MDPI

Article

Simulation of Wave Time Series with a Vector Autoregressive Method

Antonios Valsamidis [1], Yuzhi Cai [2] and Dominic E. Reeve [1,*]

1   Coastal Engineering Research Group, College of Engineering, Bay Campus, Swansea University,
     Swansea SA1 8EN, UK; antonios.valsamidis@swansea.ac.uk
2   School of Management, Bay Campus, Swansea University, Swansea SA1 8EN, UK; y.cai@swansea.ac.uk
*   Correspondence: d.e.reeve@swansea.ac.uk; Tel.: +44-1792-606-566

Abstract: Joint time series of wave height, period and direction are essential input data to computational models which are used to simulate diachronic beach evolution in coastal engineering. However, it is often impractical to collect a large amount of the required input data due to the expense. Based on the nearshore wave records offshore of Littlehampton in Southeast England over the period from 1 September 2003 to 30 June 2016, this paper presents a statistical method to obtain simulated joint time series of wave height, period and direction covering an extended time span of a decade or more. The method is based on a vector auto-regressive moving average algorithm. The simulated times series shows a satisfactory degree of stochastic agreement between original and simulated time series, including average value, marginal distribution, autocorrelation and cross-correlation structure, which are important for Monte Carlo modelling of shoreline evolution, thereby allowing ensemble prediction of shoreline response to a variable wave climate.

Keywords: VAR model; wave time series; autocorrelation; cross-correlation

1. Introduction

Simulation of time series plays an important role in many areas due to the high cost in obtaining in situ measurements. In the field of coastal engineering simulation of wave time series has been used for estimating the duration of storm events and their spacing in time, see e.g., [1,2], with the aim of assessing the risk of serious beach erosion [3,4]. Typically, a relatively short record of measured wave conditions is available, hence to perform Monte Carlo simulation of coastal flooding or erosion, wave sequences with similar statistics are required. This problem is particularly challenging and may be stated as simulating a non-stationary non-normal, correlated, trivariate stochastic process, given a sample or realisation of the process. Several methods have been proposed for tackling this type of problem. For example, Li and Winker [5] proposed a Monte Carlo method, or a quasi Monte Carlo method, to obtain simulated time series from a vector autoregressive moving average (VARMA) model fitted to sample data. Barone [6] described a simulation method that can be used to generate realisations of a VARMA process, while Shea [7] discussed a direct method of computing the initial state covariance matrix required by the simulation method. However, these methods cannot ensure the marginal distributions, and/or the autocorrelation patterns of the simulated data are the same as those of the sample. On the other hand, representative wave events obtained via a K-Means algorithm clustering technique has been used for studying port operability and long-term longshore sediment transport, but this suffers from not taking extreme wave events into account [8].

Methods to simplify the problem include a technique for simulating time series of wave height, period and direction, accounting for seasonality on a monthly mean basis [2]. This can suffer from jumps in the mean of variables at the changeover between consecutive months. Focusing on wave height and period only, bivariate autoregressive models were proposed by Soares and Cunha [9]. Recently, Cai [10] and Cai et al. [11] generalised the work

of [12,13] in order to obtain simulated multivariate time series. However, these methods are suitable for stationary time series only, and hence, they cannot be used to generate the non-stationary wave conditions observed widely in practice. Although Cai et al. [14] developed a simulation method for non-stationary time series, this method is suitable only for short non-stationary time series due to relatively higher computational cost, and therefore is not suitable for wave time series corresponding to long-term measurements.

The purpose of this paper is to extend the method of [10] to recreate both the statistics and lag correlation properties of a non-stationary, multivariate wave record spanning many years. The data used here are multivariate, correlated and non-stationary. The reader is referred to [15–17]. Furthermore, this study relaxes the constraints of describing seasonality of all wave variables with a single functional form. More specifically, the seasonality of wave height and period time series were not assumed to have the same form as the seasonality of wave direction.

This paper is organised as follows. In Section 2, the theoretical background of the model is presented; in Section 3 the measurements from the study site and the approaches we used for removing the trends and seasonality are discussed. The simulation results are demonstrated in Section 4. Finally, discussion and conclusions are presented in Section 5.

2. Methodology

2.1. Outline

Let $E_t = (E_{1t}, E_{2t}, E_{3t})$ denote observed sea condition data, where E_{1t} is the wave height, E_{2t} is wave period and E_{3t} is the wave direction, and the subscript t indicates waves at time t. As discussed in the previous section, the main objective of this research is to obtain simulated sea condition data, denoted by $\widetilde{E}_t = \left(\widetilde{E}_{1t}, \widetilde{E}_{2t}, \widetilde{E}_{3t} \right)$, such that the simulated data will have similar trend, seasonality, autocorrelation structures and marginal distributions with those of the observed E_t.

It is worth noting that the method of [10] requires that the observed data are stationary. Hence, it is important to convert the observed wave data into stationary ones. More specifically, our method includes the following steps: (i) Remove the trend and seasonality from the observed data to produce stationary series; (ii) apply the method of [10] to the series obtained from step (i); (iii) include the trend and seasonality into the simulated data obtained from step (ii) to generate the final simulated data that we require. Figure 1 illustrates our proposed methodology:
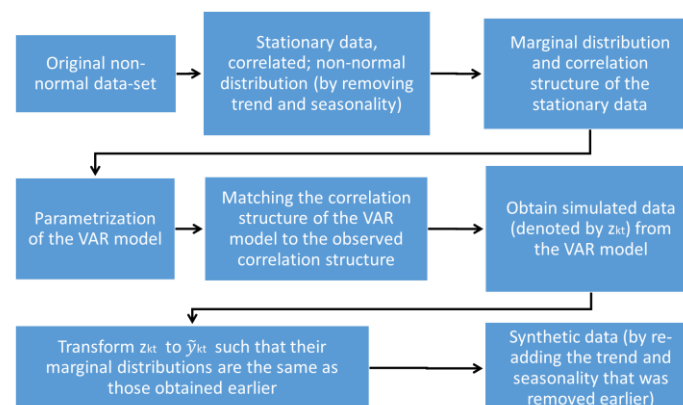

Figure 1. Schematic representation of the wave data generation methodology.

First note that in this application, we consider the trend and seasonality to be additive. That is, each E_{kt} for k = 1, 2, 3 can be expressed by $E_{kt} = T_{kt} + S_{kt} + y_{kt}$, where T_{kt} and S_{kt} represent the trend and seasonality components, respectively, and y_{kt} is the detrended and deseasonalised data that will be used in step (ii) of our method.

For wave directions, the sample was first transformed to create a variable with similar magnitude and range as that of wave height and period, (see Section 3), before the trend

was estimated and removed as per the treatment of wave height and period. Then, a linear trend for each variable and an autoregressive process model of order 5 for wave height and period were estimated, which were subtracted from the original data to give the detrended data $E_{kt} - T_{kt}$.

For seasonality S_{kt}, as waves are driven by atmospheric winds, their characteristics will reflect the seasonal changes in the weather. However, at any particular location, the wave conditions may be a combination of waves from several sources so that 'seasonality' in the literal sense may not be reflected in wave height, period and angle in an identical manner. The seasonality parameters were determined using least squares estimation. Specifically, to estimate the seasonal component of each one of the three wave time series, sums comprising of gradually increasing numbers of sine or Fourier terms were successively tested via the Augmented Dickey-Fuller test [18], until the suitable combination of seasonal components were obtained. Then, the selected seasonal components were removed from the wave time series. To this end, an existing routine in MATLAB [19] was applied for testing the stationarity of the wave time series via the Augmented Dickey-Fuller test. This routine gives as output the value 0 for non-stationary data and 1 for stationary data. Following this method, the seasonality of wave height and period time series were described via a sum of sine terms while the seasonality of wave directions was described via a sum of Fourier terms. After removal of the trend and seasonal component the detrended and deseasonalised series are given by $y_{kt} = E_{kt} - T_{kt} - S_{kt}$, which is a vector sequence, (trivariate—with wave height, period, and direction), that is stationary, correlated and non-normal.

Once the detrended and deseasonalised series y_{kt} for $k = 1,2,3$ have been computed, and successfully tested for stationarity, a vector VAR (vector auto-regressive) method, (outlined in the next section), may be applied to obtain simulated data for y_{kt}. This may then be transformed back to obtain simulated data for E_{kt} with the correct seasonality and trend.

2.2. Detailed Methodology

Following [10], let the base process of the method be a vector AR (auto-regressive) process of order p, denoted by VAR(p), defined by

$$z_t = \varphi_1 z_{t-1} + \varphi_2 z_{t-2} + \cdots + \varphi_p z_{t-p} + u_t \tag{1}$$

where $z_t = (z_{1t}, z_{2t}, z_{3t})'$, $z_{kt} \sim N(0,1)$ for $k = 1, 2, 3$, ϕ_i are fixed 3×3 coefficient matrices, $I = 1, \ldots, p$, and $u_t = (u_{1t}, u_{2t}, u_{3t})'$ is a 3-dimensional normal random variable with mean 0 and covariance matrix Σ_u such that

$$E(u_t u'_{t-h}) = \begin{cases} \Sigma_u & \text{if } h = 0 \\ 0_{3x3} & \text{otherwise} \end{cases}$$

where

$$\Sigma_u = \begin{pmatrix} s_{11} & s_{12} & s_{13} \\ s_{21} & s_{22} & s_{23} \\ s_{31} & s_{32} & s_{33} \end{pmatrix} \tag{2}$$

and

$$s_{ij} = E(u_{it} u_{jt}) \text{ for } i, j = 1, 2, 3$$

with

$$\varphi_i = \begin{pmatrix} \varphi_{i11} & \varphi_{i12} & \varphi_{i13} \\ \varphi_{i21} & \varphi_{i22} & \varphi_{i23} \\ \varphi_{i31} & \varphi_{i32} & \varphi_{i33} \end{pmatrix}$$

It follows, see e.g., [20], that the correlation matrix function of z_t is given by $\Gamma(h) = E(z_t z'_{t-h})$, where $h = 0,1,2,..., H$ and

$$\Gamma(0)=\sum_{l=1}^{p}\phi_l\Gamma(0-l)+\Sigma_u \tag{3}$$

$$\Gamma(h)=\sum_{l=1}^{p}\phi_l\Gamma(h-l)$$

$$\Gamma(h)=(r_{ijh})_{3\times3}=\begin{pmatrix} r_{11h} & r_{12h} & r_{13h} \\ r_{21h} & r_{22h} & r_{23h} \\ r_{31h} & r_{32h} & r_{33h} \end{pmatrix}, \ \Gamma(0)=(r_{ij0})_{3\times3}=\begin{pmatrix} 1 & r_{120} & r_{130} \\ r_{120} & 1 & r_{230} \\ r_{130} & r_{230} & 1 \end{pmatrix}$$

where H is a fixed number that defines the maximum lag value we would like to consider for matching the autocorrelation structures between the simulated and the observed time series, r_{ijh} is the correlation between z_{it} and z_{jt+h} and $i,j = 1, 2, 3$, and $\Gamma(-h) = \Gamma'(h)$.

The simulation method requires the estimation of ϕ_i, r_{ijh} and Σ_u. The correlations r_{ijh} can be obtained by solving the following non-linear equations

$$\rho_{ijh} = \frac{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}F_i^{-1}(\Phi(z_{it}))F_j^{-1}\left(\Phi\left(z_{jt+h}\right)\right)\xi_{r_{ijh}}\left(z_{it},z_{jt+h}\right)dz_{it}dz_{jt+h}-E(y_{it})E\left(y_{jt+h}\right)}{\sqrt{var(y_{it})var\left(y_{jt+h}\right)}} \tag{4}$$

where ρ_{ijh} is the correlation between two stationary time series corresponding to the original input data, $F_j(\cdot)$ is the marginal distribution of y_j and hence $F_j^{-1}(\cdot)$ represents its inverse function; $\Phi(\cdot)$ is the distribution function of the standard normal distribution; $\xi_{rijh}(\cdot,\cdot)$ is the joint density function of two normal variables with mean zero and correlation r_{ijh}; $E(y_{it})$ is the mean of y_{it}, and $var(y_{it})$ is the variance of y_{it}. An explanation about how Equation (4) is derived is presented in the Appendix A.

Thus, given the sample data E_t, we determined detrended and deseasonalised data y_{it}. Setting $F_j(\cdot)$ equal to the empirical distribution of y_{jt} and setting $E(y_{it})$ and $var(y_{it})$ equal to the sample mean and variance of y_{it} respectively, and replacing ρ_{ijh} with the sample autocorrelation of y_{it} and y_{jt+h}, the r_{ijh} can be obtained by solving the resulting non-linear equations using the methods detailed in the Appendix B. Once the r_{ijh}'s are available, ϕ_i and Σ_u for $i = 1,...,p$ can be obtained by solving Equation (3).

Hence, with r_{ijh}, ϕ_i and Σ_u determined the simulated data can be produced as follows: (i) Obtain simulated data from the base process (Equation (1)); (ii) use the simulated $z_t = (z_{1t}, z_{2t}, z_{3t})$ to obtain simulated y_{it} for $i = 1,2,3$. To this end, the corresponding values of the normal cumulative distribution function, $\Phi(z_t)$, are calculated given as input data the simulated values z_t. Then, the $\Phi(z_t)$ values are interpolated in the empirical marginal distribution set of values with respect to the wave parameters (height (for $i = 1$), period (for $i = 2$) and direction (for $i = 3$) to yield the simulated data y_{it}. (iii) Finally, the trend and seasonality components that had been removed before (see Figure 1) are re-added to y_{it} to yield the simulated time series E_{it} for $i = 1,2,3$ as required. The whole process is schematised in Figure 2.
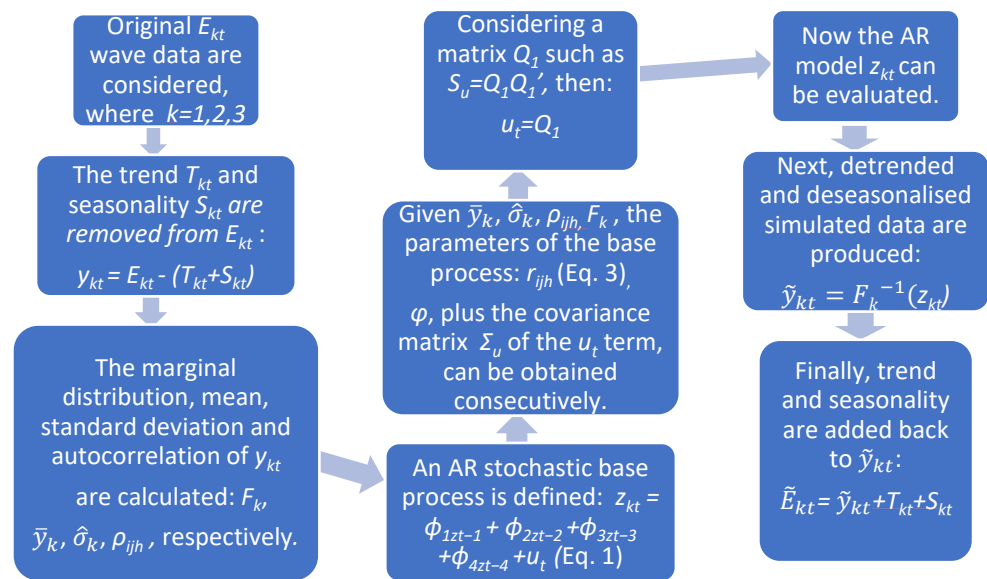
Figure 2. Schematised flow of data processing.

3. Case-Study and Data Processing

For our test case we have chosen Littlehampton which is located in Southeast England (Figure 3a).
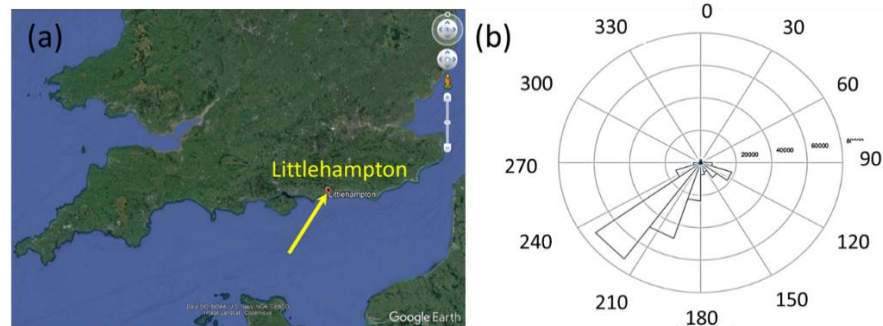

Figure 3. (**a**) Littlehampton beach lies in Southeast England; (**b**) a rose diagram illustrating the predominant wave direction from southwest to northeast,where the numbers indicate degrees from North.

Wave measurements were accessed from the Channel Coastal Observatory, a UK organisation which collects and archives coastal field-data. Specifically, time series of significant wave height (Hs), peak wave period (T_p) and wave direction (α) relative to North, between the 1st of July 2003 and the 30th of June 2016 were gathered. Integrated wave parameters were available at an interval of 30 min at a location 4 miles SSE of Littlehampton harbour entrance; observations were made with a Datawell Directional WaveRider Mk III buoy moored in approximately 10 m water. The wave rose of the 13 year record is illustrated in Figure 3b. This shows a wave climate with a predominant southwesterly approach and a secondary peak in waves from the southeast. Waves from the southwest are typically a mixture of locally generated wind waves and swell waves from Atlantic storms. Southeasterly waves are fetch-limited storm waves generated by the northern part of low-pressure systems that track to the south of the UK [21].

The wave records were averaged to create a time series of daily wave conditions, as our focus is on storm events rather than wave by wave fluctuations. The daily data are shown in Figure 4, in which seasonality is visually evident.
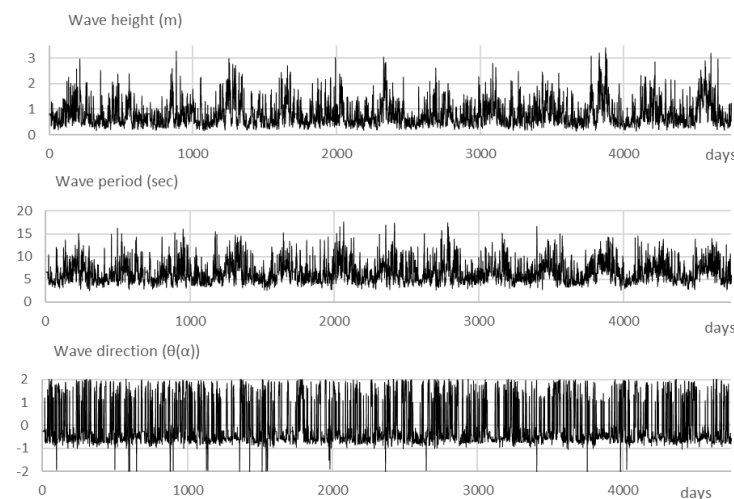
Figure 4. Time series of the observed daily wave data over the period 1 July 2003 to 30 June 2016. The horizontal axis shows days from the start of the period. The top panel shows wave heights, the middle panel shows wave periods and the bottom panel the transformed wave direction.

If the magnitudes of the variables of interest are very different, it is normal practice to standardise or transform them before beginning the modelling process in order to improve the fitting. Here the range of wave direction is much larger than the other two variables. Hence, the following transformation of wave direction in the data preparation stage was performed:

$$
\begin{aligned}
\text{if } 0 \le \alpha < \tfrac{\pi}{2}, \quad & \theta = \sqrt{1 - \cos^2(\alpha)}, \quad -0 \le \theta < 1 \\
\text{if } \tfrac{\pi}{2} \le \alpha < \pi, \quad & \theta = 1 + \sqrt{1 - \cos^2(\alpha)}, \quad -1 \le \theta < 2 \\
\text{if } \pi \le \alpha < \tfrac{3\pi}{2}, \quad & \theta = -\sqrt{1 - \cos^2(\alpha)}, \quad -1 \le \theta < 0 \\
\text{if } \tfrac{3\pi}{2} \le \alpha < 2\pi, \quad & \theta = -1 - \sqrt{1 - \cos^2(\alpha)}, \quad -2 \le \theta < -1
\end{aligned}
\tag{5}
$$

Before undergoing any further processing of the wave time series, it was important to split the available dataset into a training and a validation subset. The training subset constituted the first 75% of the whole dataset, thus, covering the time period from 1 July 2003 to 31 March 2013. In this time span, the aim was the VAR model to be properly parametrised, specifically, to estimate the trend and the seasonal elements of the wave time series, plus the φ_i and u_t parameters of Equation (1). The validation is performed on the remaining 25% of the whole dataset, namely, from 1 April 2013 to 30 June 2016.

Next, the trend and seasonal elements of the wave timeseries are assessed, and subsequently, removed to achieve stationarity.

3.1. Detrending

To estimate T_{kt}, the following model was used to represent the trend; coefficients being obtained by fitting this to the observations using least-squares estimation:

$$
x_t = a + bt + \sum_{u=1}^{p} \alpha_k x_{t-u} + \varepsilon_t
\tag{6}
$$

where ε_t are independent, identically distributed random variables, $p = 0$ for wave direction and $p = 5$ for wave height and wave period. The value of the order p was chosen to ensure the stationarity of the detrended and deseasonalised series. The estimated linear trend for each variable is shown in Figure 5, and the estimated parameter α_k values for wave height are $\alpha_1 = 0.5996$; $\alpha_2 = -0.08316$; $\alpha_3 = 0.11206$; $\alpha_4 = -0.02666$; and $\alpha_5 = 0.0373$, while the corresponding values for wave period are: $a_1 = 0.5007$; $a_2 = -0.04946$; $a_3 = 0.03786$; $a_4 = 0.0182$; and $a_5 = -0.019$. The trend, T_{kt}, was taken as the residual series ε_t of the models respectively.
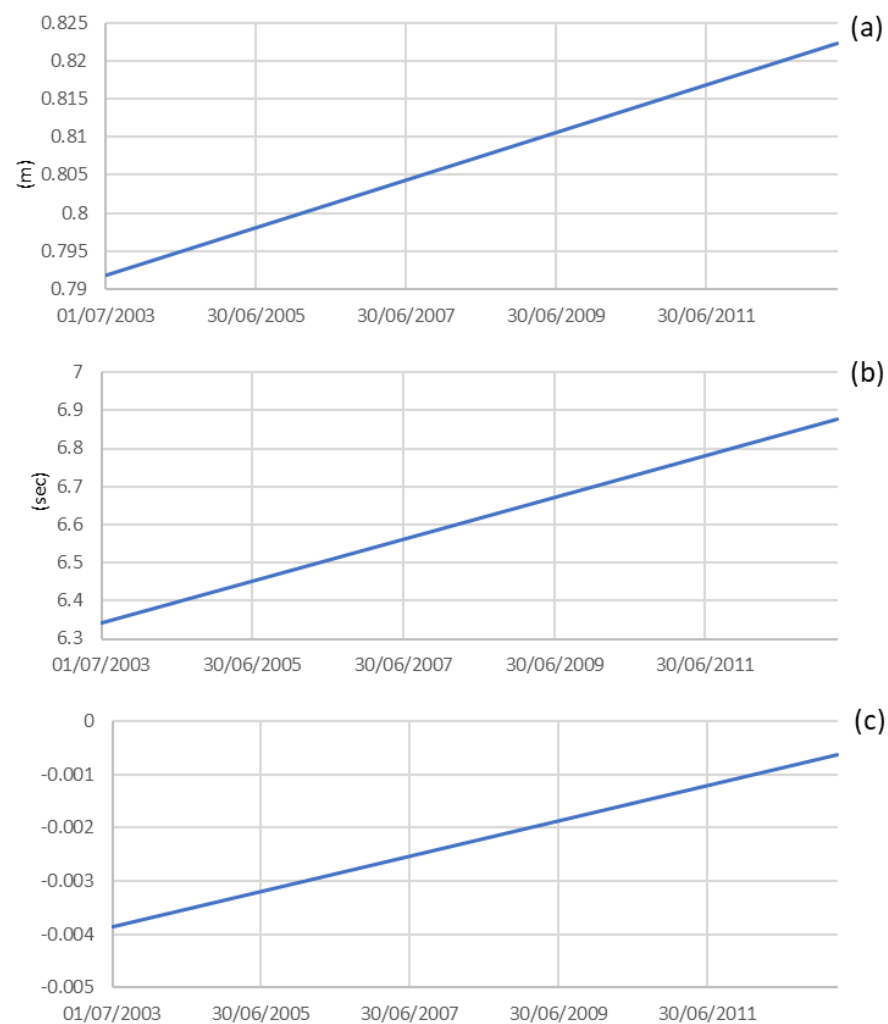
Figure 5. For 1 July 2003 to 31 March 2013, the upper panel (**a**) shows the significant wave height trend (H_s), similarly, the middle panel (**b**) depicts the peak wave period trend, and finally, the lower panel, (**c**), shows the wave direction trend. Dates are shown in day, month, year format.

3.2. Seasonality

For seasonality, a best fit curve was chosen to describe the seasonal element for each variable. This curve is given by the following equation corresponding to a sum of sine terms: $S_t = a_1 \times sin(b_1 \times n + c_1) + \ldots + a_8 \times sin(b_8 \times n + c_8)$, for significant wave height H_s and peak wave period T_p, while for the wave direction α, the seasonal element was described via an equation comprising of a sum of Fourier terms: $S_t = a_0 + a_1 \times cos(nw) + b_1 \times sin(nw) + a_2 \times cos(8nw) + b_2 \times sin(8nw) + \ldots a_8 \times cos(8nw) + b_8 \times sin(8nw)$, where n is the number of the consecutive temporal step. The fitting parameters a_i, b_i and c_i for wave height and period time series, and the corresponding values a_i, b_i and w for wave direction were estimated via the least-squares method for each of wave height, period and direction. A different formulation for the seasonal components for wave direction was required in order to remove all the non-stationarity in the series. Figure 3 suggests that seasonal components in the wave direction are different from those in the other two wave variables and this is borne out by the nature of the non-stationarity of the respective variables.

The extracted seasonal components are presented in Figure 6. Note that monthly averaged values of wave parameters have been used for visual clarity.

Finally, the augmented Dickey-Fuller test [18] was applied to the detrended and deseasonalised data y_{it} to determine whether the series was stationary. This process was performed iteratively, adding additional terms to the seasonality model, until the augmented Dickey-Fuller test indicated that y_{it} were stationary.
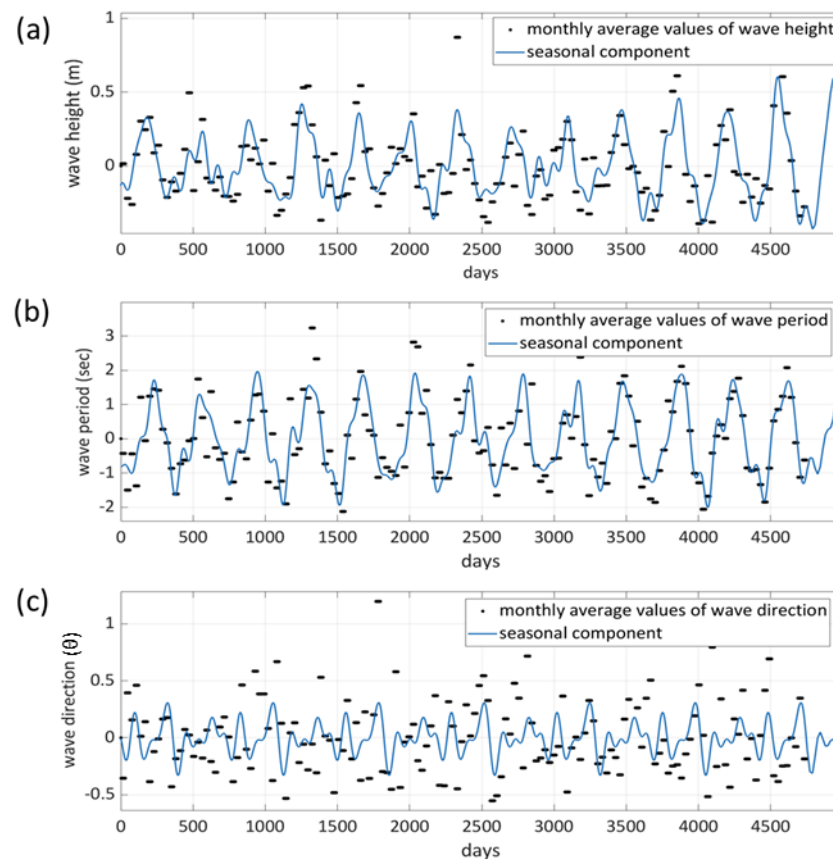
Figure 6. The upper panel (**a**) shows the seasonal component of significant wave height (H_s), similarly, the middle panel (**b**) depicts the seasonal component of peak wave period, and finally, the lower panel, (**c**), shows the seasonal component of wave direction $\theta(\alpha)$.

4. Simulation Results

As mentioned in the previous section, the trend and seasonal component of the first 75% of the available measurements at Littlehampton were taken into account. The marginal distributions of wave height, period and direction and their autocorrelations and cross-correlations were estimated from this sequence. The correlations were used to estimate the values of r_{ijh}, ϕ_i and Σ_u by solving Equations (2) and (3) (see also Appendix B), where $h = 0, 1, \ldots, 3$. Hence, in this study we let $H = 3$, corresponding to matching the correlation structure of the data up to three days, or approximately the storm duration at the site. Note that these parameter values define the correlation structure of the model (Equation (1)). Then, the estimated model (Equation (1)) was used to obtain simulated data z_t, i.e., the detrended and deseasonalised synthetic data (see Appendix B). The length of the simulated data was taken to be the same as the original series for illustration purpose. Finally, the trend and seasonality removed in the initial steps were added back to create the output synthetic time-series, $\widetilde{E}_t$ (Figure 7).

If the method is working well, the simulated data $\widetilde{E}_t$ should have similar statistical properties to those of the original data E_t. As a check on this, marginal distributions and correlations of the original and simulated series were compared. The marginal distributions of the observed and simulated data were estimated with the non-parametric kernel estimation method [22]. The estimated marginal density functions are given in Figure 8, where the blue curves correspond to the observed data and the red curves correspond to the simulated data. It can be seen that the two sets of density functions are very close for all three sea condition variables. Moreover, the estimated means of the observed and simulated data are shown by the blue and red vertical lines respectively in Figure 8, and show extremely close agreement.
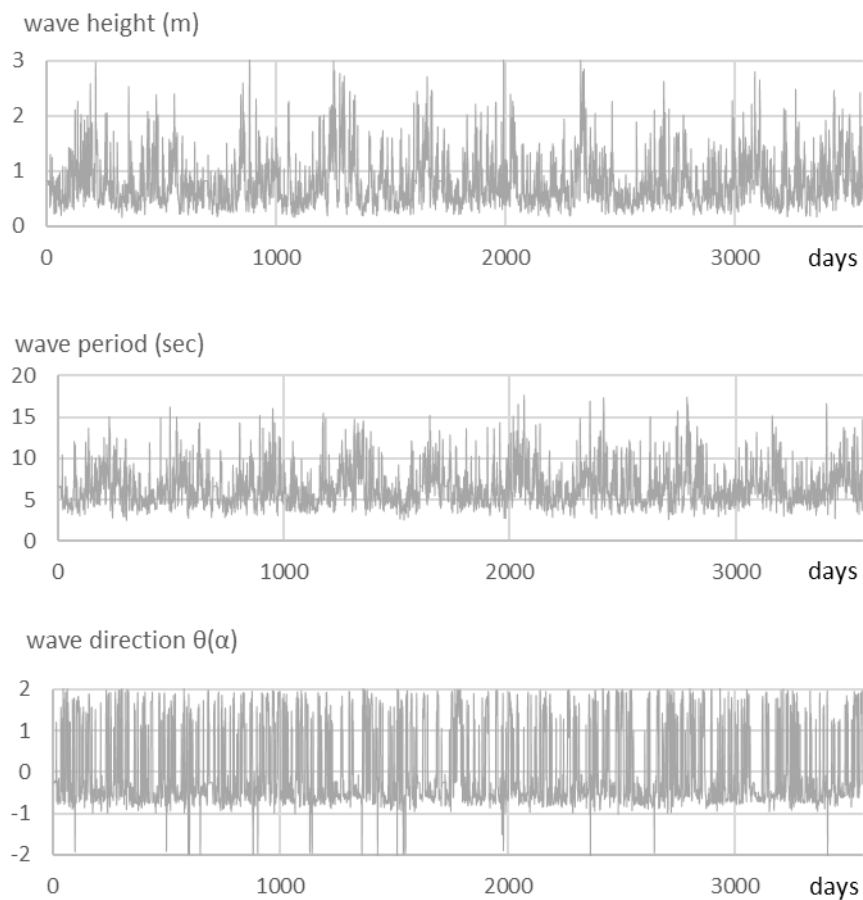
Figure 7. Time series of the simulated daily wave data in the time period: 1 July 2003 to 31 March 2013.
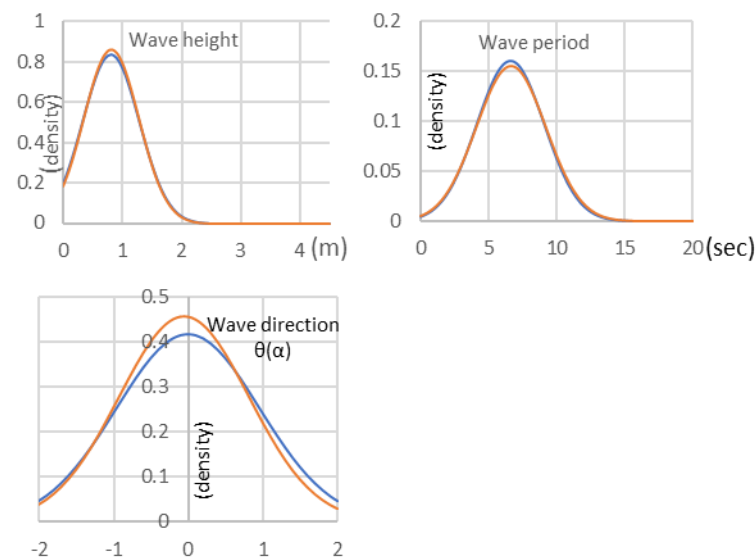

Figure 8. In the time period: 1 July 2003 to 31 March 2013, density function plots for the observed and simulated wave height (m), wave period (s) and wave direction (dimensionless transformed) respectively. Blue curves correspond to observed data and red curves correspond to simulated data. Similarly, blue and red vertical lines correspond to the means of the observed and simulated data, respectively.

Figure 9 shows the estimated autocorrelation and cross-correlation between variables for the original and simulated data series.
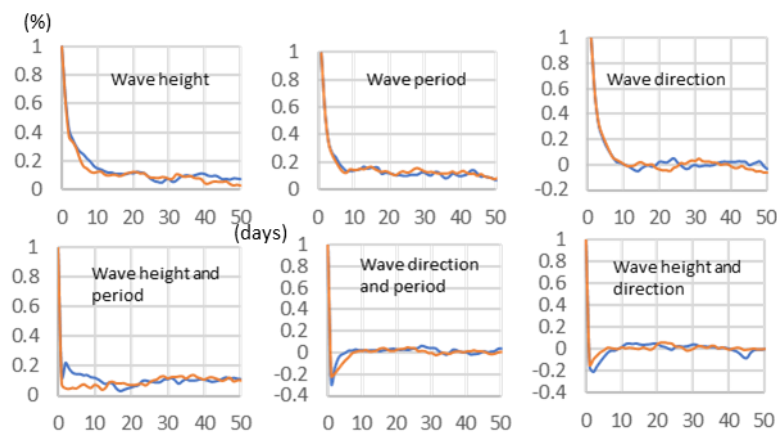
Figure 9. In the time period: 1 July 2003 to 31 March 2013, first row: autocorrelation up to 50 lags for each sea condition variable. Second row: cross-correlation between pairs of sea condition variables. Blue lines are the auto/cross correlation function of the original data, and the red lines correspond to those of the simulated data. Note: In all plots the horizontal axis is the lag time in days and the vertical axis is the normalised correlation value which lies between 1 and −1.

Overall, Figure 9 shows a very good agreement between the original and synthetic data. The final step is the validation of the VAR model on an independent section of measurements; that is, the time period from 1 April 2013 to 30 June 2016. The simulated wave time series, along with their density functions is presented in Figures 10 and 11 respectively.
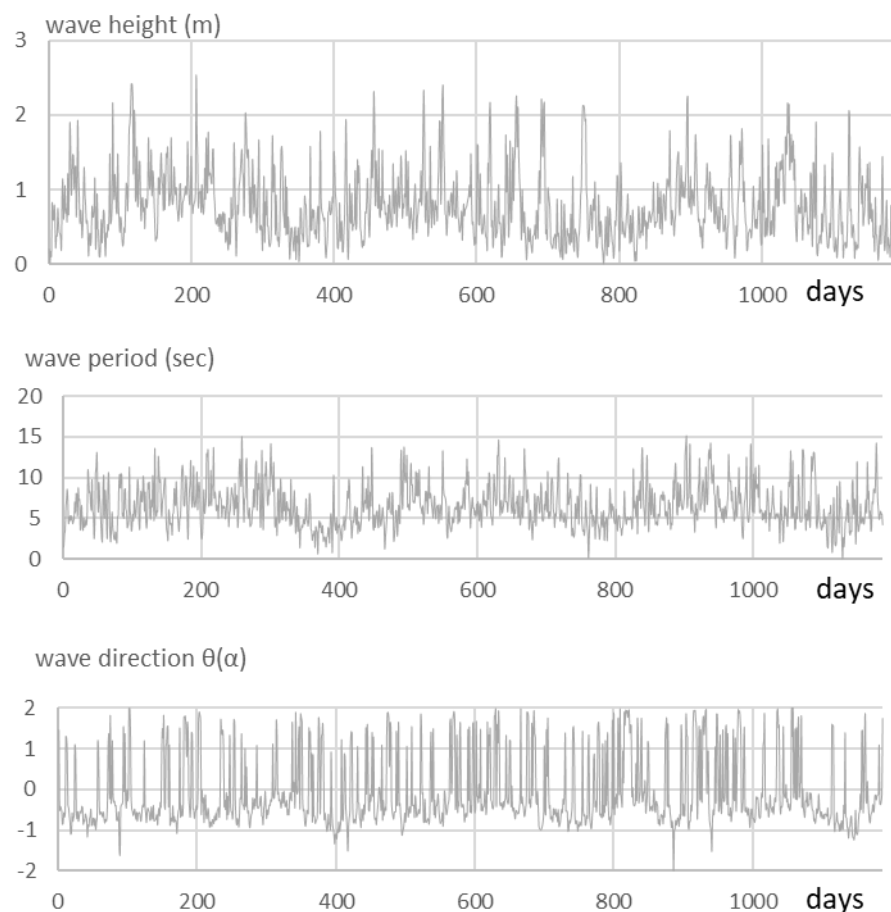

Figure 10. Time series of the simulated daily wave-data in the time period from: 1 April 2013 to 30 June 2016.
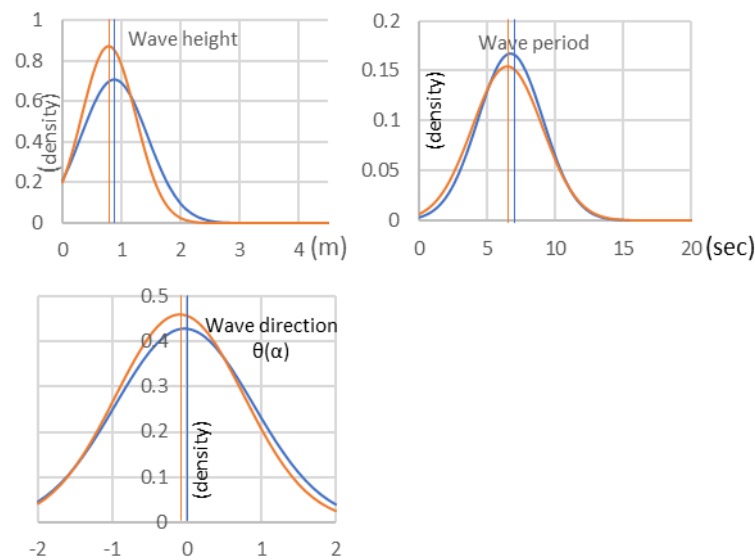
Figure 11. Density function plots for the observed and simulated wave height (m), wave period (s) and wave direction (dimensionless transformed) respectively, in the time period from: 1 April 2013 to 30 June 2016. Blue curves correspond to observed data and red curves correspond to simulated data. Similarly, blue and red vertical lines correspond to the means of the observed and simulated data, respectively.

Figure 11 illustrates good agreement between the means of wave height, period and direction. Some divergence between the original and simulated wave time series, particularly at the peaks of the density functions, is evident.

A comparison between detrended and deseasonalised original and synthetic time series was conducted. Results are shown in Figure 12 which demonstrate a good level of agreement in the correlation structure of the original and synthetic series. The temporal correlation scales in auto- and cross- correlations is captured well although the slight negative cross-correlation between wave height and period at lags up to one week is absent in the synthetic series. Example input data and processed data can be found in Supplementary Materials.
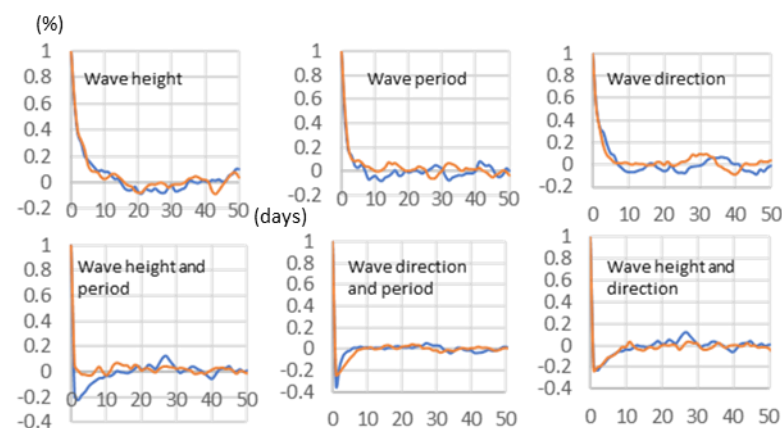

Figure 12. In the time period from: 1 April 2013 to 30 June 2016, first row: autocorrelation up to 50 lags for each sea condition variable. Second row: cross-correlation between pairs of sea condition variables. Blue lines are the auto/cross correlation function of the original data, and the red lines correspond to those of the simulated data. Note: In all plots the horizontal axis is the lag time in days and the vertical axis is the normalised correlation value which lies between 1 and −1.

5. Discussion and Conclusions

A methodology for simulating multi-variate wave sequences via a vector autoregressive (VAR) stochastic model was presented in this study and its application illustrated with measurements over a 13-year period taken at Littlehampton, UK. The measurement record was split into two non-overlapping subsets. The first one extending from 1 July 2003 to 31 March 2013 and the second one from 1 April 2013 to 30 June 2016, to provide independent training and validation data sets. The model was successfully calibrated and validated.

The utility of the correlation functions on non-stationary data is arguable and does not appear to rhyme with "A key element of the procedure is a detailed treatment of non-stationarity in the wave time sequences."

A key element of the procedure is a detailed treatment of non-stationarity in the wave time sequences. The non-stationarity may have different manifestations within each element of the wave conditions; with wave height, period and direction each exhibiting different non-stationarity. Our methodology allows for such variation and is able to create synthetic sequences of wave conditions that have very similar statistical properties to the original dataset.

We applied the method to a site in the UK that experiences mid-latitude storms that have a typical duration of several days. A crucial quantity in the method is the parameter H, which controls the number of lag correlations that are modelled. For our dataset, we found that $H = 3$, corresponding to a lag of three days, provided a good representation of the storm-scale correlation between wave parameters. Should greater fidelity in the correlation structure be required, for instance to resolve infra-storm conditions, the method allows this. It would require analysing the original data at a finer temporal resolution, say hourly or three hourly, and correlations at a larger number of lags to be found, requiring additional calculation.

We note that the purpose of removing trend and seasonality in our study is to ensure that the detended and deseasonalised series y_{kt} are stationary so that we can use our simulation method to obtain simulated data for the original wave height, direction and period data. We have used the simplest method of [20] to remove trend and seasonality, where the trend and seasonality are estimated using our methods. Furthermore, we use the ADF test to check the stationarity of detrended and deseasonalised data. On the other hand, after we re-add trend and seasonality to the simulated y_{kt}, any potential biases in the residuals will disappear.

More sophisticated approaches are available. For example, where seasonal components vary significantly in amplitude and frequency over the dataset a least-squares wavelet analysis applied in a window-wise manner may be more suitable [15]. Other methods, such as the anti-leakage least-squares spectral analysis, allow simultaneous estimation of the trend and seasonal components.

The method described in this paper does not yield unrealistic jumps in the time series, as some earlier techniques did. The vector autoregressive (VAR) stochastic model presented in this study can be developed further to optimise the modelling of the seasonal component. In addition, the choice of parameter H, corresponding to the number of lag correlations that are modelled, could be automated rather than specified via a sequence of trial simulations.

Author Contributions: Conceptualisation, Y.C. and D.E.R.; methodology, Y.C.; software, A.V.; validation, A.V., Y.C. and D.E.R.; writing—original draft preparation, A.V. and Y.C.; writing—review and editing, D.E.R.; supervision, Y.C. and D.E.R.; project administration, D.E.R.; funding acquisition, Y.C. and D.E.R. All authors have read and agreed to the published version of the manuscript.

Appendix A

The correlation between y_{it} and y_{jt+h} is given by the following equation:

$$\rho_{ijh} = \frac{cov\left(y_{it}, y_{jt+h}\right)}{\sqrt{var(y_{it}), var\left(y_{jt+h}\right)}} \tag{A1}$$

However,

$$\begin{aligned}
cov(y_{it}, y_{jt-k}) &= E[(y_{it} - E(y_{it}))(y_{jt-k} - E(y_{jt-k}))]\\
&= E[y_{it}y_{jt-k} - y_{it}E(y_{jt-k}) - y_{jt-k}E(y_{it}) + E(y_{it})E(y_{jt-k})]\\
&= E[\ y_{it}y_{jt-k}] - E[y_{it}E(y_{jt-k}) - y_{jt-k}E(y_{it})] + E[E(y_{it})E(y_{jt-k})]\\
&= E(y_{it}y_{jt-k}) - E(y_{it})E(y_{jt-k}) - E(y_{jt-k})E(y_{it}) + E(y_{it})E(y_{jt-k})\\
&= E(y_{it}y_{jt-k}) - E(y_{it})E(y_{jt-k})
\end{aligned}$$

where *cov* is the covariance between y_{it} and y_{jt+h}

Thus, Equation (A1) is modified as follows:

$$\rho_{ijh} = \frac{E(yit\ yjt+h) - E(yit)E(yjt+h)}{\sqrt{var(y_{it}), var\left(y_{jt+h}\right)}} \tag{A2}$$

The mean of the probability distribution E of the product $y_{it} \times y_{jt+h}$ can be expressed analytically via the following equation:

$$E(y_{it}y_{jt+h}) = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} F_i^{-1}(\Phi(z_{it}))F_j^{-1}(\Phi(z_{jt+h}))\xi_{r_{ijh}} dz_{it} dz_{jt+h}$$

Hence yielding Equation (4) of Section 2.2:

$$\rho_{ijh} = \frac{\int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} F_i^{-1}(\Phi(z_{it}))F_j^{-1}\left(\Phi(z_{jt+h})\right)\xi_{r_{ijh}}(z_{it}, z_{jt+h})dz_{it}dz_{jt+h} - E(y_{it})E(y_{jt+h})}{\sqrt{var(y_{it})var(y_{jt+h})}}$$

Appendix B

Marginal distribution of y_{kt}:

The marginal distribution $F_k(y)$ of y_{kt} is described via an empirical distribution function of the observed time-series y_{kt}.

Sample mean, variance and autocorrelations:

Let the sample mean of y_{kt} be $\overline{y}_k$, sample variance $\hat{\sigma}_k$ and sample autocorrelation $\hat{\rho}_{ijh}$. Then

$$\overline{y}_k = \tfrac{1}{n}\sum_{t=1}^n y_{kt}, \qquad \hat{\sigma}_k = \tfrac{1}{n-1}\sum_{t=1}^n (y_{kt} - \overline{y}_k)^2, \qquad k = 1,2,3$$

$$\hat{\rho}_{ijh} = \frac{\sum_{t=1}^{n-h} (y_{it} - \overline{y}_i)(y_{it+h} - \overline{y}_j)}{\sqrt{\sum_{t=1}^n (y_{it} - \overline{y}_i)^2 \sum_{t=1}^n (y_{jt+h} - \overline{y}_j)^2}}, \qquad i,j = 1,2; h = 0,\cdots,H$$

where H is a fixed number that defines the maximum lag value we would like to consider when matching the autocorrelation structures between the simulated and the observed time-series, and in this study has been set equal to 3.

Solve non-linear equations and Yule-Walker Equation (3):

We need to solve the non-linear Equation (4) (here Equation (A3)) for r_{ijh} for all possible i,j and h.

$$\hat{\rho}_{ijh} = \frac{\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_i^{-1}(\Phi(z_{it})) F_i^{-1}\left(\Phi\left(z_{jt+h}\right)\right) \varphi_{r_{ijh}}\left(z_{it},\, z_{jt+h}\right) dz_{it} dz_{jt+h} - \overline{y}_i\,\overline{y}_j}{\hat{\sigma}_i \hat{\sigma}_j} \tag{A3}$$

where $\Phi(\cdot)$ is the standard normal distribution function,

$$\varphi_{r_{ijh}}\left(z_{it},\, z_{jt+h}\right) = \frac{1}{2\pi\sqrt{1 - r_{ijh}^2}} exp\left\{-\frac{z_{jt}^2 - 2r_{ijh}z_{it}z_{jt+h} + z_{jt+h}^2}{2\left(1 - r_{ijh}^2\right)}\right\} \tag{A4}$$

Then r_{ijh} can be estimated by solving Equation (A3) using, an iterating technique such as the Newton–Raphson's method:

$$r_{ijh}^{(m+1)} = r_{ijh}^{(m)} - \frac{f\left(r_{ijh}^{(m)}\right)}{f'\left(r_{ijh}^{(m)}\right)}, \; m = 0,1\ldots, \tag{A5}$$

where

$$f\left(r_{ijh}^{(m)}\right) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_i^{-1}(\Phi(z_{it})) F_j^{-1}(\Phi(z_{it+h})) \varphi_{r_{ijh}}^{(m)}\left(z_{it},\, z_{jt+h}\right) dz_{it} dz_{jt+h} - \overline{y}_i\,\overline{y}_j - \hat{\sigma}_i\hat{\sigma}_j\hat{\rho}_{ijh}$$

$$f'\left(r_{ijh}^{(m)}\right) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F_i^{-1}(\Phi(z_{it})) F_j^{-1}(\Phi(z_{it+h})) \varphi'^{(m)}_{r_{ijh}}\left(z_{it},\, z_{jt+h}\right) dz_{it} dz_{jt+h}$$

$$\varphi'^{(m)}_{r_{ijh}}\left(z_{it}, z_{jt+h}\right) = \varphi_{r_{ijh}}^{(m)}\left(z_{it}, z_{jt+h}\right) \xi_{r_{ijh}}^{(m)}\left(z_{it}, z_{jt+h}\right)$$

and

$$\xi_{r_{ijh}}^{(m)}\left(z_{it},\, z_{jt+h}\right) = \frac{r_{ijh}^{(m)}}{1 - r_{ijh}^{(m)2}} + \frac{z_{it}z_{jt+h} + r_{ijh}^{(m)2}z_{it}z_{jt+h} - r_{ijh}^{(m)}z_{it}^2 - r_{ijh}^{(m)}z_{jt+h}^2}{\left(1 - r_{ijh}^{(m)2}\right)^2}$$

to evaluate two double integrations more efficiently, the following method may be applied:

$$f\left(r_{ijh}^{(m)}\right) \approx \frac{1}{M}\sum_{l=1}^M F_i^{-1}((W_l^{(1)})) F_j^{-1}\left(\Phi\left(W_l^{(2)}\right)\right) - \overline{y}_i\,\overline{y}_j - \hat{\sigma}_i\hat{\sigma}_j\hat{\rho}_{ijh} \tag{A6}$$

$$f'\left(r_{ijh}^{(m)}\right) \approx \frac{1}{M}\sum_{l=1}^M F_i^{-1}((W_l^{(1)})) F_j^{-1}\left(\Phi\left(W_l^{(2)}\right)\right) \xi_{r_{ijh}}^{(m)}\left(W_l^{(1)},\, W_l^{(2)}\right) \tag{A7}$$

where $\left(W_l^{(1)}, W_l^{(2)}\right)$ is a random sample from the bivariate normal distribution with mean 0 and correlation $r_{ijh}^{(m)}$.

Then, a bisection method is applied to find the root corresponding to $f\left(r_{ijh}^m\right) \approx 0$. The index m corresponds to the number of the consecutive applications of the bisection method until r_{ijh} is estimated.

Once r_{ijh}s are available, ϕ_i and Σ_u for $i = 1,...,p$ can be obtained directly by solving Equation (2).

Obtain simulated data for the base process z_t:

The simulated base process can be obtained by the following steps.

(1)　Construct the covariance matrix P and find Q such that $P = QQ'$, where

$$P = \begin{bmatrix} \Gamma(0) & \Gamma'(1) & \cdots & \Gamma'(p-1) \\ \Gamma(1) & \Gamma(0) & \cdots & \Gamma'(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma(p-1) & \Gamma(p-2) & \cdots & \Gamma(0) \end{bmatrix}$$

where Γ' is the transpose matrix of Γ.

(2)　Obtain the initial value $(z_{-p+1}, z_{-p+2},...,z_0)$ by simulating $v_i \sim N(0,1)$, where $i = 1, ..., 3p$, and letting $(z_{-p+1}, z_{-p+2},...,z_0) = Q(v_1, v_2,...,v_{3p})>'$.

(3)　Find matrix Q_1 such that $\Sigma_u = Q_1 Q_1'$

(4)　Obtain simulated data for u_t by letting $u_t = Q_1(v_{1t}, v_{2t}, v_{3t})'$, where $v_{kt} \sim N(0,1)$, $k = 1,2,3$ and $t = 1, 2,..., T$, where T is the length of the simulated data.

(5)　Obtain the simulated data for the base process z_t by letting

$$z_t = \varphi_1 z_{t-1} + \varphi_2 z_{t-2} + \cdots + \varphi_p z_{t-p} + u_t, \qquad t = 1, \cdots, T.$$

Obtain simulated series $\widetilde{y}_{kt}$:

$$\widetilde{y}_{kt} = F_k^{-1}(\Phi(z_{kt}))$$

where $k = 1, 2, 3$ and $t = 1, 2, ..., T$, and $F^{-1}_k(x)$ is the inverse function of the empirical distribution of the y process and $\widetilde{y}_{kt}$ can be obtained by interpolation of the empirical distribution.

Transform $\widetilde{y}_{kt}$ back to obtain simulated data for E_{kt}:

Include the trend, seasonal components and stochastic trend into the simulated process $\widetilde{y}_{kt}$ to get the simulated data $\widetilde{E}_{kt} = T_{kt} + S_{kt} + \widetilde{y}_{kt}$, where $k = 1, 2, 3$ and $t = 1, 2, ..., T$.

References

1. Jägera, W.S.; Nagler, T.; Czado, C.; McCall, R.T. A statistical simulation method for joint time series of non-stationary hourly wave parameters. *J. Coast. Eng.* **2019**, *146*, 14–31. [CrossRef]
2. Borgman, L.E.; Scheffner, N.W. *Simulation of Time Sequences of Wave Height, Period, and Direction*; Technical Report DRP-91-2, USACE-WDC; Coastal Engineering Research Center: Vicksburg, MS, USA, 1991.
3. Callaghan, D.; Nielsen, P.; Short, A.; Ranasinghe, R. Statistical simulation of wave climate and extreme beach erosion. *J. Coast. Eng.* **2008**, *55*, 375–390. [CrossRef]
4. Corbella, S.; Stretch, D.D. Predicting coastal erosion trends using non-stationary statistics and process-based models. *J. Coast. Eng.* **2012**, *70*, 40–49. [CrossRef]
5. Li, J.; Winker, P. Time Series Simulation with Quasi Monte Carlo Methods. *Comput. Econ.* **2003**, *21*, 23–43. [CrossRef]
6. Barone, P. A method for generating independent realizations of a multivariate normal stationary and invertible ARMA(p, q) process. *J. Time Ser. Anal.* **1987**, *8*, 125–130. [CrossRef]
7. Shea, B.L. A Note on the generation of independent realizations of a vector autoregressive moving-average process. *J. Time Ser. Anal.* **1988**, *9*, 403–410. [CrossRef]
8. Camus, P.; Mendez, F.J.; Medina, R.; Cofiño, A.S. Analysis of clustering and selection algorithms for the study of multivariate wave climate. *J. Coast. Eng.* **2011**, *58*, 453–462. [CrossRef]
9. Soares, C.G.; Cunha, C. Bivariate autoregressive models for the time series of significant wave height and mean period. *J. Coast. Eng.* **2000**, *40*, 297–311. [CrossRef]
10. Cai, Y. Multivariate time series simulation. *J. Time Ser. Anal.* **2011**, *32*, 566–579. [CrossRef]

11. Cai, Y.; Gouldby, B.; Hawkes, P.; Dunning, P. Statistical Simulation of Flood Variables: Incorporating Short-Term Sequencing. *J. Flood Risk Manag.* **2008**, *1*, 1–10. [CrossRef]

12. Cario, M.C.; Nelson, B.L. Numerical methods for fitting and simulating autoregressive-to-anything processes. *J. Comput.* **1998**, *10*, 72–81. [CrossRef]

13. Biller, B.; Nelson, B.L. Modeling and generating multivariate timeseries input processes using a vector autoregressive technique. *ACM Trans. Modeling Comput. Simul. (TOMACS)* **2003**, *13*, 211–237. [CrossRef]

14. Cai, Y.; Huang, J.; Tang, Y.; Zhou, G. A simulation method for finite non-stationary time series. *J. Stat. Comput. Simul.* **2014**, *84*, 1563–1579. [CrossRef]

15. Ghaderpour, E.; Pagiatakis, S.D.; Hassan, Q.K. A Survey on Change Detection and Time Series Analysis with Applications. *Appl. Sci.* **2021**, *11*, 6141. [CrossRef]

16. Looney, D.; Adjei, T.; Mandic, D.P. A Novel Multivariate Sample Entropy Algorithm for Modeling Time Series Synchronization. *Entropy* **2018**, *20*, 82. [CrossRef] [PubMed]

17. Zhang, X.; Li, Y.; Gao, S.; Ren, P. Ocean Wave Height Series Prediction with Numerical Long Short-Term Memory. *Mar. Sci. Eng.* **2021**, *9*, 514. [CrossRef]

18. Fuller, W.A. *Introduction to Statistical Time Series*, 1st ed.; John Wiley and Sons: New York, NY, USA, 1976.

19. Mathworks. Available online: https://www.mathworks.com/help/econ/adftest.html#d123e114490 (accessed on 21 January 2022).

20. Brockwell, P.J.; Davis, R.A. *Time Series: Theory and Methods*; Springer: New York, NY, USA, 1991.

21. Bradbury, A.; Mason, T.; Poate, T. Implications of the spectral shape of wave conditions for engineering design and coastal hazard assessment–evidence from the English Channel. In Proceedings of the 10th International Workshop on Wave Hindcasting and Forecasting and Coastal Hazard Symposium, North Shore, Oahu, HI, USA, 11–16 November 2007.

22. Silverman, B.W. Density Estimation for Statistics and Data Analysis. In *Monographs on Statistics and Applied Probability*; Chapman and Hall: London, UK, 1986.