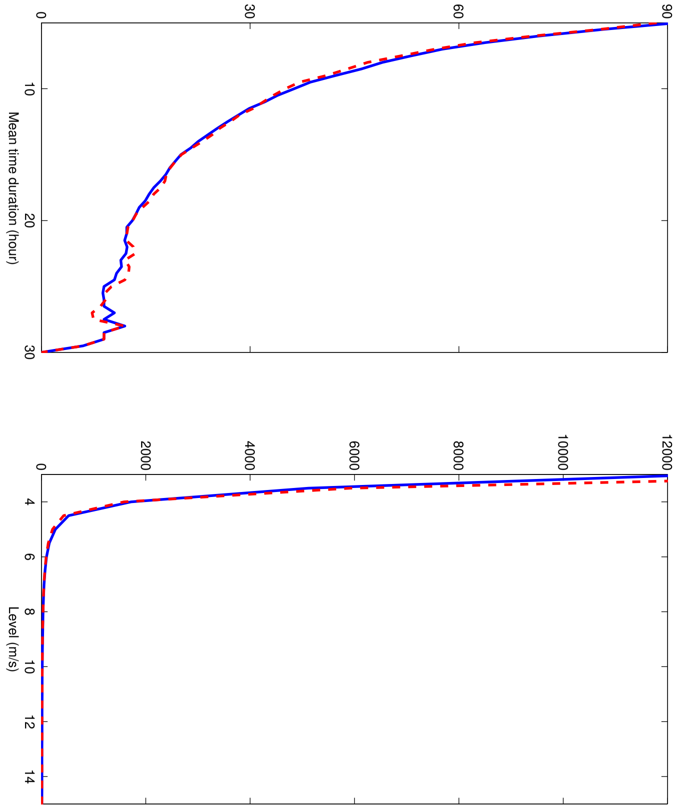

Figure 7: Mean time duration for W_s (left) and T_p (rigth)

the observed time serie which is to fit a multivaraite Markov model of order greater or equal to 2. And, in parallel, we will study the possibility of using parametric models for the transition probability. It would permit to obtain better estimation and, eventually, to perform simulations of sea state time series for area were no of few observations are available. Then it is important to extend the model to include parameters such as the propagation direction of the sea state. Models based on Hidden Markov Theory may be convenient for this problem.

REFERENCES

Ailliot P., (2000), Technical Report, DEA Université de Rennes 1, Ifremer.
Athanassoulis G.A., Stefanakos C.N., A Non-Stationnary Stochastic Model for Long-Term Time Series of Significant Wave Height, Preprint.

Borgman L.E., Scheffner N.W., (1991). Simulation of time sequences of wave height, period and direction, Technical report.

Eidsvik K.IJ (1980) Identification of Models for Some Time Series of Atmospheric Origin with Akaike's Information Criterion. *J. of Appl. Met.*; 19:357-369.

Kloeden P.E., Platen E., (1992).*Numerical Solution of Stochastic Differential Equations*; Springer-Verlag.
Langley R.S., (1987). A statistical analysis of non-linear random waves. *Ocean Eng.*; 10(5).

Monbet V., Prevosto M., (2000). Bivariate Simulation of Non-Stationary and Non-Gaussian Observed Processes: Application to Sea State Parameters. *Proc. 10th ISOPE Conf.*

Scheffner N.W., Borgman L.E., (1992). Stochastic time-series representation of wave data. *Journal of Waterway, Port, Coastal, and Ocean Engineering*; 118(4):337-351.

Vik I., (1981). The Tile Series Generating Module - Modelling aspects, *Ocean Research - Operationel Criteria, CNRD 3-2 Task 2*

Walton T.L., Borgman L.E., (1990). Simulation of non stationary, non gaussian water levels on great lakes. *Journal of Waterway, Port, Coastal and ocean engineering*; 116(6):664-685.

References for Markov Models and HMM.