Oceanological and Hydrobiological Studies

International Journal of Oceanography and Hydrobiology

Volume 47, Issue 2, June 2018
pages (190-201)

ISSN 1730-413X
eISSN 1897-3191

# Spatial variability of long-term trends in significant wave height over the Gulf of Gdańsk using System Identification techniques

by

Jordan Badur, Witold Cieślikiewicz*

DOI: **10.1515/ohs-2018-0018**
Category: **Original research paper**
Received: **August 29, 2017**
Accepted: **November 21, 2017**

*Department of Physical Oceanography, Institute of Oceanography, Faculty of Oceanography and Geography, University of Gdańsk, Al. M. Piłsudskiego 46, 81-378 Gdynia, Poland*

**Abstract**

The significant wave height field over the Gulf of Gdańsk in the Baltic Sea is simulated back to the late 19th century using selected data-driven System Identification techniques (Takagi-Sugeno-Kang neuro-fuzzy system and non-linear optimization methods) and the NOAA/OAR/ESRL PSD Reanalysis 2 wind fields. Spatial variability of trends in the simulated dataset is briefly presented to show a cumulative "storminess" increase in the open, eastern part of the Gulf of Gdańsk and a decrease in the sheltered, western part of the Gulf.

**Key words:** Gulf of Gdańsk, wave climate, significant wave height, system identification, neuro-fuzzy systems, wave modeling

* Corresponding author: *ciesl@ug.edu.pl*

DE GRUYTER

The *Oceanological and Hydrobiological Studies* is online at oandhs.ocean.ug.edu.pl

## Introduction

While the wave climate of both the open part of the Baltic Sea (Cieślikiewicz & Paplińska-Swerpel 2008; Soomere & Räämet 2011a; Soomere et al. 2012; Soomere 2008; Räämet et al. 2009) and its northern or north-eastern coastal areas (Soomere & Räämet 2011b; Zaitseva-Pärnaste et al. 2011; Kelpšaitė et al. 2011) has received enough attention to qualify as a well-established field of study (Hünicke et al. 2015), little attention has been given to the wave climate of the Gulf of Gdańsk on the southern shore of the Baltic Sea, because long-term wave studies are scarce and the information provided by open sea wave modeling is not detailed enough.

The two most recent studies of wave climate in the Gulf of Gdańsk – Cerkowiak et al. (2015) and Różyński (2010) – used wave data that cover the period of 1958–2001 and come from selected points only: a single point in the open part of the Gulf of Gdańsk in the former case and two points in the latter case, one located in the sheltered part of the Gulf of Gdańsk and one in the open part. While these analyses provide valuable information, a dataset covering a longer period of time may be more suitable to assess wave climate changes in the last century. Finer resolution also seems more suitable to show spatial changes over the Gulf of Gdańsk in data trends.

The objective of our study was to simulate the significant wave height $H_s$ field over the Gulf of Gdańsk spanning the period of 1871–2008, using data-driven System Identification techniques.

In almost every case, any recommended method to hindcast a local wave field would involve the use of a series of nested wave spectral models (WISE Group 2007; Massel 1996) and best available wind fields; the use of the spectral wave model WAM (WAMDI Group 1988; Komen et al. 1994) is one of the possible choices. Although costly in terms of time and computational effort, this would have been possible had it not been for a few deficiencies of the 20 Century Reanalysis Project 2 (Compo et al. 2011) datasets (20CRv2). The type of data (3 h mean wind velocity) and their spatial resolution (2° – approx. 120 NM) lead to a substantial underestimation of wind velocities during storm maxima compared to the output of the Unified Model for Poland (UMPL-ICM) – regional meteorological model presented in Herman-Iżycki et al. (2002). Such an underestimation would naturally result in a much lower local significant wave height during storm maxima, if the 20CRv2 wind fields were used as a direct input to any spectral wave model. Because this underestimation is typical not only for the wave fields generated by Reanalysis 1 (Swail & Cox 2000), but also for the ERA40 wind fields (Reistad et al. 2011), it seems to be caused by the coarse resolution of global hindcasts. Furthermore, the effect of the coarse wind data resolution on the generation of lower values of a significant wave is already known (Bricheno et al. 2013).

To compensate for the deficiencies of the 20CRv2 wind fields, System Identification techniques are used in a two-stage procedure: first to improve the quality of the 20CRv2 wind fields and, secondly, to simulate the significant wave height $H_s$ field over the Gulf of Gdańsk using the improved, simulated wind fields.

The common trait of the System Identification methods is their ability to identify, using available data, a model linking multidimensional input $\boldsymbol{u}(t)$ and output $y(t)$ data. As the identified model does not need to reflect the physics of the modeled system, it is possible to identify the model using long-term, lower quality input data and available, high quality output data spanning a much shorter period of time and, consequently, use the model to generate higher quality data spanning a longer period of time.

First, we use this procedure to improve the quality of long-term, coarse resolution global 20CRv2 wind fields ($\boldsymbol{u}(t)$) using high resolution, local UMPL-ICM winds ($y(t)$) to simulate long-term (1871–2008), local wind fields as depicted in the upper section of the diagram in Figure 1. Secondly, $H_s$ System Identification models link the simulated wind fields ($\boldsymbol{u}(t)$) with the short-term, high resolution $H_s$ fields ($y(t)$) generated using a nested WAM model. As the WAM model was run using a 1 NM, (nautical mile – 1852 m) grid and high quality UMPL-ICM winds, the simulation output of the $H_s$ System Identification models should retain local features while providing the significant wave height for the entire 1871–2008 period. Once the significant wave height $H_s$ time series are simulated for each grid point, spatial variability of linear trends fitted to $H_s$ monthly statistics can be presented. As the specific techniques – which will be discussed later – are quite similar, the use of generic input $\boldsymbol{u}(t)$ and output $y(t)$ denotations seems advisable to show the similarities and distinctions.

To improve the quality of the 20RCv2 dataset, a number of linked linear autoregressive models with extra input (m-ARX) and short-term, high quality wind data from the UMPL-ICM regional model were used as shown in the upper section of the diagram in Figure 1. Corresponding parts of the long term, global 20CRv2 and short-term, local UMPL-ICM wind fields were used to identify the m-ARX models, which is represented by the upper row of the upper section of the diagram in Figure 1. Once identified, the m-ARX models were used to simulate the long term local wind field as

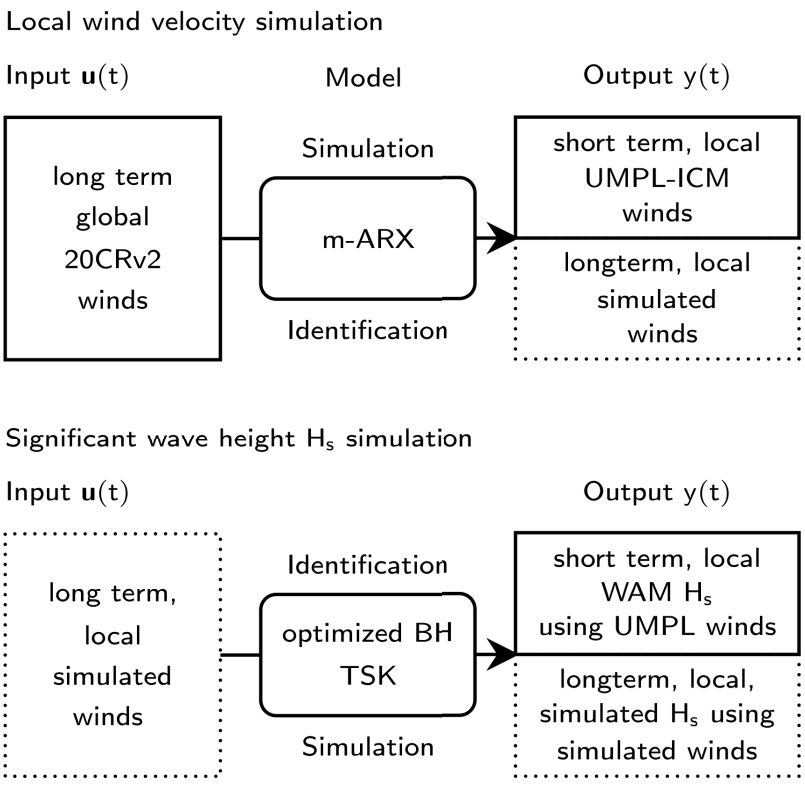


**Figure 1**

Flow diagram for wind velocity and significant wave height simulations. Non-simulated data, used for model identification and simulation, are included in solid-line rectangles. The simulated data are denoted by dotted-line rectangles; rectangles with rounded corners represent models. Dataset descriptions are provided in Table 1.

represented by the lower row of the upper section in Figure 1.

In the same way, the significant wave height models were identified and their output was simulated as represented by the lower section of Figure 1. Part of the simulated wind data and corresponding significant wave fields, previously generated with a nested WAM model (WAMDI 1988) using high quality UMPL-ICM wind data, were used to identify a set of Takagi-Sugeno-Kang (TSK) neuro-fuzzy systems

**Table 1**

Datasets used for identification, validation and simulation of local wind velocity and significant wave height $H_s$

| **Data set** Used… | Description |
|---|---|
| **UMPL-ICM winds** … to run the WAM model. As an input for the identification of wind scaling models. Half of data used for the verification. | 9 NM, 3 h, interpolated into 1 NM, 1 h grid, 1998–2001, 28 storms |
| **20CRv2 winds** … as an input for the identification and validation of wind scaling models. | 120 NM, 3 h, means interpolated into 1NM 1 h grid, 1871–2008 |
| **Simulated winds** … for $H_s$ simulation. Generated using scaling models with 20CRv2 data as an input. | 1 NM, 1 h, 1871–2008 |
| **WAM $H_s$** … to identify $H_s$ simulation models. Half of data used for verification. Generated using a locally nested WAM model and UMPL-ICM winds. | 1 NM, 3 h, 1998–2001, 28 storms |
| **Simulated $H_s$** … to generate the number of hours with $H_s$ exceeding a given threshold. Simulated using previously identified $H_s$ models and simulated winds. | 1 NM, 1 h, 1871–2008 |

(Takagi & Sugeno 1985) and to optimize a set of traditional empirical $H_s$ parametrizations (BH) as presented in the upper row of the lower section of the diagram in Figure 1. The outputs of both the TSK and the BH empirical parameterizations were simulated and joined together, depending on the weather conditions prevailing around the time step for which $H_s$ was simulated.

The use of System Identification techniques in wave modeling is a well-established concept with a number of core applications: (a) short-term wave forecasts based on measured data, (b) downscaling an offshore wave field to a nearshore area with a possible extension of the simulated time, (c) wind-based $H_s$ modeling including hindcasts – to name just a few – with various favorite techniques and a possible hybrid use of a spectral wave model. It has been shown that in the case of (a), the System Identification techniques work usually well (Özger 2010; Duan et al. 2016; Kumar et al. 2017) and can even outperform wave spectral models for short forecast lead times (Reikard & Rogers 2011; Sylaios et al. 2016). In the case of significant wave height downscaling, these techniques have an advantage of significant computation time reduction over wave spectral models while retaining good modeling quality (Herman et al. 2009; Londhe et al. 2016; Camus et al. 2011). Their performance increases significantly when $H_s$ on the area boundary is included in the input data. The primarily wind-based $H_s$ models of (c) can also perform well as demonstrated by (Mahjoobi & Mosabbeb 2009; Mahjoobi et al. 2008; Cañellas et al. 2010; Etemad-Shahidi & Mahjoobi 2009).

## Models, simulation methods and data used

### Data used

Since the proposed approach to simulation of significant wave height fields depends largely on the nature and shortcomings of the available data, these datasets are described in the following section and listed in Table 1.

As wind velocity fields of quality comparable with outputs of regional numerical weather forecast models spanning the entire last century are nonexistent, global wind velocity fields available from the 20th Century Reanalysis Project 2 (20CRv2) at NOAA/OAR/ESRL PSD were used (Compo et al. 2011). These data were generated using a global, coupled atmosphere-land model with surface pressure assimilation from the International Surface Pressure Databank, monthly sea surface temperature and ice concentration from the Hadley Centre (Compo et al. 2011). The available 3 h mean wind velocity fields are gridded with approx. 2° (approx. 120 NM) spatial resolution and span the period of 1871–2013 as presented in Table 1. However, due to the nature of the data, grid size and time resolution, the 20CRv2 global wind data are not suitable as a direct input into regional or local wave spectral models, which was demonstrated by smoothed out storm maxima and changed wind directions, when compared to the UMPL-ICM wind data. To correct the most persistent errors, m-ARX data-based models were used to identify the link between the 20CRv2 wind velocities and the high quality UMPL-ICM winds as denoted on the diagram in Figure 1.

The UMPL-ICM wind fields for the 1998–2001 period were generated at the ICM (Interdisciplinary Centre for Mathematical Computing, Warsaw University) using the UMPL (Unified Model for Poland) model, a derivative of the UM (Unified Model) developed by the ECMWF (Herman-Iżycki et al. 2002). The UMPL-ICM wind velocity fields have a relatively high temporal and spatial output resolution of 3 h and 9 NM.

Even after applying this simplified "scaling down" procedure, the quality of simulated wind velocities was not high enough to apply the WAM model directly. The final use of data-driven models, as represented by the lower row of the diagram in Figure 1, was supposed to artificially compensate for the most relevant deficiencies, while linking still deficient wind data with high quality wave height data.

The significant wave height fields generated mostly by Cieślikiewicz & Herman (2002) with the third generation spectral wave model WAM (WAMDI 1988; Komen et al. 1994) were used as the output data to train and verify parametric models. The WAM model was run on a nested grid with 1 NM resolution and 3 h model output, shallow options enabled (frequency table, depth refraction, dissipation including bottom friction, specific nonlinear interactions), and high-resolution (9 NM, 3 h) UMPL-ICM winds (Herman-Iżycki et al. 2002) as the driving input. The WAM dataset consists of model output for 29 storms which happened in 1998–2001.

Grids for WAM, UMPL-ICM and 20CRv2 datasets are shown in Figure 2. The striking contrast between the 20CRv2 global grid (approx. 120 NM) and the local 1 NM WAM grid is easily observed.

### Models and techniques used for simulation

Models described in this section are used either to simulate the wind velocity components (m-ARX) or, using the simulated wind velocity, to model the

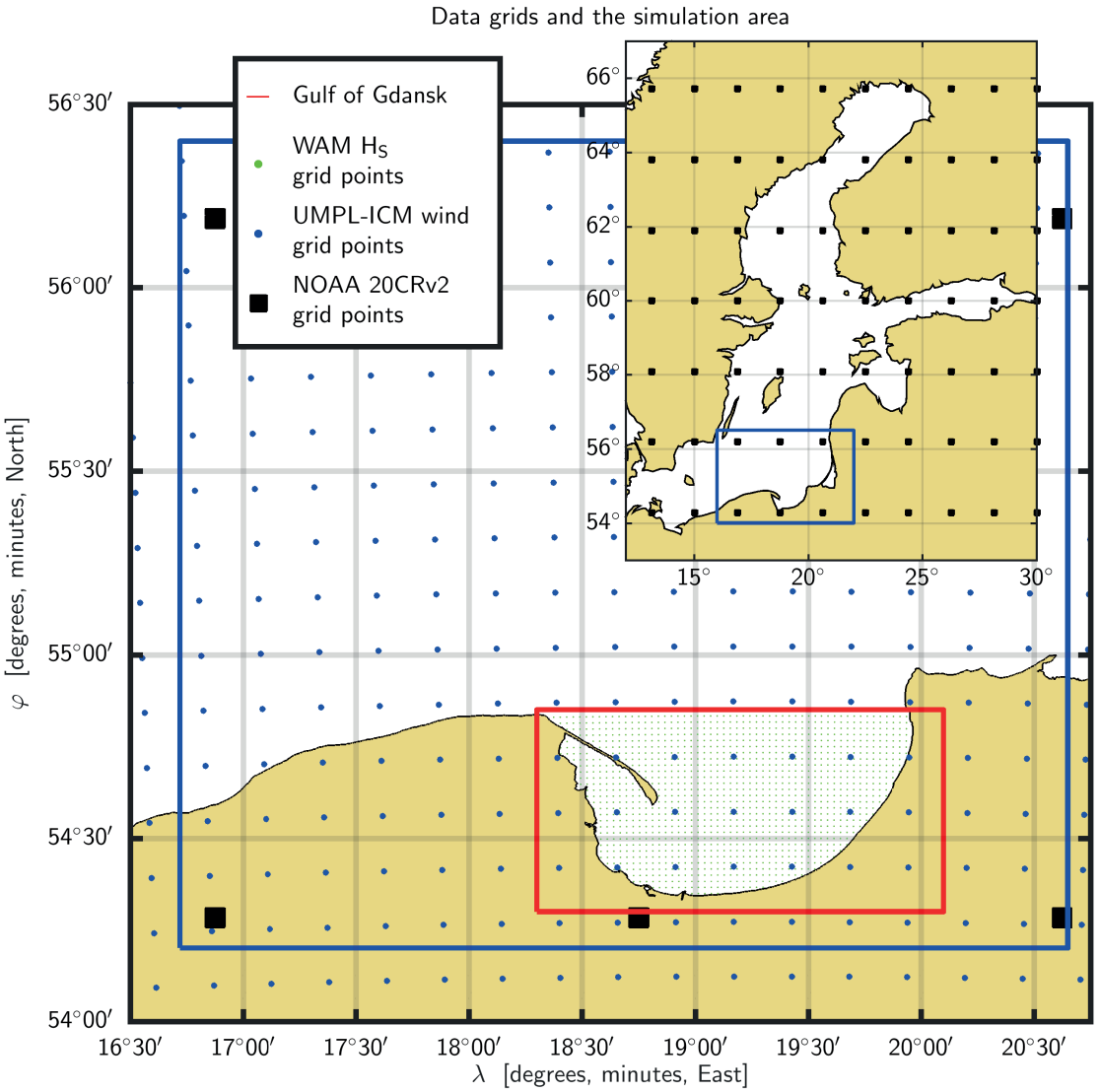


**Figure 2**

Data grids: NOAA 20CRv2 (black squares), UMPL-ICM (blue dots) and 1 NM nested WAM grid (small green dots)

significant wave height (TSK, optimized BH), according to the diagram in Figure 1. They are presented here using the generic notation for the multidimensional input $\boldsymbol{u}(t)$ and the single output $y(t)$ to show the similarities between the TSK and m-ARX as well as due to the fact that m-ARX is used to simulate both velocity components, which would call for a generic notation anyway.

### *Wind scaling using multi ARX (m-ARX)*

One of the most common ways to link the $I$-dimensional input $\boldsymbol{u}(t)$ and single output $y(t)$ data under $N$ different regimes is to associate a linear model with each of the regimes. These $n = 1,2,\ldots,N$ regimes can be described in numerous ways: providing their characteristic values ${}^{c}\boldsymbol{u}_n$, subsets or ranges of values in the input vector space. Then, the input data vector $\boldsymbol{u}(t)$ can be divided along with the corresponding output values into $N$ regimes to provide enough data to estimate $N$ linear models. The resulting output takes the form of an interpolation between $N$ linear models:

$$y(t)=\sum_{n=1}^{N} w_n(t)\sum_{i=1}^{I}\sum_{k=1}^{K} b_{ik}^{n} u_i(t-k) \qquad (1)$$

with interpolation coefficients $w_n(t)$ based on the inverse distance between $\boldsymbol{u}(t)$ and the regime centers ${}^{c}\boldsymbol{u}_n$; $w_n(t)=w_n\left(\|\boldsymbol{u}(t)-{}^{c}\boldsymbol{u}_n\|\right)$. For the $n$-th regime, $b_{ik}^{n}$

are typical coefficients of the ARX model (Ljung 1987) estimated using relevant data with $i$ indexing components of the input data vector $\boldsymbol{u}(t)$ and $k$ denoting the backward time shift. The input vector $\boldsymbol{u}(t)$ was simply equivalent to the locally interpolated 20CRv2 wind velocity vector, whereas the UMPL-ICM wind velocity components were used as the output $y(t)$ providing one model (1) for each output wind velocity component.

While it is perfectly possible to use one of the known identification (or optimization) algorithms to identify the values of $w_n$ or to use TSK to scale the wind velocities, better results were achieved using fixed, predetermined regimes with centers ${}^{c}\boldsymbol{u}_n$ and with distance-based coefficients $w_n(t)$. Local knowledge leads to an obvious choice for regime definitions as prevalently northerly and southerly winds with the mean wind velocity for both sectors used as the regime centers.

### *Significant wave height simulation using TSK systems*

Note that within the m-ARX framework given by (1), the regime characteristic values ${}^{c}\boldsymbol{u}_n$ are not optimized. Optimization is, however, possible within the frames of the Takagi-Sugeno-Kang (TSK) neuro-fuzzy system where the regime of the $n$-th linear system is defined as a fuzzy rule (Takagi & Sugeno 1985), associating the $n$-th linear model with fuzzy set centers ${}^{c}u_{ni}$, $i = 1,2,\ldots,IK$, with each $u_i(t-k)$, $k = 1,\ldots,$K treated as a separate input, following the code by Jang (1993).

For each rule, ${}^{c}u_{ni}$ centers can constitute a (pseudo) vector ${}^{c}\boldsymbol{u}_n = [{}^{c}u_{n1}, {}^{c}u_{n2}, \ldots, {}^{c}u_{nIK}]$ whose generalized distance from the input vector $\boldsymbol{u}(t)$ serves as a basis for the interpolation between the regimes. The generalized distance between $u_i(t)$ and the center of the fuzzy set $A_{ni}$ takes the form of a membership function $\mu_{ni}(t) = \mu_{ni}(\|u_i(t) - {}^{c}u_{ni}\|)$ (Jang 1993), usually of a Gaussian type. Within a TSK system, interpolation between different linear systems is provided by rule weights $w_n(t)$:

$$y(t) = \sum_{n=1}^{N} w_n(t) \sum_{i=1}^{IK} b_i^n u_i(t) \qquad (2)$$

$w_n(t) = w_n(\mu_{n1}, \mu_{n2}, \ldots, \mu_{nIK})$, which gives an opportunity to increase the influence of the $i$-th input. Note that the real difference between m-ARX and TSK systems is the automatic identification of both the location of fuzzy set centers ${}^{c}u_{ni}$ and parameters of membership functions $\mu_{ni}$. While the coefficients of linear models $b_i^n$ are estimated using the same methods as the ARX model, the ${}^{c}u_{ni}$ values are estimated using the backpropagation algorithm (Jang 1993), which supports the claim to the "neuro-" part of the neuro-fuzzy system name. Furthermore, the Radial Basis Function Networks, which are considered a simplified class of neural networks, can be shown to be equivalent to a TSK system if a radial function is used as a membership function (Jang & Sun 1993).

TSK systems are mainly used to model processes occurring under separate regimes (Bakhtyar 2011; Chaudhuri 2011) and have been applied to significant wave height hindcasting (Mahjoobi et al. 2008) and prediction (Zanageh et al. 2009) and were shown to perform similarly to artificial neural networks (Malekmohamadi et al. 2011; Mahjoobi et al. 2008). In our case, the TSK systems were chosen to simulate significant wave height for their inherent ease of identifying the underlying different regimes of the system. Because regimes are not as easily discernible, as in the case of wind scaling, an automated solution using TSK seemed advantageous.

Within our study, the TSK systems used for $H_s$ simulation employ a simulated wind velocity modulus and simulated normalized wind velocity components as an input $\boldsymbol{u}(t)$, while the WAM modeled $H_s$ was used as an output $y(t)$ to train the TSK. The precise composition of the input vector varied between the points of the Gulf of Gdańsk and contained one or two weighted mean wind velocity moduli from past 0–6 hours and normalized wind velocity components. For each point, up to four fuzzy sets with parameters identified automatically marked the regimes and the Gaussian membership function was used as an membership function to interpolate between them.

### *Optimization of significant wave height empirical parameterizations*

Significant wave height $H_s$ parameterizations of Breugem-Holthuijsen (Breugem & Holthuijsen 2006) and the Coast Engineering Manual (USACE 2008) used the simulated wind velocity as an input (Fig. 1) and were optimized to fit the previously described high quality WAM $H_s$ field. Both data-optimized engineering formulas (referred to as BH) proved to be equally useful as TSK, especially in conditions characterized by unique and rare wind velocities or calm weather. The output of the optimized BH parameterizations and TSK simulation were joined, depending on meteorological conditions (typical breezy to gale conditions were covered by TSK; calmer weather, extreme storms and winds from infrequent directions were covered by optimized BH parameterizations), to provide maximum accuracy and stability without the need to restrict the TSK simulation output by log-shaped restriction functions. Obviously, this procedure works on each

grid point separately.

Most engineering parameterizations combine $H_s$ time series generated under two idealized regimes of wave generation: limited by fetch $X$ and depth ${}^{X}H_S$ or time ${}^{t}H_S$ (Massel 1996). The first regime corresponds to a scenario when waves evolving under the wind speed $U_{10}$ influence reach the maximum wave height permitted by depth $d$ or fetch $X$, whereas the second one describes wave height evolution in time using the group velocity to convert wave evolution time into travelled distance. Because Breugem & Holthuijsen (2006) present ${}^{X}H_S$ parameterization as:

$$ {}^{X}H_S = \frac{0.24U_{10}^2}{g} tanh\left( k_3 \left( \frac{gd}{U_{10}} \right)^{1.14} \right) \left( tanh\left( \frac{0.0941\left( \frac{gX}{U_{10}^2} \right)^{0.79}}{tanh\left( k_3 \left( \frac{gd}{U_{10}^2} \right)^{1.14} \right)} \right) \right)^{0.572} \quad (3) $$

with $k_3$ denoting a depth-limiting coefficient, it must be augmented by time-limited ${}^{t}H_S$ parametrization, which is adopted from the Coast Engineering Manual (USACE 2008). It should be noted that although these parameterizations do not directly include all swell effects, it is possible to account for a slow decay of the wave height when wind direction shifts or wind speed decreases, using a longer $H_s$ memory. Furthermore, these parameterizations are optimized to reflect the behavior of WAM simulated fields, which is hoped to include typical swell conditions as modeled by the WAM, at least to some extent.

The above optimization procedure is carried out in stages, separately for each grid point. First, the fetch and depth limited parameterization (3) and regime relevant data are used to optimize geometric fetch values to enable realistic fetch $X(t)$ values, which allows to include waves incoming from directions significantly different than the wind direction during the event and partly accounts for directional wave spreading. Additionally, the depth limiting coefficient $k_3$ needs to be optimized at that stage. During the second stage, the wind velocity is adjusted $U_{10}(t) = a^s \| u(t) \|^{b_s}$, $s = 1,2$ separately for the two wind velocity sectors with sector division and both $a_s$ and $b_s$ values determined automatically for each grid point. Subsequently, coefficients of time-limited parameterization are estimated using wind and wave data corresponding to both regimes. The time-limited parameterization is augmented with an optimized parameter for the amount of change in wind direction necessary to trigger the transition to the regime of time-limited generation, which results in the inclusion of the incoming waves from directional sectors associated with former wind directions.

Finally, both ${}^{X}H_S$ and ${}^{t}H_S$ time series are joined and coefficients for the $H_s$ maxima enhancement are calculated separately for winds from the northern and southern sector. The difference between the definition of the sectors used for the adjustment of $U_{10}$ and $H_s$ is their automatic determination in the case of $U_{10}$ while in the case of $H_s$ the sectors are fixed.

Unfortunately, this procedure usually fails to provide a good estimate of $H_s$ in swell-dominated conditions, either due to topography-induced differences between the wind directions and wave propagation or fast changing meteorological conditions. This is especially pronounced in partially sheltered areas, where a small difference in wave direction is associated with a huge fetch difference. However, the optimization of selected coefficients of the BH parameterizations, using WAM-generated significant wave heights, can reduce most consistent errors. For instance, the optimization of fetch values to fit the WAM-generated $H_s$ may account for typical wave conditions (including the amount of swell) associated with a given wind direction, provided the data are consistent.

When a wave generation regime changes from fetch- to time-limited (e.g. due to wind direction change), a joining procedure must be applied to allow for a simultaneous slow decay of $H_s$ values associated with the old regime and $H_s$ built-up under a new regime, to some extent including swell.

As the optimization is carried out in stages, modifying one group of parameters at a time, there is no theoretical guarantee that a global minimum is found. However, as the search path and constraints for particular optimizations were chosen with care, using physical insight and previous runs of the genetic algorithm for a few selected points to determine typical parameter bounds, this approach may seem a reasonable compromise to decrease the computation time. A modified root mean square distance between simulated $H_s$ and the validation set of $H_S(t)$ is used as the optimization criterion.

To optimize the continuous parameters, the standard Matlab constrained solver (fmincon) with the interior-point algorithm is employed, using a relatively large number of iterations and a small required change of parameter values to stop the optimization process.

Having obtained all necessary coefficients, a simulation is carried out for the validation data set, and then for the entire 1871–2008 period.

## Simulation procedure

The outline of the simulation follows the diagram in Figure 1. First, the 20CRv2 wind velocity

components for each wet grid node were scaled down using the identified m-ARX models with better results achieved when scaling down the meridional velocity component. Next, simulated wind fields were used to simulate significant wave heights. As the TSK simulation proved to be more efficient in simulating typical stormy conditions, both the TSK and the optimized BH simulation results were joined, depending on the local wind velocity to generate the significant wave height field over the Gulf of Gdańsk for the entire 1871–2008 period.

All simulation models were identified and validated using disjoint data sets with data corresponding to WAM cold start omitted for both the identification and the validation. A short outcome of the validation process is presented in the next section "Results".

## Results

### Models validation

All models were validated against a disjoint, verification-only data set of 14 storms out of 28 (1998–2001), using the root mean square error (RMSE) and the scatter index (SI) defined as the RMSE value, normalized by the mean value of validation data. The use of SI and SI-based quality measures seems to be a common technique for wave model validation as demonstrated by its application in the validation of the third generation spectral wave models: WAM (WAMDI Group, 1988), SWAN (Ris et al. 1999) and Wavewatch III (Tolman 2002). In the case of significant wave height $H_s$, due to the small values of mean $H_s$ in coastal and nearshore waves, SI tends to underestimate the quality of fit (Ris et al. 1999).

The m-ARX simulated winds were validated against UMPL-ICM data. While the results of zonal wind component downscaling showed only a slight improvement, the scatter index (SI) values for the meridional component were significantly reduced from the maximum of 51% to a somewhat more acceptable level of 31%.

For brevity, only the validation results for the significant wave height simulation using TSK will be presented here, as values generated by this method are most likely associated with typical stormy conditions.

The scatter indices (SI) and the root mean square errors (RMSE) for each point are presented in Figure 3 and the comparison of significant wave height modeled with WAM and simulated with TSK during selected storm maxima is presented in Figure 4. Note that within the most western, sheltered part of the Gulf (Puck Bay), the scatter index values range from 5 to 27%, whereas in the open part of the Gulf, the SI values are almost constant within the range of 21–31% as observed in Figure 3. The highest SI values are found in the area occasionally protected by the Hel Peninsula and reach 50% in the nearshore zone. On the other hand, the comparison of significant wave height during storm maxima (Fig. 4) shows that the simulation retains spatial patterns, though mostly underestimates the $H_s$ values.

As the quality of both simulation methods is similar, the simulation error using both methods simultaneously does not exceed 30% of the simulated value in most areas of the Gulf of Gdańsk. Taking into account the discrepancy between the downscaled 20CRv2 wind fields and the UMPL-ICM wind used to run the WAM model generating the validation $H_s$, such an error is acceptable, especially if one is interested in the long-term statistics of the dataset and not in the actual values of $H_s$. It is worth noting that to capture a change in storminess, it is enough to obtain a good estimate of $N_H$ – the monthly number of hours with $H_s$ exceeding a given threshold that is specific for each point.

### Spatial patterns of trends

In this paper, the number of hours $N_H(t)$ with $H_s$ exceeding the point-specific threshold for each month is used as a measure of storminess. For each gridpoint, the threshold value is chosen as the $H_s$ value exceeded with a probability of 0.05 in the WAM-generated data (1998–2001). The $N_H$ values are calculated for each month and form a single time series mostly to filter out the daily variability and focus on the frequency of more significant stormy events. In this approach, the seasonal (or monthly) variability of $N_H$ over a typical year is not considered, neither is the long-time variability for a given month. The cumulative trend is calculated as a difference between the final and the initial values of the trend line fitted to $N_H$.

To compare the cumulative changes in $N_H$ attributed to the linear trend at different points more easily, these changes are normalized to the mean $N_H$ value at a given point to form a map in Figure 5. Unfortunately, the $N_H$, as defined above, is not an outstandingly consistent measure of the frequency of heavy sea conditions across all points of the Gulf of Gdańsk, especially nearshore where waves are depth limited. However, it gives similar estimates of $N_H$ for points with a similar shape of the cumulative distribution function for $H_s$. In these cases, the estimates are more consistent than when $N_H$ was defined using a fixed, common threshold for all points of e.g. 1 m. In this way, changes for each point in the

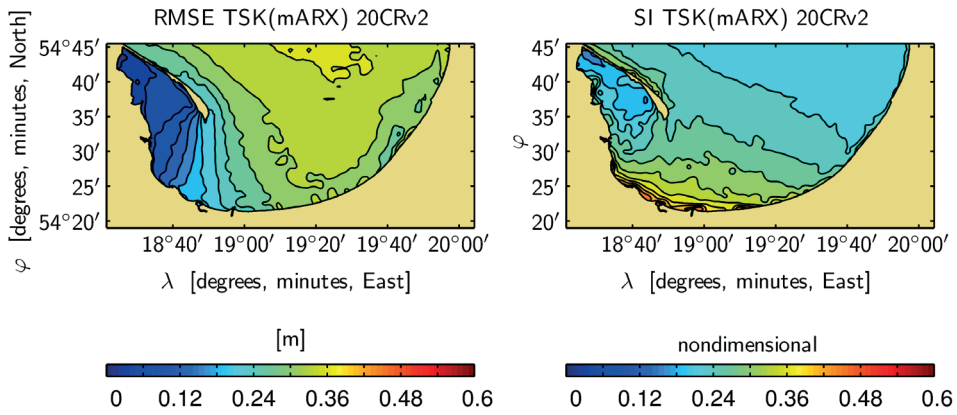


**Figure 3**

Root Mean Square Error (left) and scatter index (right) values for significant wave height simulation using the Takagi-Sugeno-Kang simulation method with m-ARX downscaled winds

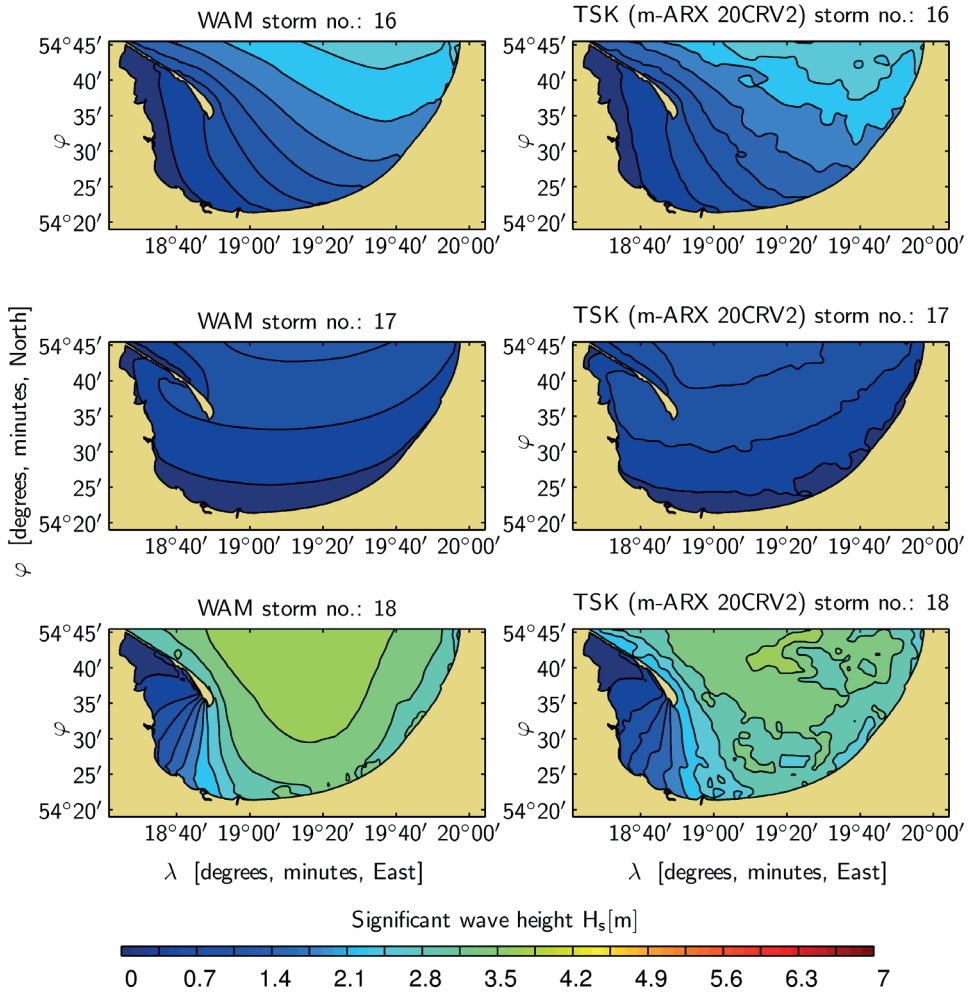


**Figure 4**

Significant wave height: Takagi-Sugeno-Kang simulated using downscaled winds (right) and WAM modeled with UMPL-ICM winds (left). Peaks of validation storms no. 16–18

Gulf of Gdańsk, when normalized with a point-mean value of $N_H$ can be compared with each other at the expense of information on absolute values. The authors believe that the presentation method applied in this study is advantageous as the commonly used change in mean monthly $H_s$ values, though interesting and readily understood, cannot be used to present spatial patterns in long-term trends. For example, a change of 0.2 m has a completely different meaning for points with a mean $H_s$ value of 1 m and, for example, 0.6 m.

As shown in Figure 5, the 1871–2008 cumulative change in $N_H$ is negative in the western part of the Gulf of Gdańsk (min. 25% of the mean value) and positive in the eastern part (max 80% of the mean value).

It is expected that further research will elaborate on long-term oscillation spatial patterns and provide a more detailed and coherent spectral description of $N_H$.

## Discussion and conclusions

A set of System Identification techniques was used to simulate significant wave height $H_s$ over the 1871–2008 period. The time-effective and relatively robust simulation is of sufficient quality to provide a reasonable estimate of $H_s$ statistics with 1 NM resolution, if not $H_s$ values alone in all areas of the Gulf of Gdańsk except the nearshore zone.

It should be noted, however, that any $H_s$ simulation technique taking a $H_s$-derived time series as an additional input would achieved a significantly better accuracy than the methods we employed in this study, as demonstrated by Herman et al. (2009) and by our trial simulations using the high quality UMPL-ICM wind field and an additional input of a $H_s$ time series at the nested grid boundary. As reported by Reikard & Rogers (2011), this feature manifests itself in the ability to outperform a spectral model in short-term forecasts based on measured $H_s$ values as the starting point. Furthermore, the performance of typical statistical System Identification methods decreases with increasing shoreline complexity, resulting in significant fetch inequalities to the point where spectral wave models are the only reasonable option – just as in the case of fast changing meteorological conditions.

The use of the point-specific threshold seems better suited to show spatial patterns than the 1 m fixed threshold as applied by (Różyński 2010), whereas the use of $N_H$ may provide a more robust estimate of storminess than the commonly used monthly mean $H_s$ (e.g. in Soomere & Räämet 2010; Soomere et al. 2012; Zaitseva-Pärnaste et al. 2011). Despite the use of different measures of storminess, the basic conclusion regarding the difference in storminess temporal trends between the eastern and western parts of the Gulf of Gdańsk corresponds to the findings of Soomere & Räämet (2011). Their study suggests a slightly stronger decrease in the annual significant wave height in the western open sea part of the Gulf of Gdańsk than in the eastern part in 1970–2007. Obviously, this should be reflected in the corresponding behavior of $N_H$.

Our analysis showed that the $N_H$ tendency to decrease with time in the western part of the Gulf of Gdańsk is weakening toward the eastern, open part of the Gulf, and eventually starts to increase.

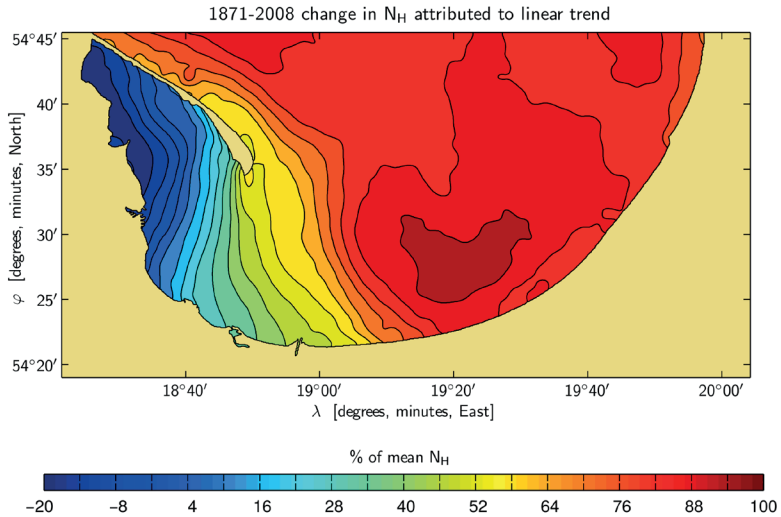


**Figure 5**

$N_H$ cumulative change due to the linear trend over the simulation period expressed as a percentage of $N_H$ mean

This easterly weakening of the decreasing temporal trend is consistent with the study by Soomere & Räämet (2011), although our analysis shows that easterly weakening in the decreasing trend results in the existence of a border area where no trend can be observed, followed by an area of a temporal increase in $N_H$ values, reaching up to 80% of the mean $N_H$ value at some points.

This difference is not surprising as our trends are fitted to the entire 1871–2008 period and data were simulated with 1 NM resolution to capture detailed changes in long-term wave field characteristics.

However, as the amount of change due to the linear trend should be compared with amplitudes of long-term oscillations, it cannot serve as a definitive climatological measure to asses storminess change over the modeled area and time. Further research is planned to elaborate on spatial patterns of long-term oscillations and their relative strength.

## Acknowledgements

The 20th Century Reanalysis Project data were provided by NOAA/OAR/ESRL PSD, Boulder, Colorado, the USA via their website http://esrl.noaa.gov/psd (Compo et al. 2011). As the support for the 20CRv2 dataset is provided by the US Department of Energy, the Office of Science, the Innovative and Novel Computational Impact on Theory and Experiment (DOE INCITE) program, and the Office of Biological and Environmental Research (BER), and by the National Oceanic and Atmospheric Administration Climate Program Office, the contribution of these institutions is gratefully acknowledged.

## References

Bakhtyar, R., Ghaheri, A., Yeganeh-Bakhtiary, A. & Jeng, D.-S. (2011). Cross-shore sediment transport estimation using fuzzy inference in the swash zone. *Journ. Franklin Inst.* 348(8): 2005–2025.

Breugem, W.A. & Holthuijsen, L.H. (2007). Generalised wave growth from Lake George. *J. Waterw. Port Coast. Ocean Eng.* 133(3): 173–182.

Bricheno, L.M., Soret, A., Wolf, J., Jorba, O. & Baldasano, J.M. (2013). Effect of high-resolution meteorological forcing on nearshore wave and current model performance. *J. Atmospheric Ocean. Technol.* 30: 1021–1037.

Camus, P., Fernando, J., Mendez, F.J. & Medina, R. (2011). A hybrid efficient method to downscale wave climate to coastal areas. *Coast Eng.* 58(9): 851–862. DOI: 10.1016/j.coastaleng.2011.05.007.

Cañellas, B., Balle, S., Tintoré, J. & Orfila, A. (2010). Wave height prediction in the Western Mediterranean using genetic algorithms. *Ocean Eng*. 37(8–9): 742–748. DOI: 10.1016/j.oceaneng.2010.02.006.

Cerkowniak, G.R., Ostrowski, R. & Szmytkiewicz, P. (2015). Climate change related increase of storminess near Hel Peninsula, Gulf of Gdańsk, Poland. *J. Water Clim. Change* 6(2): 300–312. DOI: 10.2166/wcc.2014.013.

Chaudhuri, S. & Middey, A. (2011). Adaptive neuro-fuzzy inference system to forecast peak gust speed during thunderstorms. *Meteorol. Atmos. Phys.* 114: 139–149.

Cieślikiewicz, W. & Herman, A. (2002). Wave and current modelling over the Baltic Sea and the Gdańsk Bay. In Proc. 28th Int. Coast. Eng. Conf. ASCE, 7–12 July 2002, (pp. 176–187), Cardiff, ASCE.

Cieślikiewicz, W. & Paplińska-Swerpel, B. (2008). A 44-year hindcast of wind wave fields over the Baltic Sea. *Coast. Eng.* 55: 849–905.

Compo, G., Whitaker, J., Sardeshmukh, P., Matsui, N., Allan, R. et al. (2011). The Twentieth Century Reanalysis Project. *Q. J. R. Meteorol. Soc.* 137: 1–28.

Duan, W.Y., Han, Y., Huang, L.M., Zhao, B.B. & Wang, M.H. (2016). A hybrid EMD-SVR model for the short-term prediction of significant wave height. *Ocean Eng*. 124: 54–73. DOI: 10.1016/j.oceaneng.2016.05.049.

Etemad-Shahidi, A. & Mahjoobi, J. (2009). Comparison between M5' model tree and neural networks for prediction of significant wave height in Lake Superior. *Ocean Eng*. 36(15–16): 1175–1181. DOI: 10.1016/j.oceaneng.2009.08.008.

Herman, A., Kaiser, R. & Niemayer, H.D. (2009). Wind-wave variability in shallow tidal sea – Spectral modelling combined with neural network methods. *Coastal Eng*. 56: 759–772.

Herman-Iżycki, L., Jakubiak, B., Nowiński, K. & Niezgódka, B. (2002). UMPL – the Numerical Weather Prediction System for Operational Applications. In B. Jakubiak (Ed.), *Research works based on the ICMs UMPL numerical weather prediction system results* (pp. 14–27). Warsaw: Publ. ICM.

Hünicke, B., Zorita, Z., Soomere, T., Madsen, K.S., Johansson, M. et al. (2015). Recent Change – Sea Level and Wind Waves. In The BACC II Author Team (Eds.), *Second Assessment of Climate Change for the Baltic Sea Basin* (pp. 155–185). Cham, Heidelberg, New York, Dordrecht, London: Springer International Publishing. DOI: 10.1007/978-3-319-16006-1_9.

Jang, R.J.-S. (1993). ANFIS: Adaptive-Network-Based Fuzzy Inference System. *IEEE Trans. Syst. Man Cybern.* 23(3): 665–685.

Jang, R. J.-S. & Sun, C.-T. (1993). Functional equivalence between Radial Basis Functions and Fuzzy Inference Systems. *IEEE Trans. on Neural Networks*. 4(1): 156–159. DOI: 10.1109/72.182710.

Kelpšaitė, L. Dailidiene, I. & Soomere, T. (2011). Changes in wave dynamics at the south-eastern coast of the Baltic

Proper during 1993–2008. *Boreal Environ. Res.* 16(A): 220–232.

Komen, G., Cavaleri, L., Donelan, M., Hasselmann, K., Hasselmann, S. et al. (1994). *Dynamics and modelling of ocean waves.* Cambridge: Cambridge University Press.

Krishna Kumar, N., Savitha, R. & Al Mamun, A. (2017). Regional ocean wave height prediction using sequential learning neural networks. *Ocean Eng*. 129: 605–612. DOI: 10.1016/j.oceaneng.2016.10.033.

Ljung, L. (1987). *System identification: Theory for user.* Englewood Cliffs, USA: Pretince- Hall.

Londhe, S.S., Dixit, P.R., Balakrishnan Nair, T.M., Sirisha, P. & Jain, R. (2016). A Coupled Numerical and Artificial Neural Network Model for Improving Location Specific Wave Forecast. *Appl. Ocean Res*. 59: 483–491. DOI: 10.1016/j.apor.2016.07.004.

Mahjoobi, L., Etemad-Shahidi, A. & Kazeminezhad, M.H. (2008). Hindcasting wave parameters using different soft computing methods. *Appl. Ocean Res.* 30: 28–36.

Mahjoobi, J. & Mosabbeb, E.A. (2009). Prediction of significant wave height using regressive support vector machines. *Ocean Eng*. 36(5): 339–347. DOI: 10.1016/j.oceaneng.2009.01.001.

Massel, S.R. (1996). *Ocean surface waves: their physics and prediction*. Singapore: World Scientific Publishing.

Malekmohamadi, I., Bazargan-Lari, M.R., Kerachian, R., Nikoo, M.R. & Fallahnia, M. (2011). Evaluating the efficacy of SVMs, BNs, ANNs and ANFIS in wave height prediction. *Ocean. Eng.* 38: 487–497.

Özger, M. (2010). Significant wave height forecasting using wavelet fuzzy logic approach. *Ocean Eng*. 37: 1443–1451.

Räämet, A., Suursaar, Ü., Kullas, T. & Soomere, T. (2009). Reconsidering uncertainties of wave conditions in the coastal areas of the Northern Baltic Sea. *J. Coast. Res.* 56: 257–261.

Reikard, G. & Rogers, E.W. (2011). Forecasting ocean waves: Comparing a physics-based model with statistical models. *Coastal Eng*. 58: 409–416. DOI: 10.1016/j.coastaleng.2010.12.001.

Reistad, M., Breivik, Ø., Haakestad, H., Aarnes, O.J., Furenik, B.R. et al. (2011). A high resolution hindcast of wind and waves for the North Sea, the Norwegian Sea and the Barents Sea. *Journ. Geophys. Res.* 116: C05019. DOI: 10.1029/2010JC006402.

Ris, R.C, Holthuijsen, L.H. & Booij, N. (1999). A third-generation wave model for coastal regions 2. Verification. *Journ. Geophys. Res.* 104(C4): 7667–7681.

Różyński, G. (2010). Long-term evolution of Baltic Sea wave climate near a coastal segment in Poland; its drivers and impacts. *Ocean Eng.* 37: 186–199.

Soomere, T. (2008). Extreme and decadal variations of the northern Baltic Sea wave conditions. In E. Pelinovsky & C. Kharif (Eds.), *Extreme Ocean Waves* (pp. 139–157). Amsterdam, Springer Netherlands.

Soomere, T. & Räämet, A. (2011a). Long-term spatial variations in the Baltic Sea wave fields. *Ocean Sci.* 7: 141–150.

Soomere, T. & Räämet, A. (2011b). Spatial patterns of the wave climate in the Baltic Proper and the Gulf of Finland. *Oceanologia* 53: 335–371.

Soomere, T., Weisse, R. & Behrens, A. (2012). Wave climate in the Arkona Basin, the Baltic Sea. *Ocean Sci.* 8: 287–300.

Swail, V.R. & Cox, A.T. (2000). On the use of NCEP-NCAR Reanalysis surface marine wind fields for a long term North Atlantic wave hindcasts. *J. Atmospheric Ocean. Technol.* 17: 532–545.

Sylaios, G., Bouchette, F., Tsihrintzis, V.A. & Denamiel, C. (2009). A fuzzy inference system for wind-wave modelling. *Ocean Eng.* 36(17–18 ): 1358–1365.

Takagi, R. & Sugeno, M. (1985). Fuzzy identification of systems and its applications to modelling and control. *IEEE Trans. Syst. Man Cybern.* 15: 116–132.

The WAMDI Group: Hasselmann, S., Hasslemann, K., Bauer, E., Janssen, P.A.E.M., Komen, J.G. et al. (1988). The WAM model – a third generation ocean wave prediction model. *J. Phys. Oceanogr.* 18: 1775–1810.

The WISE Group: Cavaleri, L., Alves, J.-H., Ardhuin, F., Babanin, A., Banner, M. et al. (2007). Wave modelling – The state of art. *Progr. Oceanogr.* 75: 603–674.

Tolman, H.L. (2002): *Testing of WAVEWATCH III Version 2.22 in NCEP as NWW3 Ocean Wave Model Suite*. NOAA/NWS/NCEP/OMB Technical Note Nr. 214.

USACE. (2008). *Coast Engineering Manual. USA Corps of Engineers.*

Zaitseva-Pärnaste, I., Soomere, T. & Tribštok, O. (2011). Spatial variations in wave climate change in the eastern part of the Baltic Sea. *Journ. Coastal Res.* 64: 195–199.

Zanaganeh, M., Mousavi, S.J. & Etemad-Shahidi, A.F. (2009). A hybrid genetic algorithm – adaptive network-based fuzzy inference system in prediction of wave parameters. *Eng. Appl. Artif. Intel.* 22: 1194–1202.