Donatelli, C.; Duran-Matute, M.; Gräwe, U.; Gerkema, T. (2022). Statistical detection of spatio-temporal patterns in the salinity feld within an inter-tidal basin. *Est. Coast. 45: 2345-2361*. DOI: 10.1007/s12237-022- 01089-3

Published version: https://dx.doi.org/10.1007/s12237-022-01089-3

NIOZ Repository: http://imis.nioz.nl/imis.php?module=ref&refid=352342

Statistical detection of spatio-temporal patterns in the salinity field within an inter-tidal basin

Carmine Donatelli[1,*], Matias Duran-Matute[2], Ulf Gräwe[3], Theo Gerkema[1]

Corresponding author: carmine.donatelli@nioz.nl

1. NIOZ Royal Netherlands Institute for Sea Research, Department of Estuarine and Delta Systems, Yerseke, The Netherlands
2. Eindhoven University of Technology, Department of Applied Physics, The Netherlands
3. Leibniz-Institute for Baltic Sea Research, Department of Physical Oceanography and Instrumentation, Warnemuende, Germany

*currently at: Department of Civil, Architectural and Environmental Engineering, University of Texas at Austin, Austin, TX, USA

Abstract

Salinity is a key factor affecting biological processes and biodiversity in estuarine systems. This study investigates temporal and spatial changes in salinity at a basin-wide scale for 2005-2015 in the Dutch Wadden Sea. Scan statistics is applied to track salinity variations systematically and to detect potential clusters, i.e. estuarine regions marked by anomalous high-salinity (or low-salinity) values in a certain period (i.e., strong deviations from the expected value in a statistical sense). Clusters' statistical significance has been assessed via Monte Carlo simulations. Particular attention is devoted to event-driven spatial and temporal patterns characterized by extreme salinity values since these episodes dramatically increase stress levels on organisms living in intertidal areas. Periodic components in the modeled salinity time series are identified using wavelet analysis and eventually removed from the signal before performing scan statistics. Wavelet analysis suggests that tides are the chief agent controlling salinity fluctuations in the system at within-day time scales, whereas no dominant periodicities were detected at longer time scales. Scan statistics reveal long-lasting clusters next to the main freshwater outlets and within the areas characterized by low water exchanges. In contrast, active regions of the estuary can efficiently counteract extreme events and quickly recover their pre-perturbation conditions. Finally, by analyzing the freshwater dispersal in the system, it is found that clusters' occurrence is related to episodic events characterized by extreme conditions in the southwesterly wind and freshwater discharge. This research demonstrates that scan statistics can be used as a powerful tool for spatiotemporal analyses of marine systems and for identifying data-clustering that may be indicative of emerging environmental hazards (e.g., due to climate change).

Keywords: salinity, scan statistics, estuaries, event-driven systems, extreme events

1. Introduction

Estuaries are dynamic transitional ecosystems where fresh water mixes with sea water [e.g., Van de Kreeke and Brouwer, 2017]. The ecological functioning and biodiversity of these unique coastal environments depend inherently on the spatial and temporal variability of the salinity field, which is driven by the compound actions of tides, winds, and river discharge [e.g., Cloern et al., 1989; Matsoukis et al., 2021]. Specifically, salinity varies strongly with freshwater inputs following the seasonal precipitation patterns [Teixeira et al., 2008; Teleshad and Khlebovich, 2010], and in response to extreme episodes (e.g., tropical cyclones) [Verdelhos et al., 2014]. Recent assessments forecast an increase in the frequency and intensity of extreme weather events with unforeseeable consequences for the stability of aquatic ecosystems [Doney et al., 2012; Wetz and Yoskowitz, 2013]. Rapid salinity changes dramatically increase the stress on organisms living in intertidal areas, threatening their habitats and impacting their capacity to survive and thrive [e.g., Wheatly, 1988]. Verdelhos et al., [2014] investigated mortality and behavioral responses of the bivalves *S.plana* and *C.edule* under abrupt changes in salinity, with no opportunity for them to acclimate physiologically to the new condition. These experiments revealed that both species present an optimal salinity range for their activity (for *S.plana*: 20–30; for *C.edule*: 20–25), as well as a reduction in the survival rate with salinity decline (100% mortality: for *S.plana*: <5; for *C.edule*: <10).

Despite the leading role of salinity on estuarine systems' ecology, a reliable methodology that identifies non-homogeneities in the salinity distribution is still missing. Specifically, there is a lack of knowledge in detecting singularities of the salinity field in a meaningful statistical way, i.e. regions and/or instants with a behavior different from the expected one. Furthermore, studies focusing on the spatiotemporal variability of salinity at a basin-wide scale [e.g., Ghezzo et al., 2011] and for several years [e.g., Schumman et al., 2006] are rare. Spatiotemporal analyses allow us to unravel the existence of unusual patterns within a specific study area over time and identify data-clustering that may be indicative of emerging environmental hazards. Thus, the statistical detection of a cluster (a spatial temporal phenomenon existing in a region for a certain time interval) bears important implications for descriptive and predictive purposes. For instance, Carniello et al., [2016] analyzed the spatiotemporal evolution of suspended sediment concentrations (SSC) in the Venice Lagoon, using the 'Peak-Over-Threshold' method (POT) and following the framework proposed by D'Alpaos et al., [2013] for the analysis of the wave-induced bottom shear stresses. Specifically, the POT-method is based on analyzing the data exceeding a defined threshold and understanding the system's behavior to extreme events [Cramér and Leadbetter, 1967; Leadbetter, 1990]. Their findings provided a statistically meaningful characterization of emerging SSC patterns. They unraveled the mechanisms governing sediment dynamics and the associated long-term morphological changes in the lagoon [Carniello et al., 2016].

In this study, we propose scan statistics [Naus, 1965; Kulldorff and Nagarwall, 1995; Kulldorff, 2001] as a tool to study spatiotemporal variability in marine systems. Scan statistics seeks to reveal whether the incidence of a certain event in a defined spatial-temporal subset is anomalous compared to the incidence within the entire study area [e.g., Robertson et al., 2010]. This methodology has been introduced by Kulldorff [1997, 1999a, 1999b] to detect non-homogeneities within spatial and spatiotemporal datasets in epidemiology and assess their statistical significance without making any a priori assumption about clusters' location and size. Subsequently, it has been applied with success in other contexts: wildfires [e.g., Tuia et al., 2007], water pollution [e.g., Carstensen, 2007] and astronomy [e.g., Marcos and Marcos, 2008]. To this end, we have employed high-resolution numerical modeling simulations to the Dutch Wadden Sea (DWS), a mesotidal back-barrier bay characterized by semidiurnal tides (Fig. 1). The system is a world UNESCO world heritage site because of its ecological significance. It is connected to the North Sea by several inlets (see A to E in Fig. 1) and has two primary sources of fresh water (the sluices at Den Oever and the Kornwerderzand), located at the closing dike dividing the Lake IJssel and the Dutch Wadden Sea.

Recent model studies, based on high-resolution numerical simulations covering 2009-2011, indicate that the system is strongly event-driven, primarily due to wind forcing. This driver creates a high variability in the exchange (and its constituents) between the tidal basins and the North Sea [Duran-Matute et al., 2014, 2016]. The variability even extends to annual-mean values. This means that yearly mean or median values need to be accompanied by higher-order statistics to characterize the system in a meaningful way. Given the large inter-annual variability in the transports and the system's state, long-term trends can only be identified by considering several years. Thus, the three years of modeling were not sufficient to identify the system's typical long-term state and variations thereof.

In this paper, we have extended the analysis to 11 years (2005-2015), and we have focused our study on salinity, interpreting the results concerning their biological and ecological relevance. Besides, we have employed wavelet transforms to identify the dominant periodicities in salinity within the 11-year record and determine what hydrodynamic processes govern salinity oscillations in the system. The paper is organized as follows. Section 2 describes the methodology (e.g., model setup, wavelet analysis, scan statistics) and presents a section dedicated to simple applications of scan statistics (subsection 2.4). In Section 3, we study salinity variability at different time scales to reveal potential periodic components in the time series through wavelet analysis (subsection 3.1). Further, we use scan statistics (subsection 3.2) to identify the presence of regions experiencing extreme salinity values in the DWS. In Section 3, we also explain why these extreme values occur by analyzing the freshwater dispersal in the system (subsection 3.3). Finally, the discussion is presented in Section 4 and the conclusions are outlined in Section 5.

2. Methods

We apply scan statistics to detect regions within the Dutch Wadden Sea, which experience anomalous salinity levels over 2005-2015. Anomalous values are defined here as deviations from the expected in a statistical sense. Areas within the estuary characterized by anomalous salinity values in a certain time frame are called clusters (or singularities). Before using this methodology, it is necessary to remove the dominant periodicities (i.e., patterns in a signal that occur at regular time intervals) from the original time series. This step is critical to eliminate those oscillations associated with the system's intrinsic variability (e.g., seasonal variability) that might lead to spurious singularities when performing scan statistics. After removing these components from the original signal, scan statistics will detect only clusters related to anomalous events. Here, wavelet analysis is employed to reveal the presence of periodic components in the modeled time series. The main advantage in using spectral analyses is that they enable the identification of these components straightforwardly. For instance, a strong seasonality will show up in the spectrum as a peak with a specific periodicity. The removal of this component will therefore de-seasonalize the time series.

The entire procedure can be summarized in the following steps (Fig. 2): (i) we perform high-resolution numerical modeling simulations to compute salinity values within the DWS (subsection 2.1); (ii) wavelet analysis (subsection 2.2) is used to identify the dominant periodicities in the time series at different time scales (e.g., hour-to-hour, day-to-day, seasons); (iii) after removing potential periodic components from the time series, scan statistics (subsection 2.3) is applied to detect spatio-temporal clusters in the salinity field over the studied period. Finally, we relate the occurrence of the clusters detected by scan statistics with the variability in the external forces, i.e., freshwater discharge and wind energy (see supplementary material).

2.1 Hydrodynamic model

We have simulated the circulation in the Dutch Wadden Sea with a 3D hydrodynamic model from January 2005 to December 2015, using the General Estuarine Transport Model (GETM) [Burchard and Bolding, 2002]. The freshwater dispersal in the system has been tracked using Eulerian passive tracers, adopting an approach similar to that proposed by Meier [2007] and Zhang [2009]. Two distinct tracers have been employed, one for each freshwater outlet. In this study, we have used the Framework for Aquatic Biogeochemical Models (FABM, Bruggeman and Bolding, [2014]) passive tracer module coupled with GETM to monitor the freshwater's fate in the DWS. GETM solves the advection-diffusion equation for the passive tracers employing the same method used for salinity and temperature. Bathymetric data around the year 2009 has been used for the entire period of analysis (with a closure at the most easterly watershed) identical to the one used in Duran-Matute et al., [2014]. The numerical model of the Dutch Wadden Sea is the end-member of four nested models as described by Gräwe et al., [2016]. The DWS' numerical grid has 25 vertical sigma layers and a horizontal resolution of 200 m. Water levels at the numerical domain boundaries have been obtained by superimposing astronomic tidal elevations, computed by using the Oregon State University Tidal Prediction Software (OSU-TPS), and surge levels calculated employing a vertically integrated North Atlantic model forced by surface winds and air pressure. We have used atmospheric data with a spatial resolution of 12 km and a temporal resolution of 1 hour (reanalysis data of UERRA, Ridal et al., [2017]). Time varying profiles of salinity and temperature obtained from the Climate Forecast System Reanalysis (CFSR) meteorological data of the U.S. National Centers for Environmental Prediction (NCEP) were employed for the three-dimensional boundary conditions of the North Sea model (which is forced by complete meteorological forcing, salinity, temperature and freshwater discharge), and then extracted every 2 hours and applied to the boundaries of the three-dimensional southern

North Sea model (spatial resolution of 600 m and 42 vertical layers) [Gräwe et al., 2015]. Finally, vertical profiles of salinity and temperature are extracted from the 600 m model with a temporal resolution of 1 hour and linearly interpolated to the boundaries of the DWS' model. Rijkswaterstaat has provided times series of freshwater discharges at the main sluices. The model has been validated with available observational data involving measurements of sea surface height, current velocity, temperature, and salinity. The comparisons showed that the model faithfully reproduces the hydrodynamics in the DWS [Duran-Matute et al., 2014; Gräwe et al., 2016; Gerkema and Duran-Matute, 2017]. Further details of the model validation are presented in the supplementary material.

2.2 Wavelet analysis and identification of the periodic components in the time series

Spectral analysis is a widely used tool to identify the dominant scales of variation in time series. Traditional spectral analyses (e.g., Fourier transform) decompose a time series as the sum of sine waves with different frequencies. Therefore, they are not well suited to characterize non-stationary signals (with a frequency content that varies over time). Here, we have applied wavelet analysis (WA) [Torrence and Compo, 1998] to study salinity fluctuations across different time scales and to detect the presence of periodicities in the original time series [e.g., Carstensen, 2007]. Periodic components of a time series are related to events occurring with a particular frequency over time and thus adjustments of the signal are needed before applying scan statistics.

WA uses a mother function, suitably scaled and translated in time, to calculate wavelet coefficients and the associated wavelet power spectrum. The Morlet wavelet was used as a mother function because it provides a good balance between time and frequency localization [Grinsted et al., 2004]. The wavelet outputs (e.g., amplitude) are estimated by varying the wavelet scale s and translating the scaled wavelets in time. In particular, the smallest resolvable scale, s_0, is defined as a multiple of the sampling interval, dt (e.g., $s_0 = 2dt$), while the spacing between the discrete scales, dj (scale step), was set to 1/4 (4 suboctaves per octave). The total number of scales is then computed based on dj and the number of octaves (i.e., 7) as in Torrence and Compo [1997]. The statistical significance of the results is evaluated by comparing the wavelet spectrum against the 95% confidence level of the power spectrum generated by the corresponding red noise. Statistically significant regions are shown with thick black contours in the spectra. We employed an autoregressive AR(1) model with the same autocorrelation coefficient (i.e., lag-1, this coefficient is the correlation between the time series and itself but shifted by one time step) as the observed time series (i.e., salinity time series). The wavelets are normalized to have unit energy at each scale [Torrence and Compo, 1997]. Further details about wavelet analysis can be found in Torrence and Compo [1998]. It is worthwhile noticing that the wavelet method developed by Torrence and Compo [1998] may have bias issues as previously found and rectified by Liu et al. [2007], who demonstrated that the transform coefficient squared and divided by the associated scale is a physically consistent definition of energy for the wavelet power spectrum. For this reason, we used the approach developed by Liu et al., [2007] to perform the wavelet analysis.

2.3 Spatial-temporal scan statistics

We have studied salinity changes over the 11 years, employing retrospective spatial-temporal scan statistics [Kulldorff, 1999a, b] to detect and analyze potential clusters in the Dutch Wadden Sea. The freely available SaTScan software (www.satscan.org) developed by Kulldorf [1997] has been used with this aim. The advantages of using SaTScan with respect to other software packages are discussed in detail by Robertson and Nelson [2010]. The spatial-temporal scan statistics employ a cylindrical window that scans the entire back-barrier estuary for different time intervals. The base of the cylinder reflects space, while the height represents time (Fig. 3). The scanning window is moved within a three-dimensional domain (2D space + 1D time). More specifically, for each possible geographical location, the cylinder varies its radius from zero to a specified maximum value, visiting each possible time interval (i.e., the height of the cylinder changes between 2 time-steps and a specified maximum value). Each window is considered a possible candidate cluster. The maximum spatial and temporal cluster sizes are set based on a sensitive analysis (see supplementary material). The null hypothesis is that the risk of encountering unusual patterns in the salinity field remains the same inside and outside the scanning window. The alternative hypothesis is that the risk is different. The observed cases (i.e., actual salinity values) inside and outside the scanned area are compared to the number of expected cases (i.e., expected salinity values in a statistical sense), calculated using an equal risk hypothesis for each cylinder. A likelihood ratio (LR) is calculated for each sub-area of the numerical domain scanned by the window. Under the Poisson assumption, the LR is computed as follows [Fraker et al., 2008]:

$$LR = \left(\frac{c}{\mu}\right)^{c}\left(\frac{C-c}{C-\mu}\right)^{C-c} I() \tag{1}$$

where C is the total number of cases, c is the observed number of cases within the window, and μ is the covariate-adjusted expected number of cases within the window under the null hypothesis. I() is an indicator function. If the goal is to identify only clusters with high rates, I() equals 1 when the window has more cases than expected under the null hypothesis and 0 otherwise. The opposite is true when we scan only for clusters with low rates: I() is always 1 when detecting clusters with either high and low rates is performed. We used the log-likelihood ratio (LLR) to evaluate the likelihood of clustering. Specifically, the cylinder with the maximum likelihood ratio is identified as the 'primary' cluster that is least likely to have occurred by chance. Secondary clusters are identified with an iterative process, as shown in Kulldorff [1997]. This process is as follows. In the first step, the primary cluster is detected and removed from the datasets. Then, a new analysis is performed using the remaining data. After finding the most likely candidates, the level of significance of these clusters is evaluated using Monte Carlo simulations. In particular, a large number of random datasets is generated under the null hypothesis (i.e., no anomalies in the salinity field) to determine the statistical significance of the found cluster (the *p-value* must be smaller than the selected level of significance). The *p-value* of a specific scanned area is calculated such that

$$p - value = \frac{R_B + 1}{R + 1} \tag{2}$$

where R_B is the number of replica datasets with a maximum LR higher than the maximum LR of the real data, and R is the total number of random datasets. The scanned area is significant at α=0.05 if its LLR scores higher than 95% of the random datasets. In addition to the LLR, we employed the relative risk (RR) to quantify the observed cases' unexpected degree. Specifically, it is calculated as the ratio between the observed cases and the expected number of cases inside the scanning window versus outside:

$$RR = \frac{c/\mu}{(C - c)/(C - \mu)} \tag{3}$$

For each detected cluster, SaTScan provides spatial information (e.g., coordinates, radius), the corresponding time frame, and a *p-value*.

2.4 Application of scan statistics to simple scalar fields

This section applies scan statistics to simple scalar fields defined on a square domain (Fig. 4). These fields depend on space and time and they assume a value of 0 or 1 in each cell for t>0. The following functions are used to describe the two scalar fields at a certain instant (t=$\bar{t}$):

$$S_1(x,y,t=\bar{t}) = \begin{cases} 1, & if\ 2 \leq x \leq 4\ and\ 2 \leq y \leq 4 \\ 1, & if\ x = 5\ and\ y = 5 \\ 0, & otherwise \end{cases} \quad (4)$$

$$S_2(x,y,t=\bar{t}) = \begin{cases} 1, & if\ 3 \leq x \leq 4\ and\ 3 \leq y \leq 4 \\ 1, & if\ x = 5\ and\ y = 5 \\ 0, & otherwise \end{cases} \quad (5)$$

We have used spatial scan statistics to detect anomalous trends in this field for $t=\bar{t}$. This procedure can easily be extended to the 3D field by including time as a variable. In that case, spatio-temporal scan statistics should be employed. At this stage, we have decided to work with a 2D field and scan the entire area with a circular window. The goal is to identify the primary cluster. First, we can consider only the three circles plotted in Fig. 4b as candidate clusters, regarding each cell with value 1 as a 'case'. The number of cases included in the red scanning window is 9, while the number of expected cases (μ) under the null hypothesis H_0 is 2.5. The latter is calculated as follows:

$$\mu = \bar{p} \cdot C/P \quad (6)$$

with $\bar{p}$ the population in the pixels within the full red circle (=9), C the total number of cases in the entire domain (=10) and P the total population (=36). We adopted the following population function for this test case (Fig. 4a):

$$P_{1,2}(x,y,t=\bar{t}) = 1, \text{ everywhere} \quad (7)$$

Then, the log-likelihood ratio (LLR= 9.5) is computed. The second window (yellow circle) does not respect the minimum cluster size (at least two cases), and therefore we stop the procedure and use new windows to scan the domain. The same procedure is applied to the third potential primary cluster (green circle). As the maximum cluster size is set to 50% of the total population (a common assumption in scan statistics), this window does not respect this condition. Therefore, it cannot be a potential cluster. On the contrary, the green window in Fig. 4c matches all the conditions and can be a potential cluster (LLR = 6.9). We can detect only windows with log-likelihood ratios smaller than 9.5 and 6.9 in each test case. Thus, the clusters identified by the red scanning window in Fig. 4b and by the green scanning window in Fig. 4c are the primary clusters. Table 1 summarizes the information associated with each cluster. Finally, we defined a third scalar field (S_3) and a new population function (P_3):

$$P_3(x,y,t=\bar{t}) = \begin{cases} 30, & if\ x = 5\ and\ y = 5 \\ 1, & otherwise \end{cases} \quad (8)$$

$$S_3(x,y,t=\bar{t}) = \begin{cases} 1, & if\ 2 \leq x \leq 4\ and\ 2 \leq y \leq 4 \\ 30, & if\ x = 5\ and\ y = 5 \\ 0, & otherwise \end{cases} \quad (9)$$

In this case, we detected a single cluster statistically significant within the entire domain (yellow window, Fig. 5b). The number of expected cases computed with equation (4) is 18, while the number of counted cases is 30. Table 1 summarizes the information associated with this cluster. Larger scanning windows, including ‘pixel 29', would violate the maximum cluster size restriction, and therefore, they cannot be considered candidate clusters.

The same approach will be used for the salinity field, e.g. the expected salinity value in a grid cell is computed using Eq. (6). More specifically, the salinity value in a grid cell at a certain time is defined as a ‘number of cases’. By definition, the population has to be greater than the number of cases in a cell. Thus, we consider as a ‘population’ the long-term mean salinity value experienced by a grid cell over the studied period multiplied by 2 (it is simple to demonstrate from Eq. (6) that this choice does not impact the results). Finally, C and P are defined considering the values of c and p previously computed for the entire numerical grid in each time step.

3 Results

3.1 Dominant periodicities

First, the hourly depth-averaged salinity values from the high-resolution numerical simulations were collected for the entire model run of 11 years for all grid cells. The daily averages were calculated, from which we computed a distribution with the mean and standard deviation for each pixel (i.e., grid cell) of the numerical domain. Since these statistical parameters are space-dependent, we mapped their spatial distributions. Figure 6 shows that the deep channels experience the highest salinity levels, whereas the tidal flats and the areas next to the sluices present the largest deviations from mean values. However, Fig. 6 does not reveal if these salinity fluctuations occur at a given frequency. Hence, we used wavelet analysis to identify the dominant periodicities in the modeled time series and reveal the system's chief agent governing salinity fluctuations.

It is worthwhile recalling that the vertical axis in the wavelet spectra plots represents the periodicities in a time series, while the horizontal axis depicts the time. Yellow regions surrounded by black contour lines represent statistically significant areas, whereas blue regions indicate low wavelet power values. Moreover, the cross-hatched areas identified by two thick black lines define the cone of influence; results within this region are not considered because edge effects are strong [Torrence and Compo, 1998]. Three different locations (P1, P2, and P3) were selected (Fig. 1). These points represent three distinct environments in the system: (i) deep channels (depth ~ 15 m) connecting the North Sea with the tidal basins; (ii) subtidal platforms (depth ~ 3 m) next to the main sluices; (iii) shallow area (depth ~ 1.5 m) located in the central region of the system. The results of the wavelet analysis are shown in Figs. 7 and 8. We employ the original hourly signal to identify the presence of dominant periodicities. Specifically, the wavelet power spectra (Fig. 7) reveal the existence of a peak (at 12 hours 25 minutes) throughout the entire studied period, which is statistically significant. Interestingly, the significant peaks are better defined and concentrated at 12 hours 25 minutes in locations P1 and P3, whereas the peak in point P2 is broader due to the proximity to the sluices. Since we need to eliminate those fluctuations associated with the system's intrinsic variability before performing scan statistics, we compute the daily averages of the original time series to: (i) remove this periodicity and (ii) identify the presence of dominant periodicities at longer time scales (day-to-day, seasonal and annual time scales).

Figure 8 is obtained using the daily time series and shows that salinity fluctuations present only small regions which are statistically significant at day-to-day time scale (i.e., periodicities smaller than 30 days). The significant area is spread across a wide range of periodicities, and it is not uniform throughout the studied period. The daily average (instead of the tidally average) may introduce an aliasing resulting in an oscillation with a periodicity about 15 days which might produce spurious effects resembling the spring-neap tidal cycle. However, we do not identify any region statistically significant with this periodicity. Finally, we focus on salinity variability at time scales of months and years. The spectra in Fig. 8 do not reveal any dominant periodicity but only limited areas which are statistically significant (i.e., period of ~200 days between January 2010 and January 2013 in P1 and P2). Therefore, we use the daily averages when applying scan statistics.

3.2 Detection of spatio-temporal clusters in the DWS

In this section, we employed scan statistics to identify singularities in the salinity field within the Dutch Wadden Sea over the period 2005-2015. Our analysis revealed the presence of several clusters. Here, we report only clusters statistically significant with RR > 1.20 (see Eq. (3)) for those characterized by high salinity values or with RR < 0.8 for those characterized by low salinity values. Figures 9 and 10 depict clusters with salinity levels higher and lower than the expected ones from 2005 to 2008. Cluster information (time frame, coordinates, radius, observed/expected cases, RR, LLR and p-value) from 2005 to 2015 are presented in the supplementary material.

Both high-salinity clusters (Fig. 9) and low-salinity clusters (Fig. 10) are more likely to occur in areas characterized by low flow exchanges and next to the main sluices. This result agrees with the standard deviation distribution presented in Fig. 6b, which shows a large salinity variability next to the mainland. Although the deep channels connecting the North Sea with the tidal basins experience the highest mean salinity values (Fig. 6a), clusters characterized by anomalous salinity values do not persist within this environment (the Eirlandse inlet is an exception, explained below).

We notice that long-lasting clusters (duration greater than 1 month, orange circles in Fig. 9) with an RR greater than 1.20 occur nearby the sluices. In comparison, singularities with a shorter duration (<1 month, blue circles in Fig. 9) are detected mainly in the northeastern part of the estuary, which is characterized by low water mass exchanges [Duran-Matute et al., 2014]. Moreover, we identified the presence of singularities next to the Eierlandse Inlet (e.g., cluster 3 in Fig. 9c). This result agrees with recent outcomes indicating that this inlet is less important in terms of exchange flows than Texel and Vlie inlets [Elias et al., 2012; Sassi et al., 2015].

Clusters marked by low salinity levels were detected across the estuary's entire shallow area and nearby the freshwater outlets (Fig. 10). Our results show that singularities next to Kornwerderzand develop mainly in spring and/or summer with a duration ranging from few days to 3 months. Groups of clusters were also identified in the northeastern part of the estuary and next to the Eierlandse Inlet (Fig. 10).

Finally, Fig. 11 depicts the clusters with RR>1.20 (Fig. 11a) and RR<0.8 (Fig. 11b) over the entire analysis period (2005-2015). We chose to represent only those with the highest likelihoods of clustering (LLR>1600). High-salinity clusters (Fig. 11a) occur within the northeastern part of the system in April-July, and within the central part of the system (centred in Kornwerderzand) in September-December. By contrast, low-salinity clusters (Fig. 11b) occur in the estuary's central part in summer (May-August) and the remaining area during winter (January-February).

3.3 Relation between clusters and forcing

Clusters characterized by anomalous salinity levels are not related only to extreme values in freshwater discharges. In fact, it is already known that wind can play a fundamental role in freshwater distribution and accumulation within estuarine systems [Geyer, 1997]. In the DWS, the median annual freshwater discharge has no significant year-to-year variations, but the sectorial wind energy (defined as in Gerkema and Duran-Matute, 2017, see supplementary material section 1) does (Fig. 12), with westerly and southwesterly winds having the largest energies. This suggests that freshwater dispersal might vary from year to year due to changes in the wind climate. In addition, the wind forcing exhibits substantial variability at monthly time scales. Specifically, winter and autumn are characterized by strong winds and large freshwater discharges (Fig. 13). These inter- and intra-annual variations in the external forces affect clusters' spatiotemporal variability, and anomalies are more likely to occur during those months characterized by the largest variability. However, it is necessary to explore how the external forces change at event scale to unmask the relationship between anomalies and external forces. In this subsection, we aim to relate the clusters identified by scan statistics with the variability in the wind and freshwater discharge at event scale.

We consider only two clusters, but the following approach can be applied to the other clusters as well. The first cluster that we consider took place between April 22nd and June 27th, 2008. It presents salinity values greater than expected (Fig. 11a) and the highest likelihood ratio (LLR = 5429). The mean daily freshwater discharge in 2008 is 283 m^3/s and 194 m^3/s for Den Oever and Kornwerderzand respectively, but these values are strongly reduced within the period in which the cluster is detected (i.e., 188 m^3/s for Den Oever and 40 m^3/s for Kornwerderzand). Figure 14 depicts the mean tracer concentration associated with the two sluices for the entire 2008 (Fig. 14 a, c) and the period in which the singularity is identified (Fig. 14 b, d). These maps highlight that: (i) the tracer released from Den Oever is confined in the southern part of the DWS, and (ii) the tracer coming from Kornwerdenzand presents very low concentrations between April 22nd and June 27th, 2008. Specifically, the fresh water released by Den Oever is trapped in the southern DWS during spring 2008 due to weak southwesterly winds (Fig. 15a, b). This means that the wind does not push the fresh water into the eastern part of the DWS. Therefore, the tidal basin connected to the North Sea by the Borndiep Inlet (inlet D in Fig. 1) experiences anomalous large salinity values in this period. This cluster is also compared with the anomaly occurring between September 9th and November 6th, 2009 (Fig. 11a), which exhibits similar characteristics (e.g., spatial location) during comparable forcing conditions (see supplementary material, Figs. S1 and S2).

The second cluster occurs between January 3rd and February 10th, 2005. It presents the highest likelihood ratio (LLR = 2378) among the clusters with salinity values lower than the expected ones (Fig. 11b). These weeks exhibit a mean freshwater discharge of 348 m^3/s and 248 m^3/s for Den Oever and Kornwerderzand, respectively. We notice that these values are larger than the mean freshwater discharge of the two sluices over the entire 2005 (i.e., 272 m^3/s for Den Oever and 187 m^3/s for Kornwerderzand). In addition, this period is characterized by strong southwesterly and westerly winds (Fig. 15c, d) that push the fresh water along the mainland and towards the eastern part of the DWS (Fig. 16b, d). The freshwater distribution during this period deviates largely from the mean conditions (Fig. 16a, c) as depicted in Fig. 16.

4 Discussion

The statistical detection of high-salinity clusters and low-salinity clusters within an intertidal basin was performed by using scan statistics, and the Dutch Wadden Sea as test case. The analyses are based on high-resolution numerical simulations spanning 11 years. Periodic components in the modeled salinity time series were identified by means of wavelet analysis and removed from the signal before applying scan statistics, since they can lead to the detection of spurious singularities (Figs. 7 and 8). This study showed that some anomalously low or high salinity values occur within specific estuarine sub-regions in a certain time frame. The recovery time to reach the pre-perturbation levels is larger for areas exhibiting low water exchange rates (Figs. 9 and 10). The occurrence of some anomalous salinity values can be easily linked to sporadic events characterized by winds and freshwater discharges which deviate largely from their normal/average conditions (Figs. 14 and 16). However, in other cases, the relationship between the occurrence of anomalous salinity values and the forcing might be more difficult to ascertain. This makes the use of a robust statistical method for anomalous cluster detection, such as spatial-temporal scan statistics, particularly valuable. Generally, clusters marked by salinity values greater than expected ones (Fig. 9) happen during autumn and winter, when storm frequency is higher. However, several singularities (e.g., clusters 2 and 4 in 2006, clusters 1 and 2 in 2010) are present in active biological periods of the year (i.e., spring and summer). Extreme episodes occurring in spring and/or summer can be particularly important because they influence organisms' biologically activities as well as their survival rates [e.g., Schumman et al., 2006].

The Dutch Wadden Sea has two primary agents in water movements: the highly predictable tides and the wind, which is episodic in nature and strongly variable from year to year (Fig. 12). These two drivers affect the freshwater fate and retention in the system. In particular, the wind pushes the fresh water into different areas, depending on the wind speed, direction and duration of each event, making salinity values anomalously low where fresh water accumulates or anomalously high where the amount of fresh water decreases. Thus, the wind forcing impacts on clusters' spatio-temporal variability. By contrast, changes in freshwater discharge contribute solely to clusters' temporal variability since they influence mainly the overall amount of fresh water in the system (Fig. 17).

Our results show that the wind creates an enormous variability in the freshwater fate and affects clusters' occurrence even in the easternmost part of the Dutch Wadden Sea (Fig. 16). In particular, scan statistics shows that clusters present marked spatio-temporal variations (Figs. 9 and 10), which reveals the event-driven nature of the system. In other words, winds play an important role in the DWS' dynamics and storms can significantly alter the response of the system with respect to its long-term typical state (i.e., median and/or mean conditions). Since episodic events have a paramount effect on the long-term mean state, the DWS cannot be studied as statistically steady (i.e., as if tides were the main agent in water movements). Thus, long-term numerical modeling simulations and advanced statistical methods are needed to identify the typical state of the system and the cumulative impact of anomalous events on the estuary's long-term characteristics. This finding has implications not only for the freshwater transport and mixing processes, but also for the exchange of sediments, larvae and nutrients between the Dutch Wadden Sea and the adjacent North Sea.

Numerous statistical approaches devoted to understanding salinity patterns in estuarine and coastal areas are present in the literature [e.g., Guerra-Chanis et al., 2019; Eslami et al., 2019]. However, a methodology which allows the statistical detection of clusters in the salinity field is still missing. Identifying anomalous behaviors in a meaningful statistical way is fundamental to relate the occurrence of extreme events to global climate change [e.g., van Oldenborgh et al., 2019], to understand event-driven systems' dynamics [e.g., Duran-Matute et al., 2014], and to detect data-clustering which may be indicative of potential emerging environmental hazards [e.g., van Oldenborgh et al., 2015]. For instance, cluster detection may help reveal the environmental factors that drive anomalies within a system and evaluate how the number of these anomalies evolves in time. In this study, we propose scan statistics as a tool to identify (statistically) singularities in the salinity field, link the anomalies to the forcing conditions, and unravel intricate estuarine processes underlying clusters' occurrence. In addition, this data mining technique is a valuable method for fast explorations of large amount of data, since it enables the understanding of complex dynamics (e.g., relationship between clusters and external forces) within coastal systems by focusing on specific spatio-temporal windows.

The major advantage of scan statistics with respect, for instance, to the POT-method is that the selection of a threshold is not needed since anomalous values in a specific location are detected by comparing the actual salinity values with the expected ones. In addition, scan statistics consider in cluster detection that the salinity field presents a certain spatial variability within the estuary, whereas the POT-method uses the same censoring critical value for the entire system [e.g., Carniello et al., 2016]. Another advantage of the proposed approach is identifying pixels that experience anomalous values with respect to the expected one within the same temporal window and assess their statistical significance.

The main shortcoming of scan statistics is related to the use of cylindrical scanning windows. As the cluster's shape becomes more irregular (e.g., next to locations where the morphology is more complex), the efficiency of this methodology decreases. Attempts to employ irregularly-shaped search areas are present in literature [Patil and Taillie, 2004; Tango and Takaashi, 2005], but these methods are very computationally intensive [Robertson and Nelson, 2010]. Another limitation is that scan statistics can identify clusters that are singularities only in an average sense. The algorithm can consider as a single singularity two adjacent clusters by including all the pixels between them, albeit these locations do not experience anomalous salinity values.

5. Conclusions

Using high-resolution numerical simulations and scan statistics, we identify regions within the Dutch Wadden Sea that experience anomalous salinity values (i.e., strong deviations from the expected in a statistical sense) within a certain temporal window over 2005-2015. This study shows that a mathematical method widely tested in epidemiology can be applied for performing spatio-temporal analysis in back-barrier basins. The proposed methodology is also suitable for other scalar fields (e.g., temperature) and other coastal systems.

The Dutch Wadden Sea behaves like an event-driven system (i.e., the highly variable wind forcing is an important agent in water movements) as such it exhibits a substantial temporal and spatial variability in clusters' occurrence. In particular, these clusters do not occur in the entire system but within subregions. The exact location of these areas depends on how the external forcing (i.e., winds, freshwater discharge) is anomalous compared to the average conditions. Specifically, we show that high-salinity clusters and low-salinity clusters are related to sporadic episodes characterized by extreme southwesterly winds and/or anomalous amounts of fresh water discharged by the two main sluices (Figs. 14 and 16). In addition, scan statistics suggest that estuarine regions characterized by low water exchanges present long-lasting clusters. These areas are less dynamic, and therefore they require more time to recover their pre-perturbation conditions after the occurrence of a particular extreme event (Figs. 9, 10 and 11). Finding clusters characterized by anomalous behaviors in space and time is useful for understanding coastal systems' dynamics, and for analyzing the occurrence of extreme events from a global climate change prospective. This research underlines that the dynamics of event-driven systems cannot be studied as a steady predictable system (i.e., governed by the repetitive tides where the wind forcing is just a perturbating factor), but we must take into account the episodic character in the hydrodynamics to properly study the freshwater dispersal, mixing and transport processes in a meaningful way.

Acknowledgements

This study was supported by the NWO/ENW project: ‘The Dutch Wadden Sea as an event-driven system: long-term consequences for exchange (LOCO-EX)’. The work was supported by the North-German Supercomputing Alliance (HLRN). Data are available online (https://zenodo.org/search?page=1&size=20&q=carmine%20donatelli).

We thank Dr. David Ralston, the two anonymous Reviewers and Dr. Alessandro Di Bucchianico (Department of Mathematics and Computer Science, Eindhoven University of Technology, the NL) for their useful comments and suggestions.

Figure captions

Figure 1. Study area. Bathymetry of the Dutch Wadden Sea. Points P1, P2, and P3 represent three typical environments in the estuary: deep channels (P1), subtidal area next to the sluices (P2) and tidal flats (P3).

Figure 2. This diagram summarizes the main steps, followed by scan statistics to detect potential spatio-temporal clusters.

Figure 3. Scan statistics. The cylinders represent two scanning windows centred in two different points of the domain. These cylinders can scan different time intervals by varying their heights and different geographical locations by changing their radius.

Figure 4. Application of scan statistics on a simple domain: (a) population and (b, c) number of cases. Colored circles indicate the scanning windows.

Figure 5. Application of scan statistics on a simple domain: (a) population and (b) number of cases. The colored circle indicates the detected cluster.

Figure 6. These maps represent the (a) mean of the daily average salinity and (b) standard deviation of the daily average salinity in the DWS. The analysis period is 2005-2015.

Figure 7. Wavelet analysis of the hourly salinity time series: (a, b, c) wavelet power spectrum for point P1, P2, and P3. Wavelet power varies from low power (blue) to high power (red). The regions of greater than 95% confidence are shown with thick black contours. Cross-hatched regions indicate the "cone of influence," where edge effects become important.

Figure 8. Wavelet analysis of the daily salinity time series: (a, b, c) wavelet power spectrum for point P1, P2, and P3. Wavelet power varies from low power (blue) to high power (red). The regions of greater than 95% confidence are shown with thick black contours. Cross-hatched regions indicate the “cone of influence,” where edge effects become important.

Figure 9. Application of scan statistics on the DWS over the period 2005-2015. Detected clusters with salinity values higher than the expected ones in (a) 2005, (b) 2006, (c) 2007, and (d) 2008. Blues circles represent clusters shorter than 1 month. Orange circles represent clusters longer than 1 month. The circles' contour's color indicates the season: red for spring and summer, purple for autumn and winter.

Figure 10. Application of scan statistics on the DWS over the period 2005-2015. Detected clusters with salinity values lower than the expected ones in (a) 2005, (b) 2006, (c) 2007, and (d) 2008. Blues circles represent clusters shorter than 1 month. Orange circles represent clusters longer than 1 month. The circles' contour's color indicates the season: red for spring and summer, purple for autumn and winter.

Figure 11. Application of scan statistics on the DWS over 2005-2015. Detected clusters with salinity values (a) higher (b) lower than the expected ones with a LLR>1600 over the entire period of analysis.

Figure 12. Sectorial annual mean energy for all individual years. Wind energy is defined according to Gerkema and Duran-Matute [2017] (see supplementary material).

Figure 13. Monthly variability in the (a) wind climate and in the (b) total amount of freshwater discharge.

Figure 14. The maps represent the mean tracer concentration distribution within the DWS for: (a, b) Den Oever and (c, d) Kornwerderzand in the (a,c) 2008 and (b, d) in the period in which the cluster occurs (April 22nd-June 27th, 2008).

Figure 15. Sectorial mean wind energy: (a) for 2008, (b) for the period: April 22nd-June 27th, 2008, (c) for 2005 and (d) for the period: January 3th-February 10th, 2005. Wind energy is defined according to Gerkema and Duran-Matute [2017] (see supplementary material).

Figure 16. The maps represent the mean tracer concentration distribution within the DWS for: (a, b) Den Oever and (c, d) Kornwerderzand in the (a,c) 2005 and (b, d) in the period in which the cluster occurs (January 3th-February 10th, 2005).

Figure 17. This diagram explains how the freshwater discharge and the wind forcing influence the spatio-temporal variability in clusters' occurrence.

References

Blauw, A.N., Benincà, E., Laane, R.W.P.M., Greenwood, N., Huisman, J., (2012). Dancing with the tides: fluctuations of coastal phytoplankton orchestrated by different oscillatory modes of the tidal cycle. PLoS ONE 7 (11), e49319. http://dx.doi.org/10.1371/journal.pone.0049319.

Blauw, A.N., Benincà, E., Laane, R.W.P.M., Greenwood, N., Huisman, J. (2018). Predictability and environmental drivers of chlorophyll fluctuations vary across different time scales and regions of the North Sea. Prog. Oceanogr., 161, 1–18

Bruggeman, J., and K. Bolding (2014), A general framework for aquatic biogeochemical models, Environ. Model. Softw., 61, 249–265, doi: 10.1016/j.envsoft.2014.04.002

Burchard, H., and Bolding, K., (2002). GETM: A general estuarine transport model, Scientific documentation, Tech. Rep. EUR 20253 EN, Eur. Comm., Ispra, Italy

Carniello, L., D'Alpaos, A., Botter, G., Rinaldo, A., (2016). Statistical characterization of spatio temporal sediment dynamics in the Venice lagoon
J. Geophys. Res. Earth Surf., 121, pp. 1049-1064, 10.1002/2015JF003793

Cloern, J.E., (1984). Temporal dynamics and ecological significance of salinity stratification in an estuary (South San Francisco Bay, USA), Oceanol. Acta, 7, 137–141.

Cramér, H., and Leadbetter, M. R., (1967). Stationary and Related Stochastic Processes, 348 pp., John Wiley, New York.

Doney, S.C., Ruckelshaus, M., Duffy, J.E., Barry, J.P., Chan, F., English, C.A., Galindo, H.M., Grebmeier, J.M., Hollowed, A.B., Knowlton, N., Polovina, J., Rabalais, N.N., Sydeman, W.J., Talley, L.D., (2012). Climate change impacts on marine ecosystems Annu. Rev. Mar. Sci., 4, pp. 11-37

Duran-Matute, M., and Gerkema, T., (2015), Calculating residual flows through a multiple-inlet system: The conundrum of the tidal period, Ocean Dyn., 65(11), 1461–1475, doi:10.1007/s10236-015-0875-1.

Duran-Matute, M., Gerkema, T., de Boer, G.J., Nauw, J.J., and Gräwe, U., (2014). Residual circulation and freshwater transport in the Dutch Wadden Sea: A numerical modelling study, Ocean Sci., 10(4), 611–632, doi:10.5194/os-10-611-2014.

Duran-Matute, M., Gerkema, T., and Sassi, M.G., (2016), Quantifying the residual volume transport through a multiple-inlet system in response to wind forcing: The case of the western Dutch Wadden Sea, J. Geophys. Res. Oceans, 121, 8888–8903, doi:10.1002/ 2016JC011807.

D'Alpaos, A., Carniello, L., and Rinaldo, A., (2013). Statistical mechanics of wind wave-induced erosion in shallow tidal basins: Inferences from the Venice lagoon, Geophys. Res. Lett., 40, 3402–3407, doi:10.1002/grl.50666.

Elias, E., Van der Spek, A., Wang, Z., (2012). Morphodynamic development and sediment budget of the Dutch Wadden Sea over the last century. Neth J Geosci 91(3):293–310

Fraker, S.E., Woodall, W.H., and Burkom, H.S., (2008) “A note on the Poisson likelihood ratio test statistic for Kulldorff’s scan methods,” Communications in Statistics—Theory and Methods, vol. 37, no. 7, pp. 998–1001.

Gerkema, T., Nauw, J.J., and Van der Hout, C.M., (2014). Measurements on the transport of suspended particulate matter in the Vlie Inlet, Neth. J. Geosci., 93(3), 95–105, doi:10.1017/njg.2014.7.

Gerkema, T., and Duran- Matute, M., (2017). Interannual variability of mean sea level and its sensitivity to wind climate in an inter- tidal basin. Earth System Dynamics, 8(4), 1223–1235. https://doi.org/10.5194/esd- 8- 1223- 2017

Gerkema,T., (2019). An introduction to tides. Cambridge: Cambridge University Press. https://doi-org.proxy-ub.rug.nl/10.1017/9781316998793

Geyer, W.R., (1997). Influence of wind on dynamics and flushing of shallow estuaries. Estuarine, Coastal and Shelf Science 44:713- 722.

Ghezzo, M., Sarretta, A., Sigovini, M., Guerzoni, S., Tagliapietra, D., and Umgiesser, G., (2011). Modeling the inter-annual variability of salinity in the lagoon of Venice in relation to the water framework directive typologies. Ocean & Coastal Management. doi:10.1016/j.ocecoaman.2011.06.007.

Grinsted, A., Moore, J.C., Jevrejeva, S., (2004). Application of the cross wavelet transform and wavelet coherence to geophysical time series. Nonlinear Processes Geophys. 11, 561–566.

Gräwe, U., Floser, G., Gerkema, T., Duran-Matute, M., Badewien, T.H., Schulz, E., and Burchard, H., (2016). A numerical model for the entire Wadden Sea: Skill assessment and analysis of hydrodynamics, J. Geophys. Res. Oceans, 121, 5231–5251, doi:10.1002/2016JC011655.

Kulldorff, M., (1997). A spatial scan statistic. Comm. Statist.: Theory Methods 26, 1481–1496.

Kulldorff, M., (1999a). An isotonic spatial scan statistic for geographical disease surveillance. J. Nat. Inst. Public Health 48 (2), 94–101.

Kulldorff, M., (1999b). Spatial scan statistics: models, calculations and applications. In: Glaz, Balakrishnan (Eds.), Scan Statistics and Applications. Birkhauser, Boston, pp. 303–322.

Kulldorff, M., (2001). Prospective time periodic geographical disease surveillance using a scan statistic, J R Stat Soc Ser A, 164 (1), pp. 61-72

Kulldorff, M., Heffernan, R., Hartman, J., Assuncao, R.M., Mostashari, F., (2005). A space–time permutation scan statistic for the early detection of disease outbreaks. PLoS Med, 2, pp. 216-224

Leadbetter, M. R., (1991): On a basis for “peaks over threshold” modeling. Stat. Prob. Lett., 12, 357–362.

Liu, Y., Liang, X.S., and Weisberg, R.H., (2007). Rectification of the bias in the wavelet power spectrum. J. Atmos. Oceanic Technol. 24, 2093–2102. (doi:10.1175/2007JTECHO511.1).

Marcos, R.D.L.F. and Marcos, C.D.L.F., (2008). From star complexes to the field: open cluster families, Astrophysical Journal, 672, 342–351.

Matsoukis, C., Amoudry, L.O., Bricheno, L. et al. Investigation of Spatial and Temporal Salinity Distribution in a River Dominated Delta through Idealized Numerical Modelling. Estuaries and Coasts (2021). https://doi-org.proxy-ub.rug.nl/10.1007/s12237-021-00898-2

Meier, H.E.M. (2007): Modeling the pathways and ages of inflowing salt and freshwater in the Baltic Sea, Estuar. Coast. Shelf S., 74, 610–627.

Naus, J.I., (1965). The distribution of the size of the maximum cluster of points on a line J Am Stat Assoc, 60, pp. 532-538

Patil, G.P., Taillie, C., (2004). Upper level set scan statistics for detecting arbitrarily shaped hotspots, Environ Ecol Stat, 11, pp. 189-197

Ridal, M., Olsson, E., Unden, P., Zimmermann, K., and Ohlsson, A., (2017), Deliverable D2.7 : HARMONIE reanalysis report of results and dataset. http://www.uerra.eu/component/dpattachments/?task=attachment.download&id=296

Rupp, G.S., Parsons, G.J., (2004). Effects of salinity and temperature on the survival and byssal attachment of the lion's paw scallop Nodipecten nodosus at its southern distribution limit. J. Exp. Mar. Biol. Ecol., 309, pp. 173-198

Sassi, M.G., Gerkema, T., Duran-Matute, M., and Nauw, J.J., (2015), Residual water transport in the Marsdiep tidal inlet inferred from observations and a numerical model, J. Mar. Res., 74(1), 21–42, doi:10.1357/002224016818377586.

Schumann, R., Baudler, H., Glass, Ä., Dümcke, K., Karsten, U., (2006). Long-term observations on salinity dynamics in a tideless shallow coastal lagoon of the southern Baltic Sea coast and their biological relevance, J. Mar. Syst., 60, pp. 330-344

Telesh, I.V., Khlebovich, V.V., (2010). Principal processes within the estuarine salinity gradient: a review. Mar. Pollut. Bull., 61, pp. 149-155

Takahashi, K., Kulldorff, M., Tango, T., Yih, K., (2008). A flexibly shaped space–time scan statistic for disease outbreak detection and monitoring Int J Health Geogr, 7, p. 14

Teixeira, H., Salas, F., Borja, A., Neto, J.M., Marques, J.C., (2008). A benthic perspective in assessing the ecological status of estuaries: the case of the Mondego estuary (Portugal). Ecol. Indic., 8, pp. 404-416

Torrence, C., Compo, G.P., (1998). A practical guide to wavelet analysis. Bull. Am. Meteorol. Soc. 79, 61–78. http://dx.doi.org/10.1175/1520-0477(1998) 079, 0061:APGTWA>2.0.CO;2.

Tuia, D., Lasaponara, R., Telesca, L., Kanevski, M., Emergence of spatio-temporal patterns in forest-fire sequences, Physica A, 387 (2008), p. 3271

van de Kreeke, J., Brouwer, R.L., (2017). Tidal Inlets: Hydrodynamics and Morphodynamics. Cambridge University Press

van Oldenborgh, G.J., Haarsma, R., de Vries, H., and Allen, M.R., (2015). Cold extremes in North America vs. mild weather in Europe: The winter 2013–14 in the context of a warming world. Bull. Amer. Meteor. Soc., 96, 707–714, doi:10.1175/BAMS-D-14-00036.1

van Oldenborgh, G.J. et al., (2019). Pathways and pitfalls in extreme event attribution EMS Annual Meeting Abstracts 15 EMS2018-476

Verdelhos, T., Marques, J.C., Anastácio, P., (2015). The impact of estuarine salinity changes on the bivalves Scrobicularia plana and Cerastoderma edule, illustrated by behavioral and mortality responses on a laboratory assay. Ecol. Indic., 52, pp. 96-104

Wheatly, M.C., (1988). Interated responses to salinity function. Am. Zool. 28:65-77.

Whitfield, A. K., Elliott, M., Basset, A., Blaber, S. J. M., and West, R. J., (2012). Paradigms in estuarine ecology—the Remane diagram with a suggested revised model for estuaries: A review. Estuarine, Coastal & Shelf Science 97: 78–90.

Zhang, W.G., Wilkin, J.L., and Chant, R.J., (2009): Modeling the pathways and mean dynamics of river plume dispersal in the New York Bight, J. Phys. Oceanogr., 39, 1167–1183.

Figure

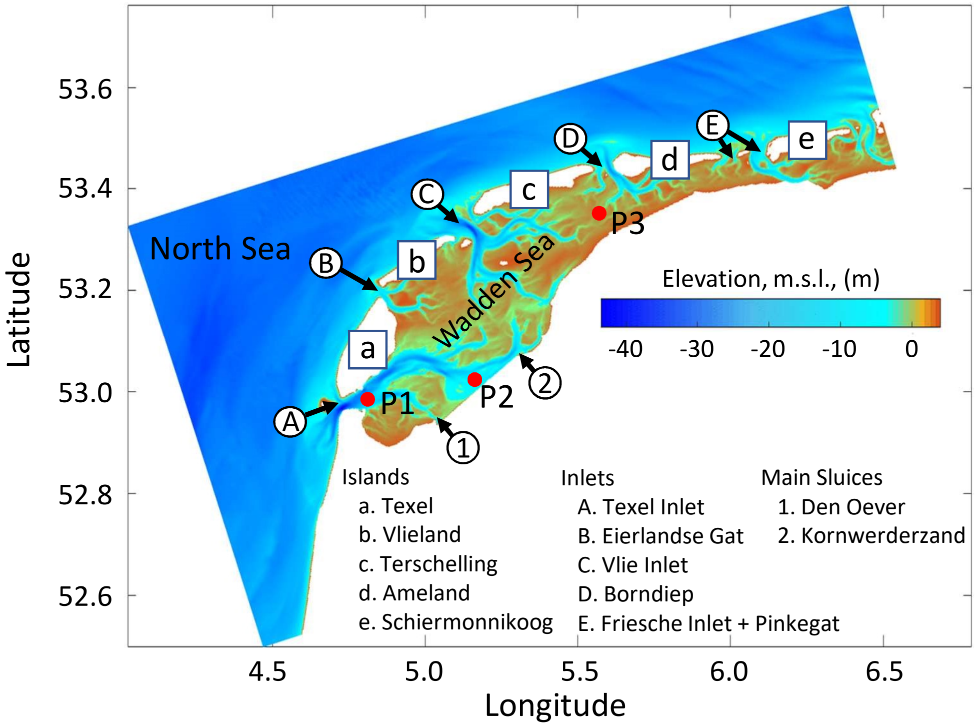


Figure 1

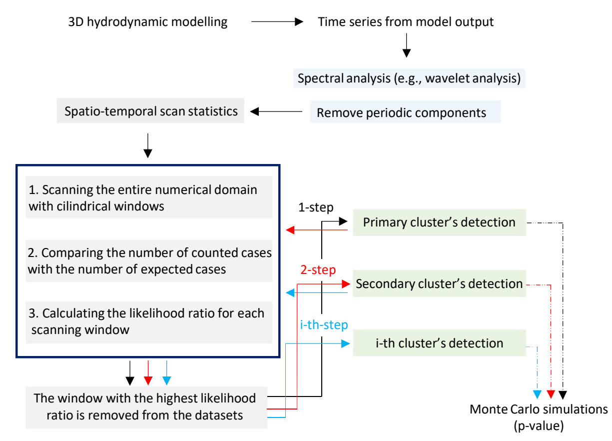

3D hydrodynamic modelling
Time series from model output
Spectral analysis (e.g., wavelet analysis)
Spatio-temporal scan statistics
Remove periodic components
1. Scanning the entire numerical domain with cilindrical windows
2. Comparing the number of counted cases with the number of expected cases
3. Calculating the likelihood ratio for each scanning window
1-step
Primary cluster’s detection
2-step
Secondary cluster’s detection
i-th-step
i-th cluster’s detection
The window with the highest likelihood ratio is removed from the datasets
Monte Carlo simulations (p-value)

Figure 2

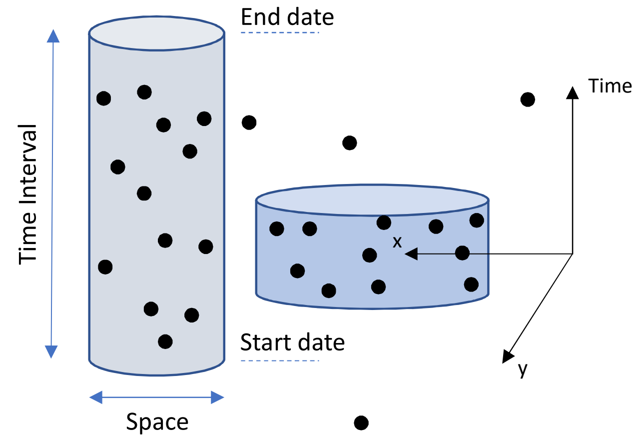


Figure 3

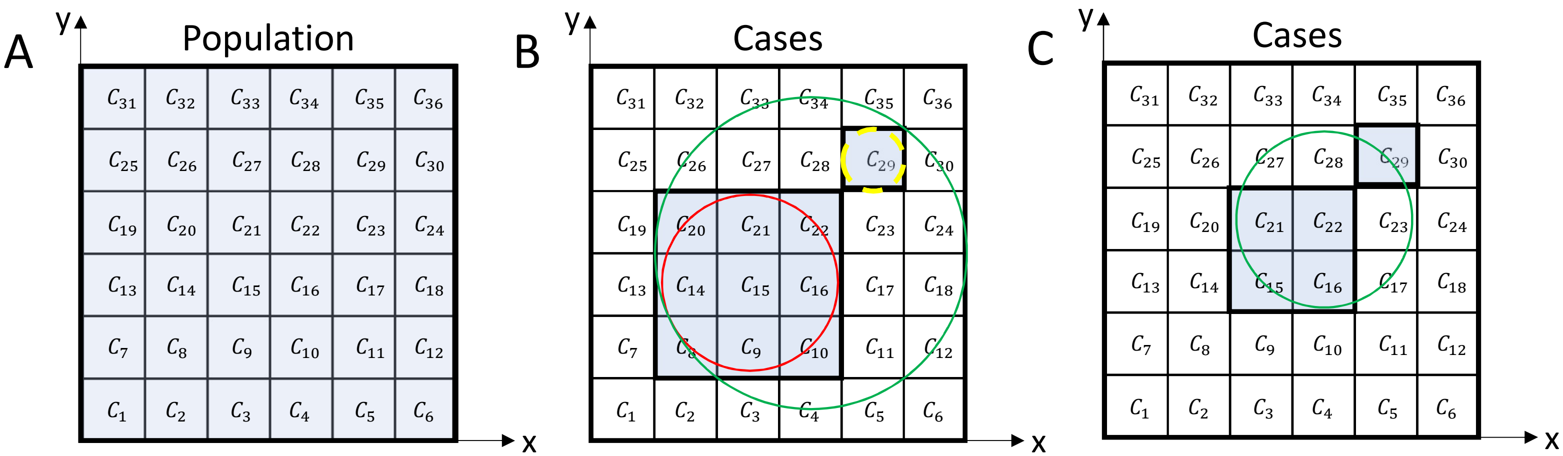
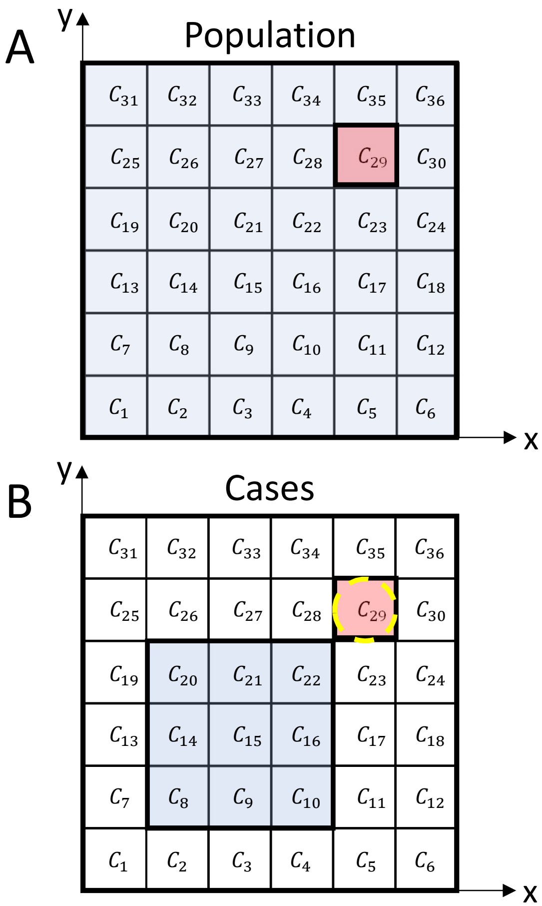


Figure 4



A
Population
y
x
C_{31} C_{32} C_{33} C_{34} C_{35} C_{36}
C_{25} C_{26} C_{27} C_{28} C_{29} C_{30}
C_{19} C_{20} C_{21} C_{22} C_{23} C_{24}
C_{13} C_{14} C_{15} C_{16} C_{17} C_{18}
C_{7} C_{8} C_{9} C_{10} C_{11} C_{12}
C_{1} C_{2} C_{3} C_{4} C_{5} C_{6}
B
Cases
y
x
C_{31} C_{32} C_{33} C_{34} C_{35} C_{36}
C_{25} C_{26} C_{27} C_{28} C_{29} C_{30}
C_{19} C_{20} C_{21} C_{22} C_{23} C_{24}
C_{13} C_{14} C_{15} C_{16} C_{17} C_{18}
C_{7} C_{8} C_{9} C_{10} C_{11} C_{12}
C_{1} C_{2} C_{3} C_{4} C_{5} C_{6}

Figure 5

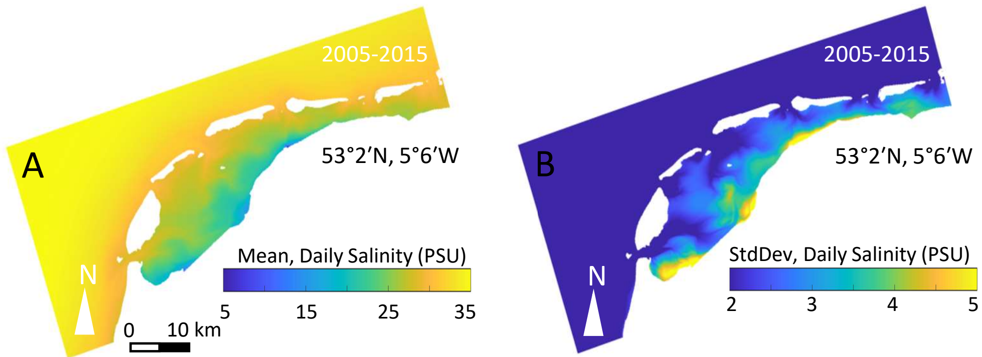

A
2005-2015
53°2'N, 5°6'W
Mean, Daily Salinity (PSU)
5
15
25
35
N
0
10 km
B
2005-2015
53°2'N, 5°6'W
StdDev, Daily Salinity (PSU)
2
3
4
5
N

Figure 6

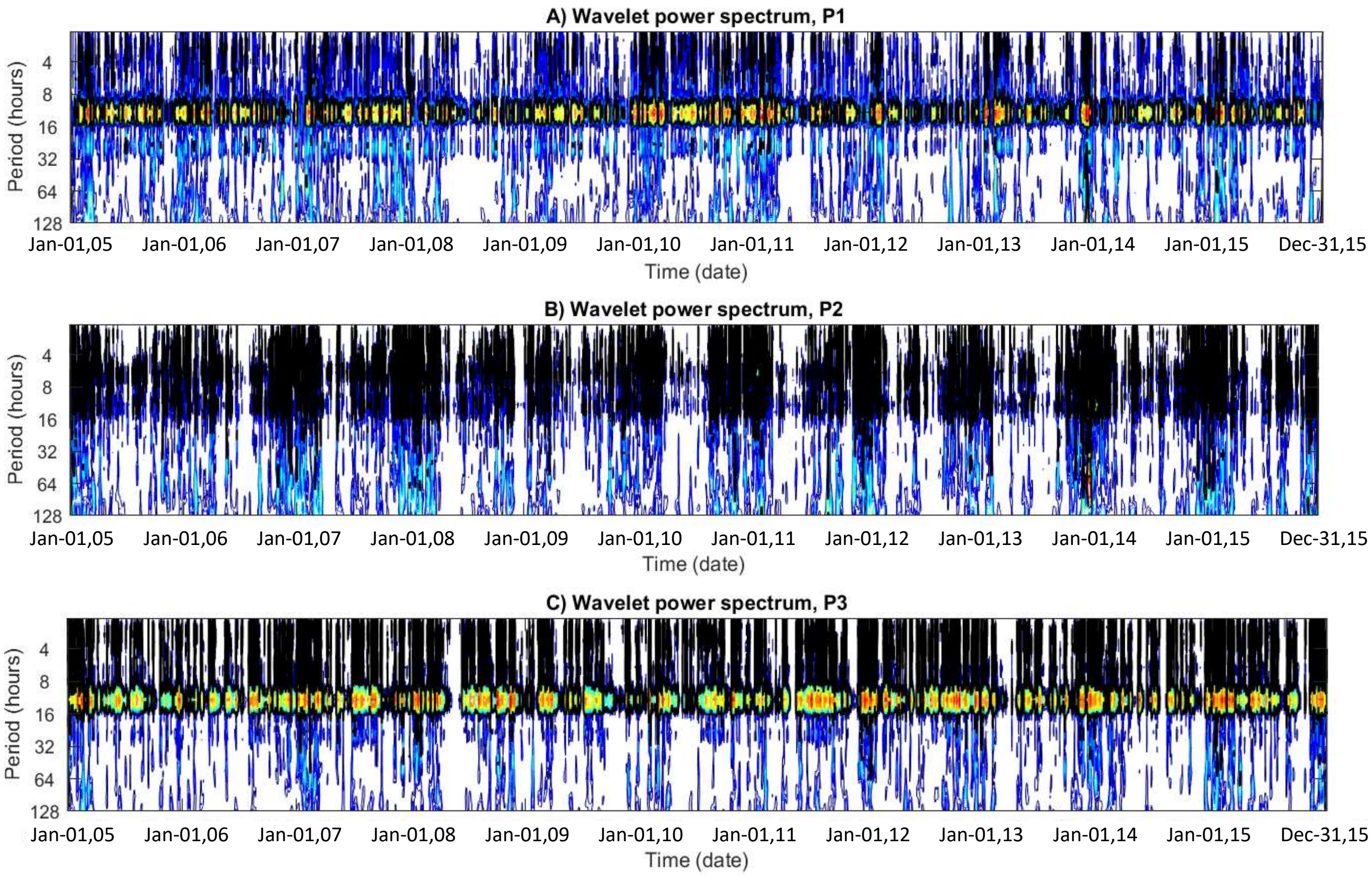

A) Wavelet power spectrum, P1
B) Wavelet power spectrum, P2
C) Wavelet power spectrum, P3
Period (hours)
4
8
16
32
64
128
Jan-01,05
Jan-01,06
Jan-01,07
Jan-01,08
Jan-01,09
Jan-01,10
Jan-01,11
Jan-01,12
Jan-01,13
Jan-01,14
Jan-01,15
Dec-31,15
Time (date)

Figure 7

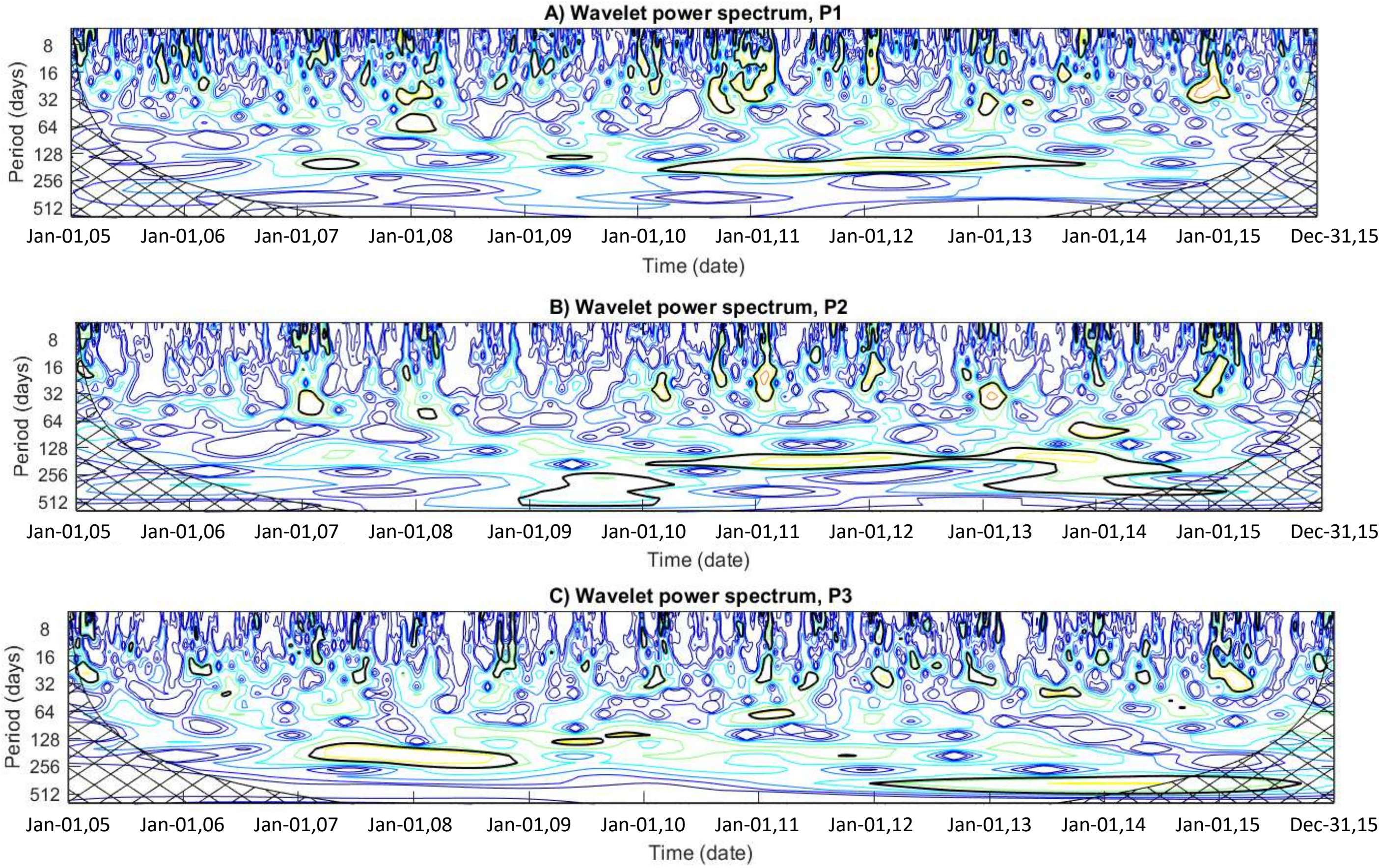

A) Wavelet power spectrum, P1
B) Wavelet power spectrum, P2
C) Wavelet power spectrum, P3
Period (days)
8
16
32
64
128
256
512
Jan-01,05
Jan-01,06
Jan-01,07
Jan-01,08
Jan-01,09
Jan-01,10
Jan-01,11
Jan-01,12
Jan-01,13
Jan-01,14
Jan-01,15
Dec-31,15
Time (date)

Figure 8

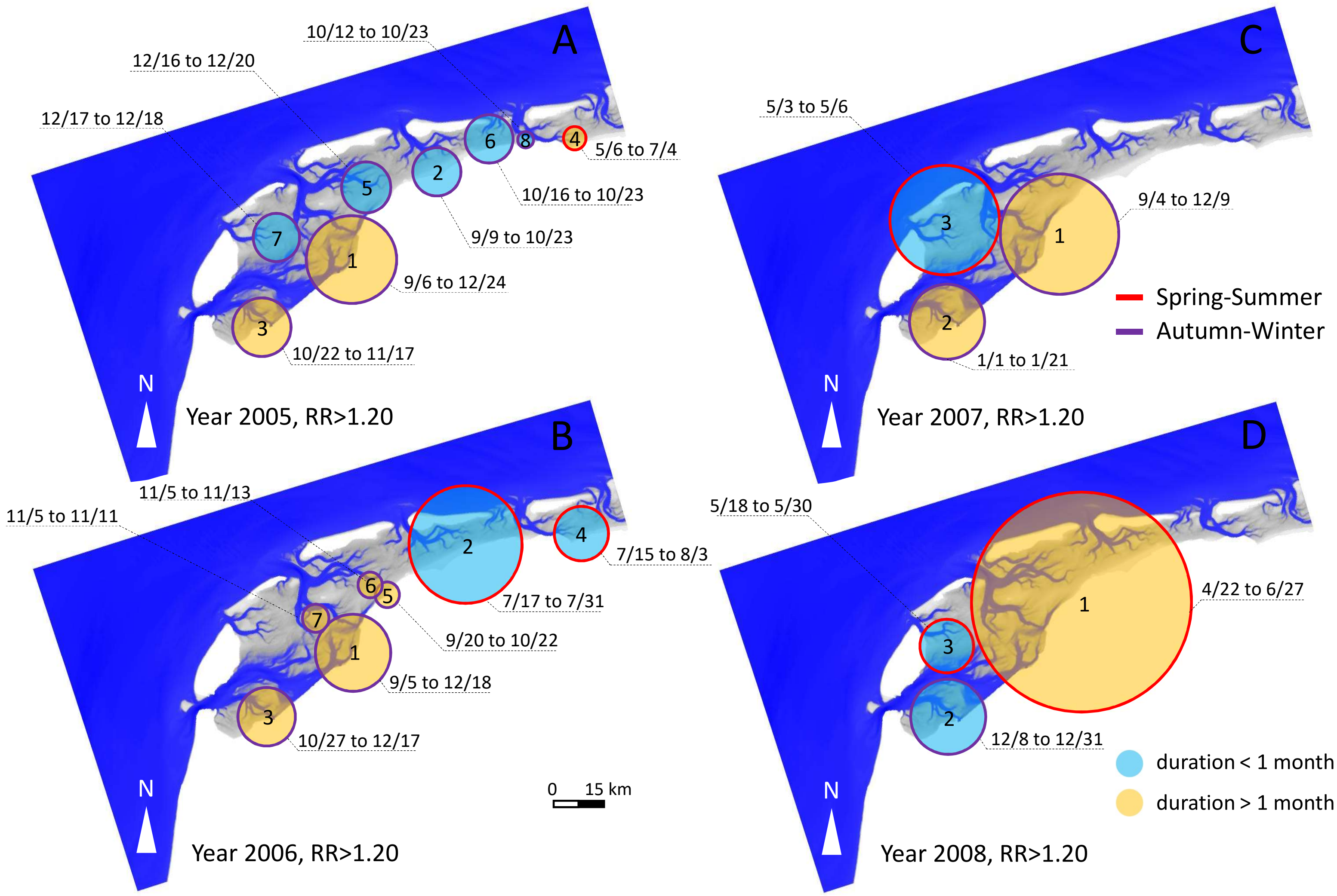

A
Year 2005, RR>1.20
12/17 to 12/18
12/16 to 12/20
10/12 to 10/23
5/6 to 7/4
10/16 to 10/23
9/9 to 10/23
9/6 to 12/24
10/22 to 11/17
B
Year 2006, RR>1.20
11/5 to 11/11
11/5 to 11/13
7/15 to 8/3
7/17 to 7/31
9/20 to 10/22
9/5 to 12/18
10/27 to 12/17
0 15 km
C
Year 2007, RR>1.20
5/3 to 5/6
9/4 to 12/9
1/1 to 1/21
Spring-Summer
Autumn-Winter
D
Year 2008, RR>1.20
5/18 to 5/30
4/22 to 6/27
12/8 to 12/31
duration < 1 month
duration > 1 month
N

Figure 9

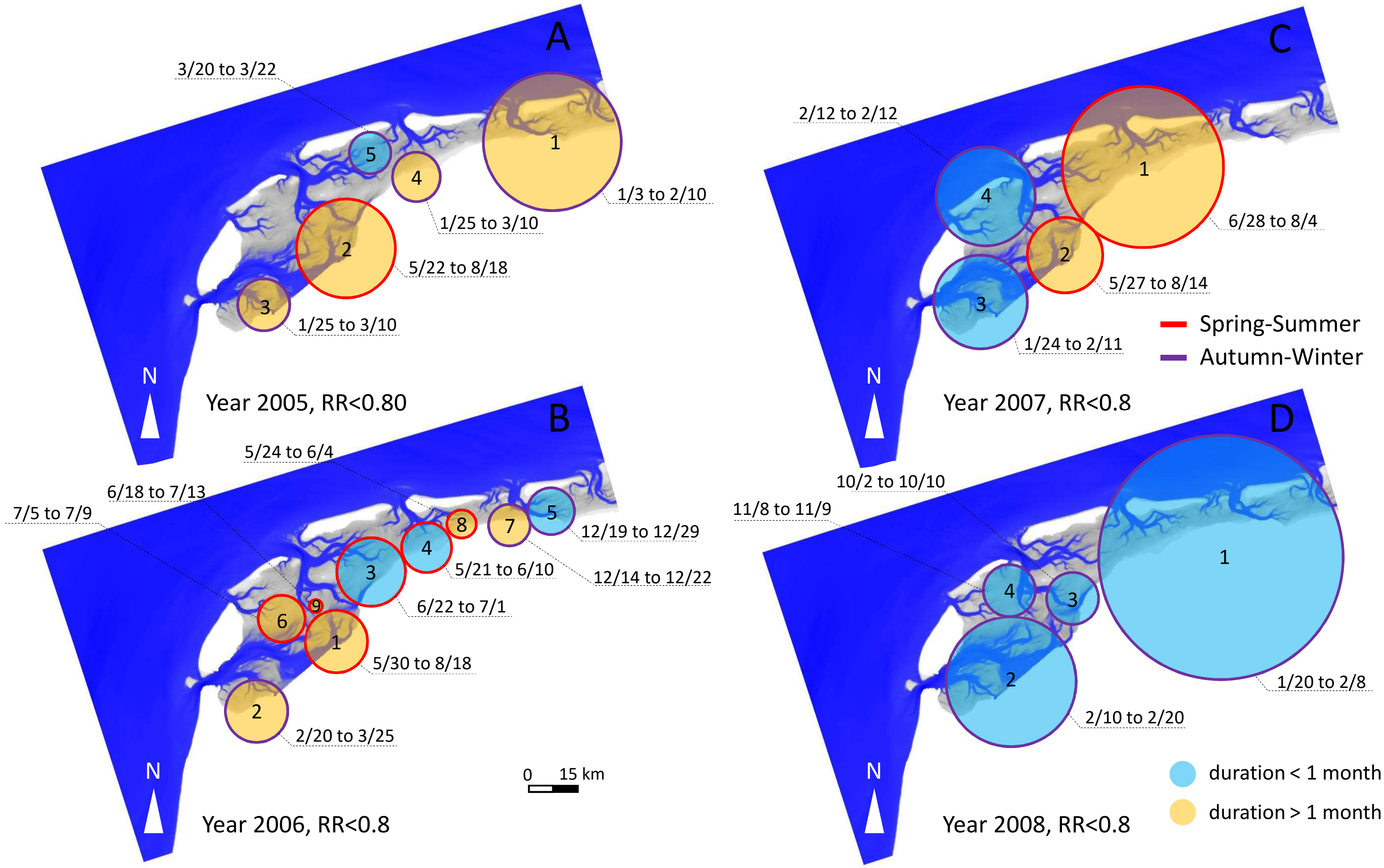

A
3/20 to 3/22
5
4
1
2
3
1/3 to 2/10
1/25 to 3/10
5/22 to 8/18
1/25 to 3/10
N
Year 2005, RR<0.80
C
2/12 to 2/12
4
1
2
3
6/28 to 8/4
5/27 to 8/14
1/24 to 2/11
Spring-Summer
Autumn-Winter
N
Year 2007, RR<0.8
B
5/24 to 6/4
6/18 to 7/13
7/5 to 7/9
5
8
7
4
3
9
6
1
2
12/19 to 12/29
5/21 to 6/10
12/14 to 12/22
6/22 to 7/1
5/30 to 8/18
2/20 to 3/25
0 15 km
N
Year 2006, RR<0.8
D
10/2 to 10/10
11/8 to 11/9
4
3
1
2
1/20 to 2/8
2/10 to 2/20
duration < 1 month
duration > 1 month
N
Year 2008, RR<0.8

Figure 10

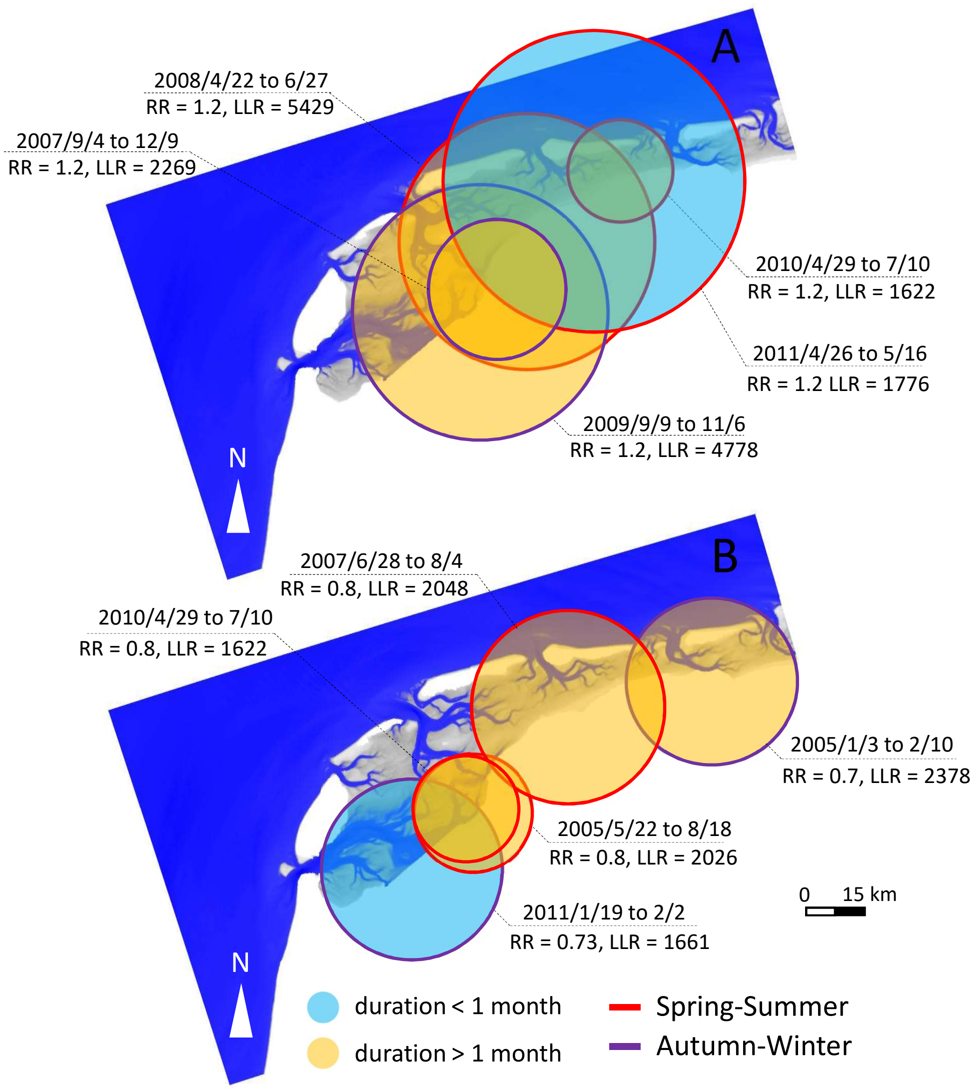
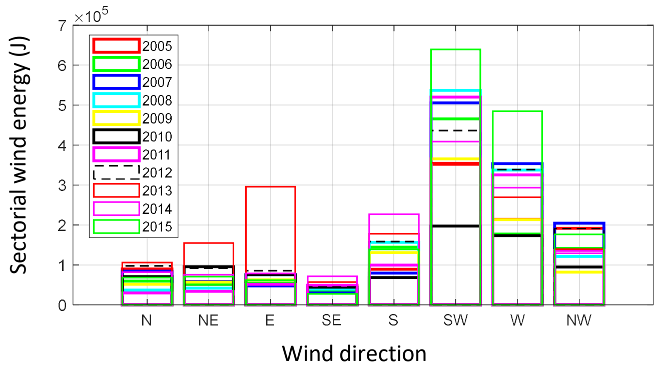
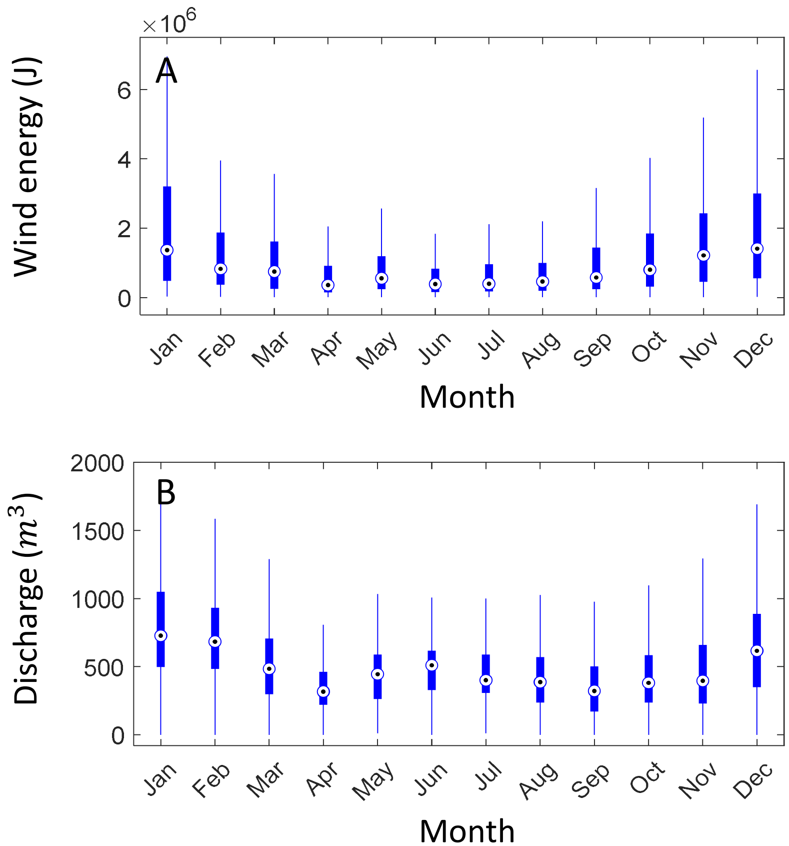

A
2008/4/22 to 6/27
RR = 1.2, LLR = 5429
2007/9/4 to 12/9
RR = 1.2, LLR = 2269
2010/4/29 to 7/10
RR = 1.2, LLR = 1622
2011/4/26 to 5/16
RR = 1.2 LLR = 1776
2009/9/9 to 11/6
RR = 1.2, LLR = 4778
N
B
2007/6/28 to 8/4
RR = 0.8, LLR = 2048
2010/4/29 to 7/10
RR = 0.8, LLR = 1622
2005/1/3 to 2/10
RR = 0.7, LLR = 2378
2005/5/22 to 8/18
RR = 0.8, LLR = 2026
2011/1/19 to 2/2
RR = 0.73, LLR = 1661
0 15 km
N
duration < 1 month
duration > 1 month
Spring-Summer
Autumn-Winter

Figure 11



×10^5
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
Sectorial wind energy (J)
Wind direction
N
NE
E
SE
S
SW
W
NW

Figure 12



×10^6
A
Wind energy (J)
6
4
2
0
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month
B
Discharge (m^3)
2000
1500
1000
500
0
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month

Figure 13

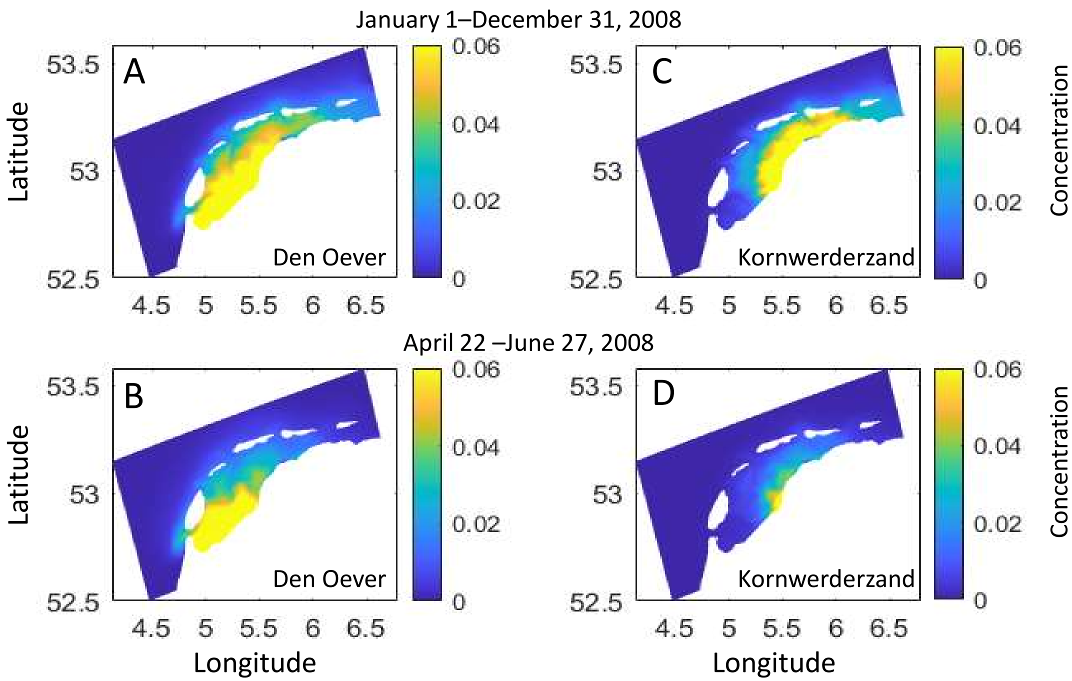

January 1–December 31, 2008
April 22 –June 27, 2008
A
B
C
D
Den Oever
Den Oever
Kornwerderzand
Kornwerderzand
Latitude
Latitude
Longitude
Longitude
Concentration
Concentration
53.5
53
52.5
4.5
5
5.5
6
6.5
0.06
0.04
0.02
0

Figure 14

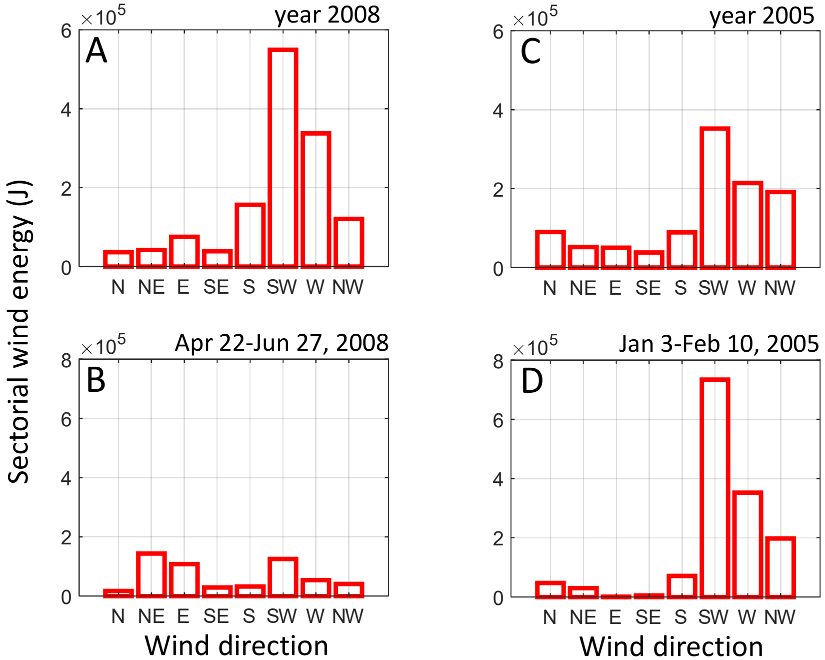

Sectorial wind energy (J)
×10^5
year 2008
A
6
4
2
0
N NE E SE S SW W NW
×10^5
year 2005
C
6
4
2
0
N NE E SE S SW W NW
×10^5
Apr 22-Jun 27, 2008
B
8
6
4
2
0
N NE E SE S SW W NW
Wind direction
×10^5
Jan 3-Feb 10, 2005
D
8
6
4
2
0
N NE E SE S SW W NW
Wind direction

Figure 15

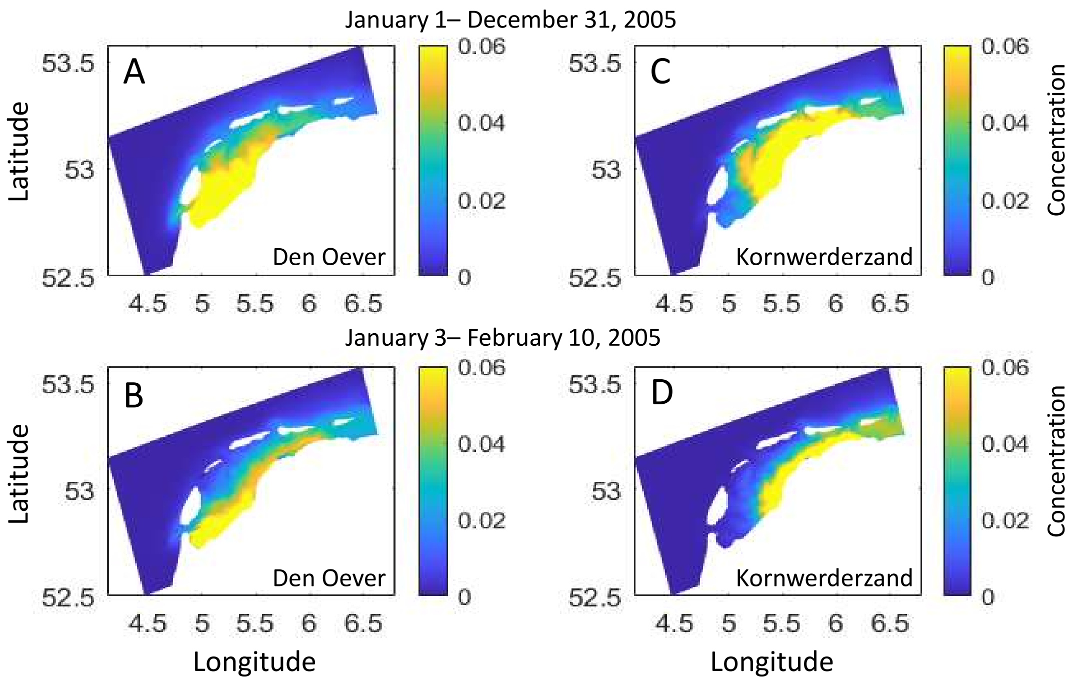

January 1– December 31, 2005
A
Den Oever
C
Kornwerderzand
January 3– February 10, 2005
B
Den Oever
D
Kornwerderzand
Latitude
Longitude
Concentration
53.5
53
52.5
4.5
5
5.5
6
6.5
0.06
0.04
0.02
0

Figure 16

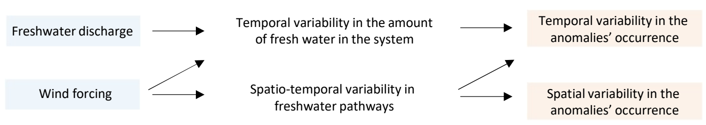

Freshwater discharge
Wind forcing
Temporal variability in the amount of fresh water in the system
Spatio-temporal variability in freshwater pathways
Temporal variability in the anomalies' occurrence
Spatial variability in the anomalies' occurrence

Figure 17

Table_1

	Pixels	Coordinates	Radius	LLR	p-value
Primary cluster Figure 4b	8,9,10,14,15,16,20,21,22	3,3	1.41	9.5	<0.05
Primary cluster Figure 4c	15,16,17,21,22,23,27,28,29	4,4	1.41	6.9	<0.05
Primary cluster Figure 5	29	5,5	0	7.7	<0.05

Table 1. Detected clusters in Figures 4 and 5: pixels, coordinates, radius, LLR and p-value. These clusters present a number of cases higher than the expected one.