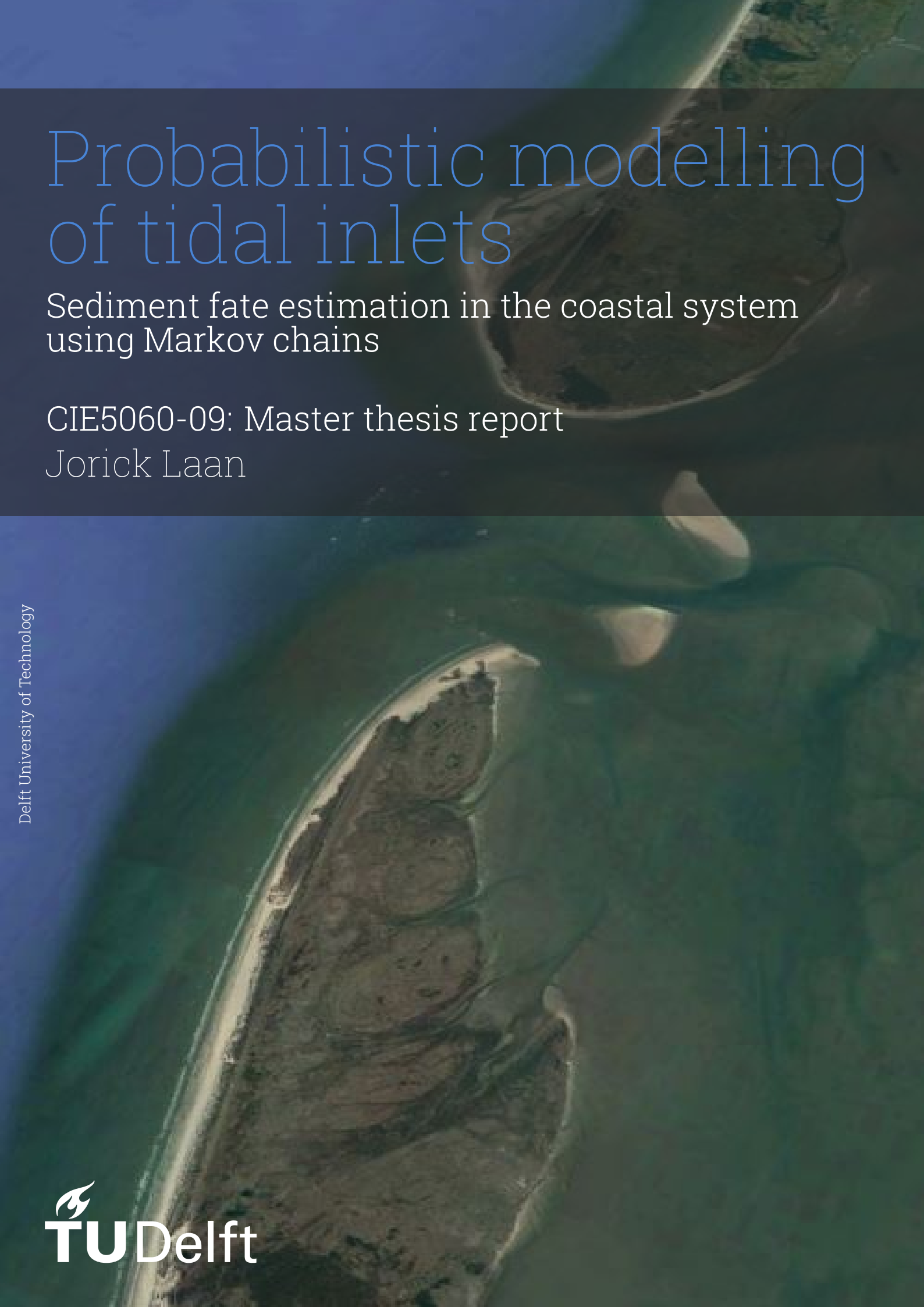


Probabilistic modelling of tidal inlets

Sediment fate estimation in the coastal system using Markov chains

CIE5060-09: Master thesis report

Jorick Laan

Delft University of Technology

TU Delft

Probabilistic modelling of tidal inlets

Sediment fate estimation in the coastal system using Markov chains

by

Jorick Laan

to obtain the degree of Master of Science
at the Delft University of Technology,

Student number:	4343409	
Thesis committee:	Dr. S.G. Pearson	TU Delft, chairman
	Dr. J.A. Álvarez Antolínez	TU Delft
	Dr. ir. O. Morales Napoles	TU Delft
Project Duration:	September, 2023 - August, 2024	
Faculty:	Faculty of Civil Engineering and Geosciences, Delft	

Cover: Snapshot of Terschelling and Ameland tidal inlet (Netherlands) using Google Earth

TU Delft

Acknowledgments

This thesis marks for me the end of a long journey in completing my Master of Science in Civil Engineering at the TU Delft. Doing the Hydraulic Engineering track with both specialisations Environmental Fluid Mechanics and Coastal Engineering was not the easiest, but I really enjoyed following my curiosity this way into complex problems. I am really glad with the opportunity I got to do this project. It really feels like it has helped move the field one step forward into doing something new and innovative.

First I would like to thank my committee for all the help, guidance, patience and enthusiasm on every step during the project. Special thanks here to my chairman Stuart G. Pearson, who gave me such an interesting thesis topic to work on and delve into. The weekly meetings and positive interactions were a huge help that this thesis would probably not have been possible without it. The project did not dive too much into trying different wave climates, but a thank you to José A.Á. Antolínez for the ecplanations and suggestions that were made. Also thank you to Oswaldo Morales Napolez for all his expertise on Markov Chains and the advice to develop the model in the direction of a Dynamic Bayesian Network.

Also thank you to all the friends and family that were there for me during this period. The students of 3.95 were great company to work through the days and enable me to learn some of the practical sides of working on a thesis. Thank you to my family for the support and patience that I would be able to finish my Master eventually. A last thank you to Johan Reyns who although I did not have to use it, gave me access to the SedTRAILS repository.

There was so much to work on during this period that I am not yet sure what will be next for me. One thing is certain though, that I will find my pathway through life to follow my curiosity.

Jorick Laan
Delft, August 2024

Abstract

Under tidal currents and wave action, sediment particles show complex transport patterns for coastal systems like tidal inlets. To model the sediment transport in such coastal systems, models have been derived that can predict that transport using Eulerian (grid-based) or Lagrangian (particle-tracking) methods. The sediment transport model SedTRAILS uses a Lagrangian method to predict the pathways that sediment particles will follow. As a process-based model SedTRAILS is limited in how fast it can compute results. Here we show a methodology to make a faster probabilistic metamodel from the process-based model SedTRAILS by creating an *Auto-regressive Hidden Markov Model* (ARHMM). Creating a probabilistic model of the sediment transport in a coastal system using Markov chains is a novel approach in that it builds further on the connectivity approach of SedTRAILS, which is still quite a new development in the field of sediment transport at present. For different versions of the model, correspondence with physical properties of the system were found, that the development to estimate sediment pathways accurately look promising. The method of describing sediment transport pathways of a coastal system as an ARHMM is a form of probabilistic *Machine Learning* allowing for faster computation times by describing the patterns the system can make. This is in contrast to a process-based model that has to compute all of the underlying processes. The autoregressive component in this approach is expected to be essential in describing any type of dynamical system that consists of trackable pathways. Taking into account the position of a particle at the current time step highly limits the position a particle can have at the next time step in favor of accurately estimating sediment pathways. The modelling approach discussed can find uses in engineering applications where particle pathways are of importance like the dispersal of nourishments, but also other dispersion events like dispersed goods from when a vessel loses a container, as long as there is a process-based model available to create the necessary Lagrangian data.

Contents

List of Abbreviations

A
ARHMM *Auto-regressive Hidden Markov Model.* ii, 14, 23, 25–28, 34–36, 42–45, 47–52, 58, 60–62

B
BN *Bayesian Network.* 7–12, 24, 26, 42, 50

C
CPD *Conditional Probability Distribution.* 8, 12, 13

D
DBN *Dynamic Bayesian Network.* 7, 11, 12, 23–26, 28, 34, 42, 47, 50

E
EM *Expectation Maximization.* 15, 27, 28, 42, 43

H
HMM *Hidden Markov Model.* 7, 10–14, 25, 42, 50

K
KFM *Kalman Filter Model.* 10, 25

M
MC *Markov Chain.* 7, 9
ML *Machine Learning.* ii, 26, 43, 49

P
PRNG *pseudorandom number generator.* 43

R
RWS Rijkswaterstaat. 1, 27, 31, 33, 51

S
SON Schiermonnikoog Noord. 17
SSM *State-Space Model.* 7, 10, 25, 26, 43, 50

1 Introduction

1.1. Sand nourishments in the Netherlands

The Dutch coast uses a dynamic conservation strategy as maintenance policy for flood safety consisting of sand nourishments to reinforce the sandy coastline (Lodder & Slinger, 2022). Since 1990 more than 300 nourishments have been done, making it one of the most nourished coasts globally (Brand et al., 2022). The current policy is part of the Delta programme of the Netherlands, which is taken care of by the Dutch Ministry of Infrastructure and Water Management and executed by Rijkswaterstaat (Glas, 2023). To predict the effects of dredging, nourishment, and natural processes involving sediment, a large history of *sediment transport* theory and models have been developed over decades. Present day predictions of sediment transport still have their limitations, because of the complexity of hydrodynamical processes and the empirical nature of equations in this field of research (Bosboom & Stive, 2021). *This is why although simple cases of sediment transport can be modelled, complex cases take more time to model or are too computationally expensive to be modelled at all.*

1.2. Regular shoreline maintenance

The existing options for *dynamic coastline preservation* by sand nourishments consist of beach nourishments, shoreface nourishments and channel wall nourishments (Brand et al., 2022). Of these measures, shoreface nourishments are the cheapest, but beach nourishments are still to be considered based on morphology, the local setting and the purpose of the nourishment (Brand et al., 2022). The placement of such nourishments is based on theory about cross-shore and longshore sediment transport.

For cross-shore transport, summer-winter seasonality (Wright & Short, 1984) explains that sand lost after storms in the winter is not lost forever, but just moved to the foreshore. During the calmer months, like in the summer, this sediment moves towards the coast again to build up the coastline again resulting in a yearly cyclic behaviour. Increase of the water level by sea-level-rise has a different effect on the cross-shore profile, which is described by *Bruun rule* (Bruun, 1954, 1962). This rule explains that the cross-shore profile tends to move landwards with sea-level-rise, but because cities are often built very close behind the dunes that is not possible. As a result, maintenance of the Dutch coastline mainly consists of supplying sediment to the active coastal zone that the dune profile can adapt to sea-level-rise (Lodder & Slinger, 2022).

Longshore transport of sediment is caused by the fact that waves coming in to the coast from an angle will cause a force on the sediment in alongshore direction. The angle at which waves come in is not always the same, such that the longshore transport is the residual effect of the waves over time resulting from the local *wave climate*. For the Dutch coast this causes that sediment supplied by rivers moves in northwards direction by alongshore transport like a conveyor belt. This concept was also used in the *Sand Motor* project (Stive et al., 2013), where a large nourishment ($> 20Mm^3$) was done at a specific location making longshore transport distribute the sediment along the Delfland coast over time.

1.3. Tidal inlets

Between many stretches of sandy coastline there are interruptions that connect the sea to basins behind the coastline, which together are called *tidal inlets*. Tidal inlets were formed during post-glacial periods by sea level rise flooding low-lying areas (Bosboom & Stive, 2021). These gaps in the shoreline exchange water and sediment with the basin behind the shoreline and are prevented from closing by tidal currents.

Currents and main components

Under the influence of ebb and flood the water level in the basin changes over time constantly drying and flooding areas in the basin called *tidal flats*. As currents can not flow over dry areas during ebb there is a distinctions in *residual currents* that shoals and tidal flats are generally flood-dominant in contrary to channels that are generally ebb-dominant.

In Figure 1.1 a sketch is shown of an example of a tidal inlet system. The system mainly consists of tidal flats and channels defining an ebb-tidal delta at the sea-side and a flood-tidal delta at the basin side, which are connected by a tidal gorge in between. The basis of how these elements work is that flows going into the gap accelerate that they can come from all angles, but when going out of the gap velocities are high that the flow forms a more straight *jet*. As a result the ebb-tidal delta shows *marginal flood channels* (Fig. 1.1, [1]) near the coastlines and a deeper *ebb channel* that is in the center. The other areas of the ebb-tidal delta consist of shallows that causes waves from the sea to break and shelter the basin (Fig. 1.1, [4,5]). In the flood-tidal delta currents will flood the tidal flats (Fig. 1.1 , [6]). Waves that do get into the basin will propagate mostly through the channels by what is called *channeling* (Bosboom & Stive, 2021). During ebb water levels drop that currents will be mainly in the ebb-dominant channels.

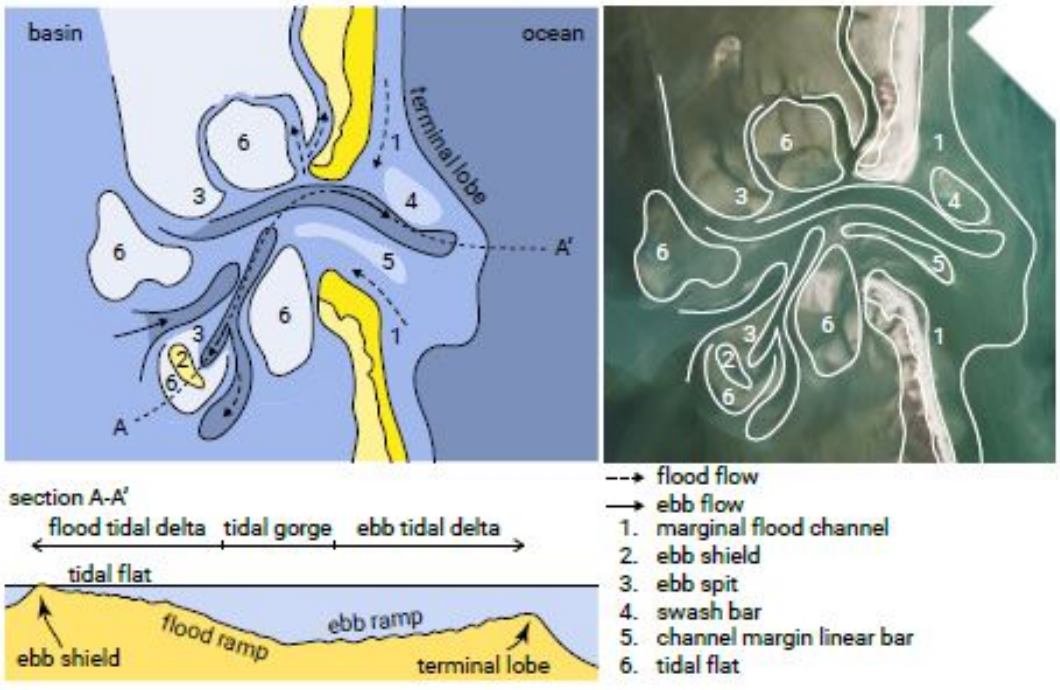


Figure 1.1: Sketch of the morphological elements of a tidal inlet based upon a photo from an inlet near Sheep Island, North Carolina (2016 imagery from NC OneMap Geospatial Portal (2020)). The lower part of the figure shows a cross-section of the depth variation along the inlet. From Bosboom & Stive (2021)

Empirical relationships

Shore maintenance for tidal inlets is less simple than for regular shorelines. The forces of wind, waves and tides causes complex 3D current patterns in the system that the relations we use for normal coastlines are not applicable. For this reason most of the formulas developed for tidal inlets are empirical. More recent works use computational models that can better describe the processes, but have computational limitations in space and time resolutions.

Some of the first works were done by Escoffier (1940) determining how stable the existence of a tidal inlet is and O'Brien (1931, 1969) determining a relation between the *tidal prism* and inlet cross-sectional area. The *tidal prism* is a valuable characteristic to define tidal inlets being the amount of water entering the basin during flood and leaving during ebb (Bosboom & Stive, 2021). This has been defined in most works using the geometry of the basin in combination with the tidal range (Elias et al., 2019).

Another empirical relationship is that between the tidal prism and ebb-shoal volume by Walton Jr & Adams (1976). Increasing the basin area due to sea-level-rise, called the *accommodation space*, will cause extra sediment to be imported into the basin due to the basin trying to maintain equilibrium volume (Ranasinghe et al., 2013). In (Elias & van der Spek, 2006) this causes the Wadden Sea to take sediment from the ebb-tidal delta and the beaches surrounding the tidal inlet and in (Zhang et al., 2004) it made it not possible to apply Bruun rule in almost 70% of their $220km$ study area. Conceptual models and empirical formulas have been developed over time, but as a shortcoming they often lack descriptions of the underlying physics, which are essential for understanding for human interventions at a smaller scale (Elias & van der Spek, 2017). These coastal systems are found in many areas in the world (Mulhern et al., 2017) that will be influenced by sea-level-rise and increased storm activity caused by climate change. Better understanding is expected to become increasingly important to keep their function as coastal defence.

To understand the processes leading to these results the morphology of the ebb-tidal delta takes a major role. Being at the most active part of the system the sediment stored in this part participates the most in sediment exchanges happening in and around the system (Elias et al., 2019). The main process of sediment traversing from the updrift to the downdrift island of the tidal inlet is called *Sediment bypassing*. Bruun & Gerritsen (1959) were one of the first to use the concept of sediment bypassing to understand the patterns of shoals and channels on the ebb-tidal delta (Elias et al., 2019). They noticed the ratio between longshore sediment transport and tidal inlet currents is of influence on the way sediment traverses the tidal inlet. This ratio determined two mechanism, which are *flow bypassing* and *bar bypassing*. In Figure 1.2 both the sediment drift and wave-driven currents entering over the shoals and returning through the channels are shown to visualize *bar-bypassing* on the outer delta. In case of *flow bypassing* sediment follows the channel.

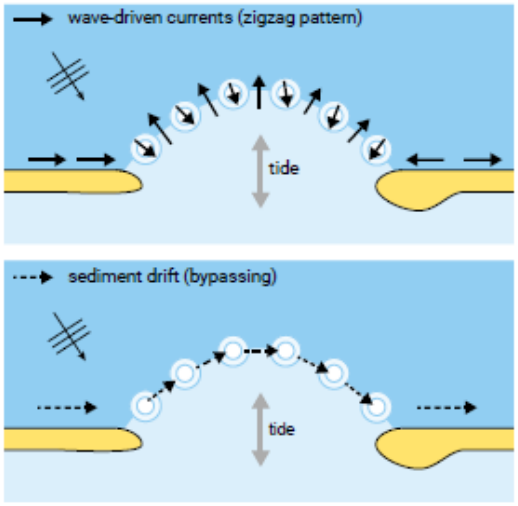


Figure 1.2: Sketch visually explaining the concept of sediment bypassing by shoal migration from (Bosboom & Stive, 2021).

Various other conceptual models were developed to explain the variability of shoals on tidal inlets. Some that are used a lot are Davis & Hayes (1984) who determined a relation classifying the inlet morphology using the relative importance of wave versus tidal energy. Also for mixed-energy coasts FitzGerald et al. (1978) used the relationship between the stability of the inlet throat and the movement of the main ebb-channels to created 3 conceptual models for sediment bypassing. These are *inlet migration and spit breaching*, *stable inlet processes* and *ebb-tidal delta breaching*.

Sediment transport:
More recent research has the availability of a lot more tools to analyse the behaviour of the shoals and the corresponding sediment transport. For example Ridderinkhof et al. (2016) used satellite images to analyse changes of shoals in the ebb-tidal delta seawards of Osterems. Other research using process-based numerical models has found the bypassing mechanism from Bruun & Gerritsen (1959) as different pathways of sediment transport. Herrling & Winter (2018) show a classification of 3 possible sediment bypassing mechanisms that can be seen in Figure 1.3. The flow-bypassing by the main channel (A) and bar-bypassing by the outer delta (C) were again found, but also a mechanism of sediment recirculation (B). These mechanisms were shown to be different for different sediment sizes as shown in Figure 1.4. For Otzum ebb-tidal delta and Texel inlet it was seen that the process of sediment recirculation is dominant (Elias et al., 2019). This means that for some tidal inlets sediment may not bypass to the downdrift side of the tidal inlet.

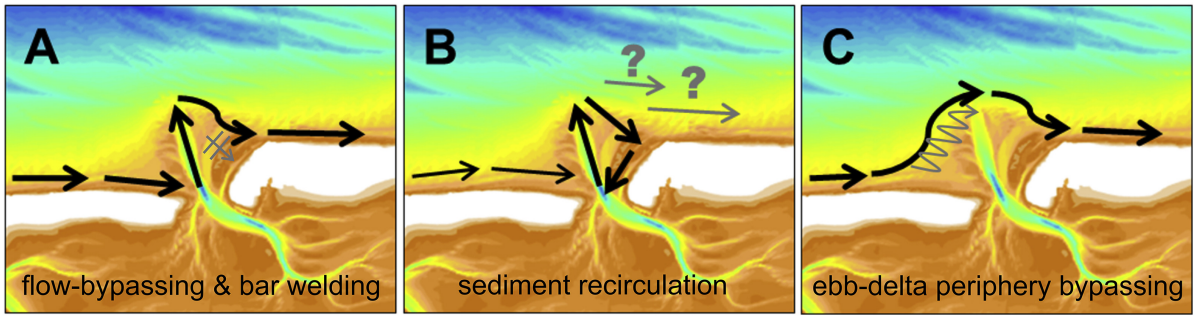


Figure 1.3: Sediment bypassing mechanisms by Herrling & Winter (2018).

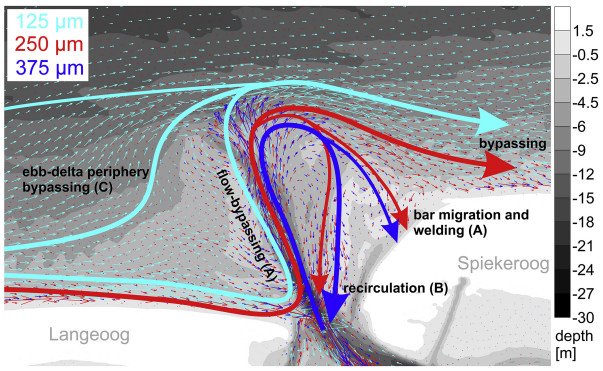


Figure 1.4: Example different sediment sizes influencing bypassing mechanisms by Herrling & Winter (2018).

In Elias et al. (2019) the large documenting history of Ameland inlet was used to analyse the bypassing processes happening. From this it was found that for a full conceptual model of tidal inlets a *scale-cascade model* using the concept of De Vriend (1991) gives a useful structure to help understand changes in the system. In Figure 1.5 this *scale-cascade model* for Ameland inlet is shown. Four different levels were recognised being the Wadden Sea, the tidal inlet system, the ebb-tidal delta and individual shoals. Different scales can interact with each other that the small scale individual shoals can cause the larger scale main ebb-channel to switch. The opposite also happens that at the baisn scale land reclamations and levee building since 1600 CE have caused the tidal divides to move eastwards, but the inlet was prevented to move the same way. As a result the inlet gorge at the ebb-tidal delta scale moved eastwards.

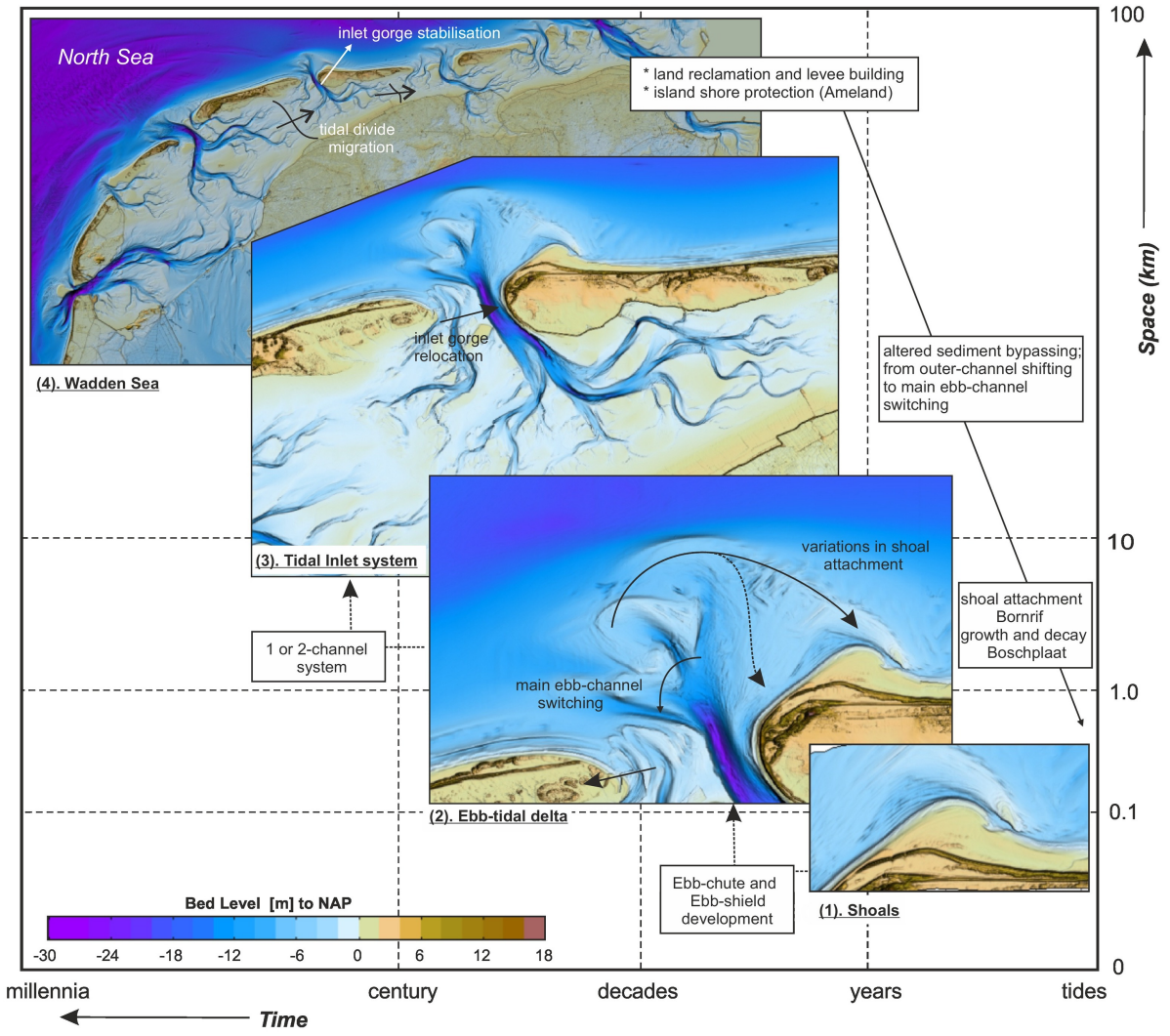


Figure 1.5: Dynamics of Ameland inlet over different scales from (Elias et al., 2019)

1.4. Sediment pathways

In recent years much research has been done on the topic of tidal inlets. As part of the Dutch research programme Kustgenese 2.0 (Lodder & Slinger, 2022) Pearson et al. (2020) made big steps in understanding the underlying physics of sediment transport in tidal inlets by describing the system in terms of connectivity and the development of the Lagrangian model SedTRAILS (Pearson et al., 2023).

We learned from that, that dividing the complex system into a network of nodes and links gives a quantitative framework to analyse the sediment transport patterns in the system. Connectivity in the system was found to be a function of sediment grain size identifying major bypassing paths across the tidal inlet (Pearson et al., 2020).

SedTRAILS is a *Lagrangian model* that uses hydrodynamic data and computes *sediment pathways* (Pearson, 2022). Lagrangian models are different from Eulerian models in that they follow the particles and calculate trajectories from the forces acting on these particles. These sediment pathways describe the position of particles released at different areas acting as sources in the model. This information can be put into a connectivity network, which gives a description of tidal inlets that can be analysed using various mathematical techniques that are available for networks including graph theory.

1.5. Markov chains

The connectivity network divides the system into some defined areas and tracks how many particles are in which area in comparison to the area the particles started in. Another representation of the sediment pathway data is in the form of an *adjacency matrix*. This type of matrix describes the pathways of all particles in the system by putting in the rows the source areas of the particles and in the columns the position of those particles at that moment in time (Pearson et al., 2020). When instead of the amount of particles probabilities are calculated a transition matrix is obtained that mathematically is equivalent to a *Markov chain*. Markov chains are discrete state Markov processes (Ibe, 2013) that describe one or multiple random variables in which each state only dependent on the state before it (Everitt, 2006). This is also called the *Markov property*.

Having the transitions as probabilities creates a lot of opportunities to describe the system using probability, which can be used to create a probabilistic model of the system. One of the biggest advantages of probabilistic models is that they do not need to describe the exact processes, which is why a probabilistic model of the system using Markov chains is expected to be much faster than how SedTRAILS can predict sediment pathways currently. Other advantages can be found in how often results in terms of probability are desired. Input parameters like the wave climate can have uncertainty in itself, which would be easier to implement into a probabilistic model than a deterministic model. Also for engineering purposes results are expected to include some uncertainty that having a probabilistic model could directly include those effects that can not be computed deterministically.

1.6. Research questions and objectives

To approach this the main research question in this thesis is defined as the following:

How can we use sediment pathways data of complex tidal systems to create a probabilistic model that can make fast and accurate estimates of those pathways using Markov chains?

To answer this the first 2 research questions go into how to create the model by explaining the type of model chosen and the steps that are necessary to get to that model:

RQ1: *What type of probabilistic model could be made for sediment pathways by using Markov chains?*

RQ2: *What processing steps are necessary to create the probabilistic model?*

The last 2 research questions tie the thesis back to the main research question by looking at the accuracy and computational speed of the obtained model and compare it to the original sediment pathways model used SedTRAILS:

RQ3: *How accurate are results of the probabilistic model in comparison to SedTRAILS results?*

RQ4: *How fast can the probabilistic model compute results compared to SedTRAILS?*

1.7. Outline

This thesis will first in Chapter 2 bridge the knowledge gap between readers experienced with the relevant probabilistic models for this thesis and those who are not. Next Chapter 3 will discuss the methodology for developing the probabilistic model using Markov chains of Ameland tidal inlet. This includes the steps: schematizing the wave climate, modelling the hydrodynamics with Delft3D, transforming the hydrodynamic data with SedTRAILS to obtain a synthetic database of sediment pathways and developing the probabilistic model from that data. In Chapter 4 the obtained model will be compared for different settings for the model. In Chapter 5 the results are discussed to analyse how the model performed in terms of strengths and limitations the model has to discuss the opportunities this will bring for future research. In Chapter 6 a conclusion will be made answering the research questions that were defined.

2

Theoretical background: probabilistic models

This chapter is to bridge the gap between readers that are not very experienced with probabilistic graphical models and readers that are. Most of the theory is based on (Murphy, 2002), which is why for a more extensive coverage of *Dynamic Bayesian Network* (DBN) models the reader is recommended to read that thesis. Especially for the last section there are more algorithms that were not discussed for brevity.

First will be discussed how probabilistic models are different from process-based models. After that *probabilistic graphical models* will be explained by covering *Bayesian Network* (BN) and *Markov Chain* (MC). The following *State-Space Model* (SSM) will be a different group of models with latent variables, which will discuss the specific example of a *Hidden Markov Model* (HMM) and how it can be described as a DBN. The last section will discuss for probabilistic graphical models the general concepts of the algorithms to learn parameters and obtain information from the model using *inference*.

Some of concepts in this chapter are very theoretical that the connection to sediment transport can be difficult to grasp. For those it may be mentioned that this connection becomes more clear in Chapter 3 explaining the methodology. Any reader that has a good understanding of the topics of this chapter may also continue to that chapter immediately.

2.1. Process-based vs probabilistic

To estimate sediment transport pathways in complex coastal systems, we want to use a probabilistic method to get a model that is faster and can handle stochastic variables more efficiently than the existing process-based models available.

Existing models for sediment transport in tidal inlets can be classified into process-based models, aggregated (semi-) empirical models and idealised models which are simplified process-based models (Ranasinghe, 2020). Some examples of these models are the idealised model of Schuttelaars & de Swart (2000), the process-based models Delft3D (Elias & Hansen, 2013) and the aggregated (semi-) empirical model ASMITA (Wang et al., 2020). An overview of many other sediment transport models can be found in (Hoagland et al., 2023). Ranasinghe (2020) mentions that there is a need for new models in that the existing process-based models are only accurate for simulation times of less than 5 years and that highly aggregated models lack providing much insight into the processes governing morphological evolution.

Beside the increase in computational power a new generation of models was also mentioned that there is a demand for probabilistic models for risk informed coastal management (Jongejan et al., 2016; Wainwright et al., 2015). Also for the input of variables it was mentioned by Pearson (2022) that wave forcing, morphodynamic feedbacks, sensitivity of initial conditions and climate change are of a stochastic nature that probabilistic modelling approaches are a promising way forward. In Hirschberg et al. (2011) the importance of uncertainty information is also mentioned for weather forecasting showing that this information is of importance. Some work has already been done on probabilistic models for sediment

transport. For fluvial sediment transport Schmelter et al. (2012), Schmelter et al. (2011), and Schmelter & Stevens (2013), Schmelter et al. (2015) developed a Bayesian model for sediment transport. More recently Kroon (2024) analysed uncertainties in predicting large-scale interventions in the coastal zone.

These methods have in common that they take into account uncertainty by making variables in the process-based model stochastic to get probabilistic results. As a way to get the uncertainty of results these methods are good for validity, but to speed up computations a different representation of the system would be needed. That is where in this thesis the connectivity description of SedTRAILS comes in. Where the other methods could be seen as a bottom up approach taking into account the uncertainty of initial variables to compute the uncertainty of the final variables, the connectivity approach can be seen as more of a top down approach by computing the sediment particle pathways and obtaining probabilities when describing those pathways using connectivity. In that matter there could be here a trade-off between the two methods of creating probabilistic models in terms of validity and computational speed.

2.2. Probabilistic graphical models

To describe dependencies between random variables in our model we will use what is called a *probabilistic graphical model*. This is a type of model that can be seen in Figure 2.1 consisting of *vertices*, which are also called nodes or points, connected by *edges*, which are also called links or lines. The vertices represent the states of the system being the random variables, while edges represent dependencies within the system that are described using *Conditional Probability Distribution* (CPD). This representation we can also call a *graph* and this terminology is directly from the mathematical field of *graph theory*. The connectivity description of sediment pathways by SedTRAILS is also a *graph*, which explains why a *probabilistic graphical model* is a logical choice for the model.

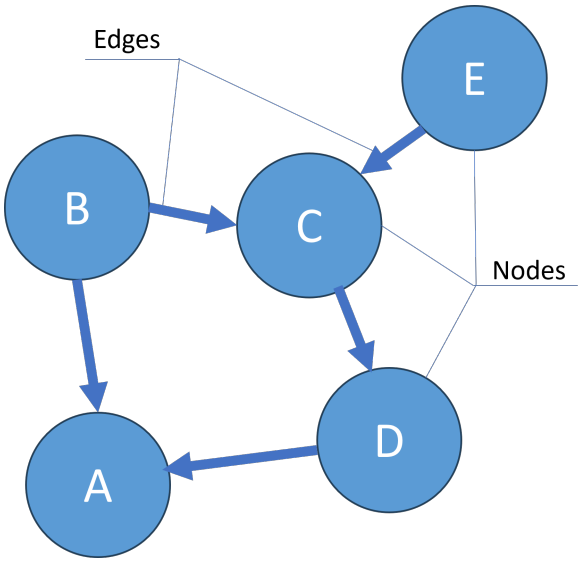


Figure 2.1: Example probabilistic graphical model of a BN showing the edges and nodes

The basic idea of these type of models is that by using a prior distribution of a random variable and the CPD relating it to another random variable, you can get the joint probability distribution of both random variables. When there is a description of the joint probability distribution of the full system, obtaining the probability distribution of any random variable in the system is a matter of marginalizing.

2.2.1. Bayesian networks

The most basic form of a probabilistic graphical models is a *Bayesian Network* (BN). In this case the network describes dependencies between random variables in the system without time playing any role. The network is Bayesian in that this model uses *Bayes rule* shown in Equation 2.1 to compute the joint probability distribution $P(B \cap A)$, but for all the random variables in the network. By this same equation also the probability of a variable given some evidence can be obtained by calculating $P(A|B)$ when B is the evidence.

$$P(A|B) = \frac{P(B \cap A)}{P(B)} = \frac{P(B|A)P(A)}{P(B)} \tag{2.1}$$

This type of model is used by many fields to describe causal dependencies between different random variables in a system. The structure of such systems can be learned from data describing the parameters of interest (Ekanayake & Zois, 2022). A common example is that from the medical field determining the relations between symptoms leading to a disease. Symptoms in such a system would be the discrete random variables that can have a value true or not true. When there is evidence of a symptom the probabilities in the system are updated to take into account the confirmed information about the system. For sediment transport we would want these to be variables that are naturally stochastic like wave climate parameters or the effect of sea level rise.

The relations in a BN are causal in that they only have to go in one direction and have no directed cycles. The graph being directional in this way gives the advantage that from the structure it can be seen which sets of variables are independent of other sets of variables. A criteria called *d-separation* also helps with this definition by inspecting which sets of variables are independent given a third set (Pearl, 1988). In Figure 2.1 an example of a BN was shown. It can be seen that if edge B to A was the other way around there would be a cyclic structure, which would result in a system that is not a BN.

2.2.2. Markov chains

A different form of probabilistic graphical model can be created by using *Markov Chain* (MC). *Markov Chain* consist of one or multiple random variables of which the states show the *Markov property* (Ibe, 2013). This property says that every state in the system only has limited memory. In the case of the *strong Markov property* this means that each next state is only dependent on the current state shown in equation 2.2. This assumption may not be true for all processes, but for processes varying in time this is often a good assumption.

$$P(X_{n+1} = x|X_1 = x_1, X_2 = x_2, ..., X_n = x_n) = P(X_{n+1} = x_{n+1}|X_n = x_n) \tag{2.2}$$

Different from BN *Markov Chain* does allow cyclic structures to describe a system. This is useful for physical systems like sediment transport as transitions are possible by allowing sediment particles to stay in one area or to return to an area depending on the forces acting on it.

A Markov chain can be visually described by a *state transition diagram*. In Figure 2.2, two examples of transition diagrams representing one random variable with three possible states are shown. In both examples a cyclic structure can be seen, but over time they behave quite differently. In the left example states A,B and C are *recurrent* that over time they can keep happening, but if $p > 0.5$ in the full system the probability of state C to occur will be higher than state B and the probability of state B to occur will be higher than state A. In the right example it can be seen that states A and B are only *transient* before going to state C being the recurrent state. The definitions of recurrent and transient states can also be described mathematically with the descriptions below:

- Recurrent states: $P(T_i < +\infty|X(0) = i) = 1$
- Transient states: $P(T_i < +\infty|X(0) = i) < 1$

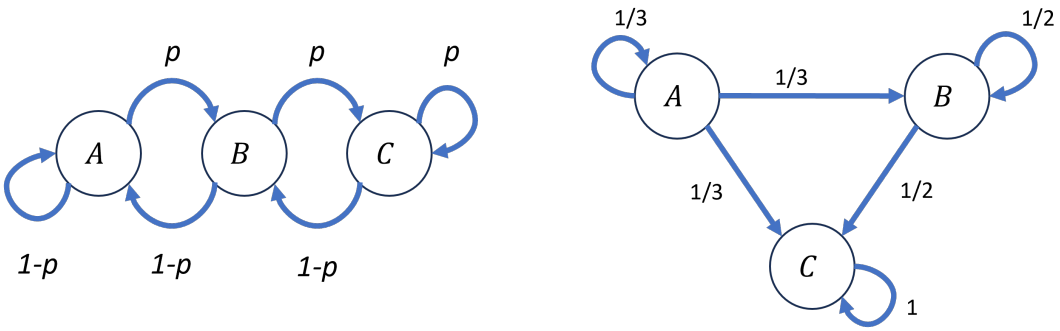


Figure 2.2: Two examples of Markov chains as state-transition diagrams. The left example shows a system with only recurrent states in contrary to the right example showing a system in which nodes A and B are transient states and node C is the only recurrent state.

To describe all the different transition probabilities from a state transition diagram they can be written more mathematically and compact as a matrix. The obtained matrix is called a *transition matrix* like shown in equation 2.3. The rows of the transition matrix represent the current state and the columns give for each row the probabilities of where it can end up at the next time step. As a result each rows sums up to a probability of one. Using an initial probability distribution between the nodes of the system the transition matrix can repeatedly be applied to find the probability distribution over the system in the next time steps.

$$\underbrace{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}}_{\text{Initial distribution}} \underbrace{\begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1n} \\ p_{21} & p_{22} & \cdots & p_{21} \\ \vdots & \vdots & \ddots & \vdots \\ p_{n1} & p_{2n} & \cdots & p_{nn} \end{bmatrix}}_{\text{Transition matrix}} = \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}_{\text{Next timestep}} \tag{2.3}$$

Like with a BN a model can be made to describe the relations between multiple random variables using Markov chains. Such a system is called a *Markov random field* or *Markov network*. Different from a BN the edges in a Markov random field are undirectional allowing cyclic structures. As a consequence the joint probability distribution is also differently described by Equation 2.4 (Murphy, 2002). In this equation the ψ represents potentials between random variables and Z is a normalisation factor.

$$\begin{aligned} P(x) &= \frac{1}{Z} \prod_{C \in \mathcal{C}} \psi_C(x_C) \\ Z &= \sum_x \prod_{C \in \mathcal{C}} \psi_C(x_C) \end{aligned} \tag{2.4}$$

To describe sediment transport pathways in the complex system of a tidal inlet using a Markov random field, the connectivity approach developed by Pearson (2022) can be used. It was suggested to use the method of ensemble modelling by running the Lagrangian model SedTRAILS multiple times for different hydrodynamic conditions and create a transition matrix using those conditions. By analogy of the Snakes and Ladders example by Althoen et al. (1993), this could be used to create a state absorbing Markov chain in discrete time. An absorbing Markov chain is a Markov chain in which every state can reach an absorbing state, which is a state that once entered cannot be left.

An ecological example of using Markov chains for connectivity can be found in Bacher et al. (2016). Different from the suggested method they used an Eulerian approach to create the transition matrix.

2.3. State-space models

State-Space Model (SSM) are different from probabilistic graphical models in that they always include some underlying hidden states. Having hidden states in the system gives the advantage that variables can be included in the system that are desirable to know but can not be measured (Murphy, 2002).

2.3.1. Hidden Markov models

Very common SSM are *Hidden Markov Model* (HMM) and *Kalman Filter Model* (KFM). The KFM is not explained here as the research for this thesis was mainly focused on Markov models. The structure of an HMM can be seen at the left side in Figure 2.3. The upper row of the HMM shows the hidden states A,B and C behave as a Markov chain, which causes some the observed states 0 and 1 to happen. The transitions between the hidden states are described using a transition matrix like in the *Markov random field* before. The relations between hidden variables and observed variables are described using an *emission matrix*, which is a conditional probability distribution. Similarly to the *transition matrix*

each row of the *emission matrix* has to sum up again to one, but as the latent variables are the rows and the observed variables are the columns this matrix does not have to be square in shape.

The observed values allow extra information to be added to get better estimate of the sequence of hidden states described by the transition matrix. The obtained model this way can be used in various ways which are classified by the term *inference*. Some examples of *inference* are computing the most likely sequence of hidden states and both hind- and forecasting observations.

2.3.2. Dynamic Bayesian networks

A different way to describe the HMM structure is by using the *Dynamic Bayesian Network* (DBN) structure at the right side of Figure 2.3. Like a BN, a DBN describes causal relationships between random variables in the system that the edges only go in one direction. As a difference, DBN do change in time by repeating the BN structure for each time step. Most state-space models can be described as a DBN giving the advantage that structure changes can be made that algorithms can easily be chosen to adapt to these changes (Murphy, 2002). Similarly to the BN, the joint probability distribution of the DBN can be described by adding the time dependency as another product resulting in equation 2.5. In this equation $Pa(Z_i^T)$ are the parent nodes of Z_t^i.

$$P(Z_{1:T}) = \prod_{t=1}^{T} \prod_{i=1}^{N} P(Z_t^i | Pa(Z_t^i)) \tag{2.5}$$

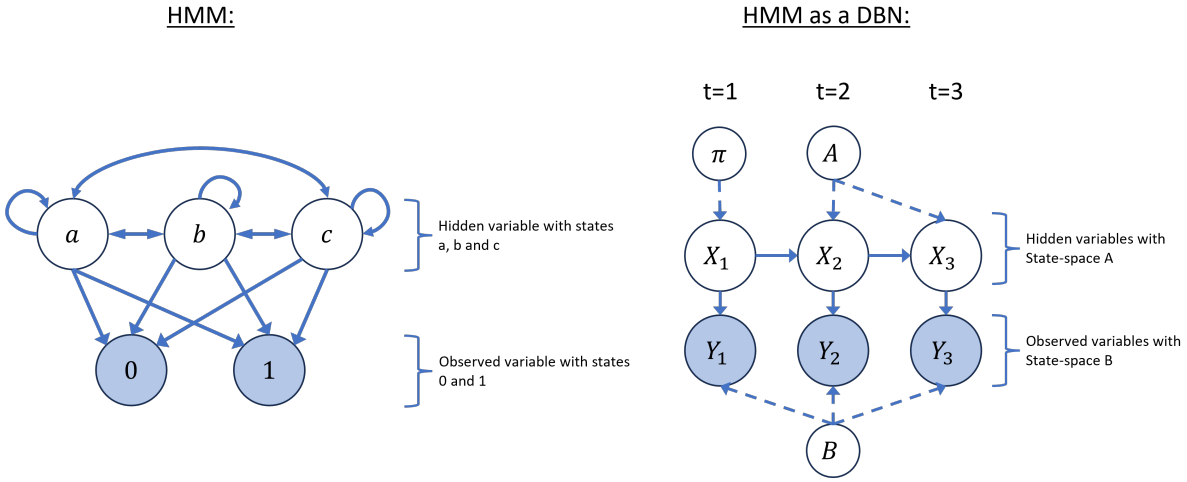


Figure 2.3: Example structure HMM as DBN. π represents the initial distribution of variable x at $t = 1$, A represents the conditional probabilities of transitions from X_t to X_{t+1}, B represents the conditional probabilities describing the relation between X and Y and π represents the initial probability distribution of X.

2.4. Inference and learning for probabilistic graphical models

So far only descriptions of the different types of models were discussed, but to use probabilistic graphical models some algorithms are necessary.

The first type of algorithms are for *inference*, which is the process of drawing conclusions about a population by using observations of a sample of individuals from that population (Everitt, 2006). For probabilistic graphical models this practically means all the methods to compute the probabilities of a set of query variables X_Q given some evidence X_E. Using *Bayes rule* this is done by computing the conditional probability distribution as described in equation 2.6.

$$P(X_Q | X_E) = \frac{P(X_Q, X_E)}{P(X_E)} \tag{2.6}$$

Having a description of the joint probability distribution of the full system *inference* mainly consists of the process of marginalizing joint distributions (Murphy, 2002). There are many types of inference, which are: *filtering*, *smoothing*, *prediction*, *control*, *Viterbi decoding* and *classification* (Murphy, 2002). The goals of these methods are different, but they are all variations to the main method for *inference* that will be discussed in the next section.

There is also the distinction between *exact* and *approximate inference*. *Approximate inference* is necessary for large structures like grids as in this case exact inference may quickly become computationally expensive (Murphy, 2002).

The second type of algorithms necessary are for *learning*. There are two types of learning, which are *structure learning* and *parameter learning*. *Structure learning* is less important for this thesis, but the general concept of it is that the structure of a *probabilistic graphical model* can be learned from a dataset by first assuming all variables are connected and finding the zeros in the system. When not all the *Conditional Probability Distribution* (CPD) in the system are known *parameter learning* can be used that for given data determines the parameters for the system that gives the *maximum likelihood* or the *maximum a posteriori* of the model. For *learning* algorithms *inference* is a subroutine making it more computationally expensive than *inference*.

2.4.1. Exact inference

There are two main methods to do *exact inference* on a probabilistic graphical model. The first method called *variable elimination* does this by using the description of the joint probability distribution of the system and repeatedly "eliminate" a node from the system by marginalization. In a system with N nodes this computation takes $O(N)$ time, but to do this computation for each of the N nodes it would take $O(N^2)$ time (Murphy, 2002). The second method reduces computing the marginals of all the nodes to $O(N)$ time by storing the terms calculated from "eliminating" nodes to re-use them for computing the other marginals. This is done by turning the system into a *junction tree* data structure and using a method called *message passing* to compute all the marginals of the system. To make these methods clear, they will be discussed in more detail including a simple numerical example.

Variable elimination

For variable elimination of a BN the joint probability distribution can be described by Equation 2.7. In this equation $P(X_1, ...X_N)$ is the joint probability distribution of all N random variables in the system. $P(X_i|Pa(X_i))$ is the probability of a random variable i given the parent nodes node(s) of that variable. For example in Figure 2.1 node D only has one parent being node C, but node C itself has two parents being nodes B and E.

$$P(X_1, ..., X_N) = \prod_{i=1}^{N} P(X_i|Pa(X_i)) \tag{2.7}$$

To *marginalize* variables have to be *summed out*, which is described by Equation 2.8. *Summing out* a variable is done by summing the different states of that random variable in the conditional or joint probability table that, that variable is not present anymore in that distribution. The sum for each elimination can be put into separate λ structuring this process. The order in which variables are summed out affects how computationally expensive this calculation is. Finding the optimal order is a separate problem most of the time taken care of by the software used. less important for this thesis, but is extensively discussed in Murphy (2002).

$$P(X_1) = \sum_{X_2,...,X_N} P(X_1, ..., X_N) \tag{2.8}$$

In Huang & Darwiche (1996) it was addressed that some of the explanations of these methods can be difficult to understand. For that reason in this thesis the choice was made to explain *variable elimination* by means of Example 2.4.1 on the next page. To stay close to the specific structures considered in this thesis, we continue from Figure 2.3 by choosing the HMM as DBN *unrolled* for 3 time steps for the example. Also random variables are kept simple by giving them only 2 possible states.

Example 2.4.1: Calculating $P(X_1|Y_2{=}1)$ for a HMM by using variable elimination

For this example of variable elimination Figure 2.4 shows an HMM structure for 3 time slices with 2-state random variables. The values for $P(X_1)$ and the CPD were arbitrarily chosen. Writing out the joint probability distribution for this structure gives Equation 2.9.

$$P(X_1, X_2, X_3, Y_1, Y_2, Y_3) = P(X_1)P(Y_1|X_1)P(X_2|X_1)P(Y_2|X_2)P(X_3|X_2)P(Y_3|X_3) \tag{2.9}$$

It was chosen to compute $P(X_1, Y_2 = 1)$, which is the probability distribution of variable X_1 given the evidence variable $Y_2 = 1$. In equation 2.10 it can be seen that variables are summed out only over terms that include that variable starting from the right side. Generally terms that are summed out without influence of the evidence will sum to 1 not affecting the solution. For the evidence it is different that only the terms of the evidence are used for summing giving a result that is not equal to 1.

$$P(X_1, Y_2 = 1) = P(X_1) \sum_{Y_1} P(Y_1|X_1) \sum_{X_2} P(X_2|X_1) \sum_{Y_2=1} P(Y_2|X_2) \sum_{X_3} P(X_3|X_2) \sum_{Y_3} P(Y_3|X_3) \tag{2.10}$$

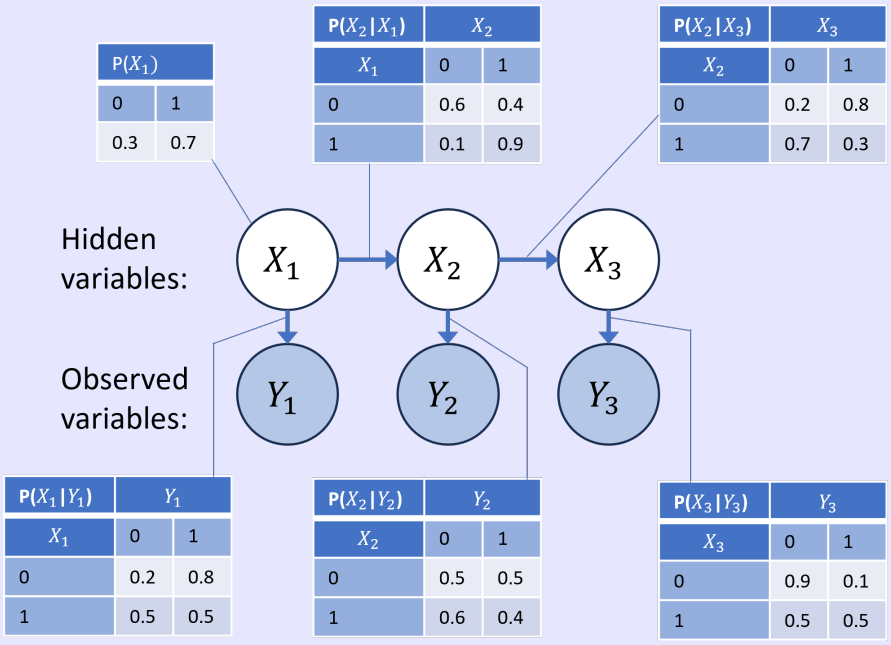


Figure 2.4: Example of an HMM with some arbitrarily chosen discrete random variables, each being able to only take on values of 0 or 1 for simplicity. $P(X_1)$ in this representing node X_1 and the conditional probability distributions representing the edges of the model.

To make the computation more efficient each sum term is put into a term λ. The subscript first shows what variable is summed and after the arrow shows what variable it will be summed next. Underneath the computations for the different lambda terms and the final result $P(X_1|Y_2 = 1)$ are shown.

$\lambda_{Y_3 \to X_3}(X_3) = \sum_{Y_3} P(Y_3|X_3) = [0.9 + 0.1 \quad 0.5 + 0.5] = [1 \quad 1]$

$\lambda_{X_3 \to X_2}(X_2) = \sum_{X_3} P(X_3|X_2)\lambda_{Y_3 \to X_3}(X_3) = [1 * (0.2 + 0.8) \quad 1 * (0.7 + 0.3)] = [1 \quad 1]$

$\lambda_{Y_2 \to X_2}(X_2) = \sum_{Y_2=1} P(Y_2|X_2) = [0.5 \quad 0.4] = [0.5556 \quad 0.4444]$

$\lambda_{X_2 \to X_1}(X_1) = \sum_{X_2} P(X_2|X_1)\lambda_{X_3 \to X_2}(X_2)\lambda_{Y_2 \to X_2}(X_2)$

$= [1 * 0.5556 * 0.6 + 1 * 0.4444 * 0.4 \quad 1 * 0.5556 * 0.1 + 1 * 0.4444 * 0.9]$

$= [0.5112 \quad 0.4556] = [0.5288 \quad 0.4712]$

$\lambda_{Y_1 \to X_1}(X_1) = \sum_{Y_1} P(Y_1|X_1) = [0.2 + 0.8 \quad 0.5 + 0.5] = [1 \quad 1]$

$P(X_1, Y_2 = 1) = P(X_1)\lambda_{X_2 \to X_1}(X_1)\lambda_{Y_1 \to X_1}(X_1)$

$= [0.3 \quad 0.7][0.5288 \quad 0.4712][1 \quad 1]$

$= [0.1586 \quad 0.3298] = [0.3247 \quad 0.6753]$

In the example some of the lambda terms are normalised, but that was just a choice to prevent rounding errors by numerical underflow (Huang & Darwiche, 1996). Normalizing can also be done at the end, which represents the denominator step in equation 2.6.

The summing out of variables to marginalize the joint probability distribution of the system has also been called the *sum-product algorithm*, when instead the *max-product* is used it results in *Viterbi's algorithms* computing the most likely sequence of states.

Junction trees and message passing

The second method turns the model into a *junction tree* data structure and use *message passing* to do the inference. For a HMM this is similar to the *forward-backward algorithm* (Murphy, 2002). Determining the marginals of the system is called *sum-product message passing* also better known as *belief propagation*.

The general concept of *message passing* is that messages are collected from the outer nodes, also called the *leaves* of the probabilistic graphical model, and will propagate towards a node that is chosen to be the *root* of the model. After that messages are distributed again from the root back to the nodes that makes sure that every node in the model is affected by possible evidence anywhere in the system. This can also be seen in the example of a junction tree for a HMM in Figure 2.5 by first the normal arrows collecting evidence to the root at node X_2X_3 followed by the dotted arrows distributing the messages back to the leaves. A limitation of this method is that in case of loops message passing is not exact. Creating a *junction tree* from the probabilistic graphical model has the function to turn a model that contains loops into a tree structure.

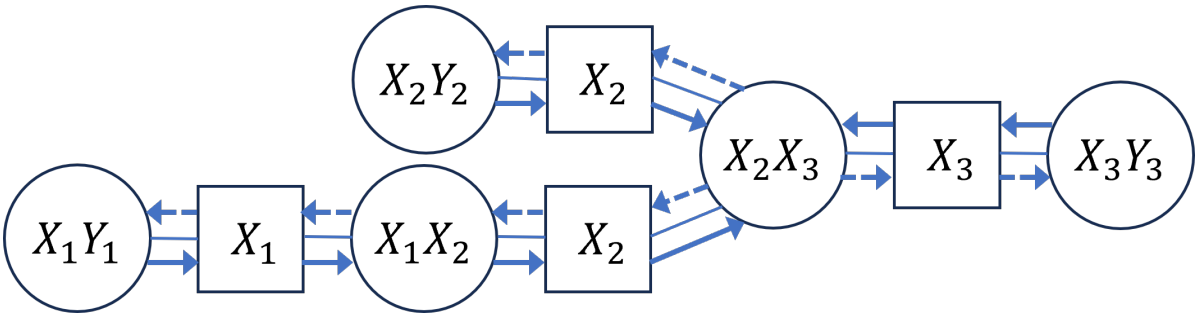
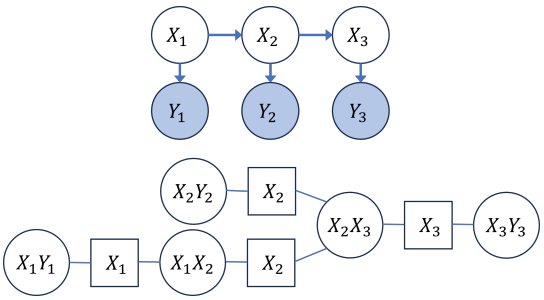
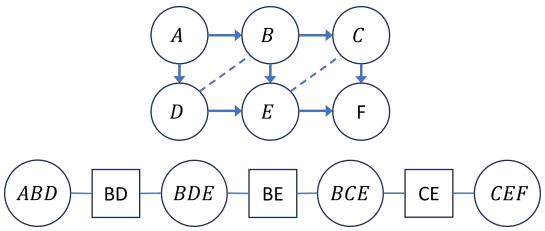


Figure 2.5: Example of message passing/belief propagation on the junction tree for a HMM. Messages are first passed by the normal arrows to the chosen root node X_2X_3 and after that the messages are distributed back by the dotted arrows.

The junction tree structure is that of a *factor graph* consisting of round nodes having sets of nodes of the original structure and square factor nodes representing connections between those sets of nodes. There are many methods to create a junction tree that optimization is important for the fastest computational speed. For this thesis the way to find the most optimized junction tree is of less importance, but for understanding of the algorithms for this process can be found explained in (Murphy, 2002). For a HMM and ARHMM examples of junction trees are shown in Figure 2.6. It can be seen that in the case of a HMM in Figure 2.6a the junction tree structure is less effective only combining sets of 2 nodes than in the case of a *Auto-regressive Hidden Markov Model* (ARHMM) in Figure 2.6b where sets of 3 nodes can be combined.



(a) HMM with a possible junction tree.



(b) ARHMM with a possible junction tree. Dotted lines are from the moralizing step to create the junction tree.

Figure 2.6: Junction tree examples

What a junction tree does is in fact encode the data structure before using belief propagation on it. For the message passing this has the consequence that the definition of the joint probability distribution of the system changes into equation 2.11 (Huang & Darwiche, 1996). The ϕx in this equation are the belief belief potentials of sets of parameters and the ϕs are the belief potentials of the factor nodes. For a more in depth mathematical explanation see (Murphy, 2002) and in (Huang & Darwiche, 1996) an elaborate example of message passing can be found.

$$P(U) = \frac{\prod_i \phi x_i}{\prod_j \phi s_j} \tag{2.11}$$

2.4.2. Approximate inference

In computer science the computational expense can be expressed in terms of the millennium problem P vs NP (Cook, 2000). Problems that are P can be solved in polynomial time being problems that if such a problem is not solvable at the moment the steady increase of computational power of computers over the years will eventually be able to solve it. Problems that are NP are not that easy to solve that increase in size generally means a fast increase in computational expense. Exact inference is #P-hard, which is harder than NP-hard making it necessary to approximate solutions using approximate inference (Murphy, 2002).

The most common method of *Approximate inference* is *Loopy belief propagation*, which applies belief propagation to the original graph even if it has loops in the structure. This gives a risk of double counting information and the solution not converging or converging in the wrong solution, but in practice is seen to often work well (Murphy, 2002).

2.4.3. Learning

Learning parameters is a step necessary for most state-space models as having no measurements of the hidden states also comes with limited knowledge about the dependencies between those hidden states. Two option discussed in (Murphy, 2002) are *gradient descent* and *Expectation Maximization*. The difference between these methods is that gradient descent when using maximum likelihood tries to minimize the gradient explicitly in contrary to EM trying to minimize the gradient implicitly. Again a more elaborate explanation on these methods can be found in (Murphy, 2002) and other sources.

3

Methodology

The goal of this thesis is to see if a probabilistic model can be made using Markov chains that estimate sediment pathways like SedTRAILS and if it does that in a faster way. To create a probabilistic model of the system the following questions had to be answered:

1. **Data:** Which and how much data is needed for the probabilistic model?
2. **Model components:** How to transform this data that it can be used in the probabilistic model?
3. **Model creation:** How to create the probabilistic model from the chosen data format?
4. **Validation:** How well does the model perform?

These questions resulted in the following framework for the model shown in figure 3.1. Most of these steps have been applied, but due to time constrains and some limitations in the Python packages some alternative to the learning step had to be chosen and instead of a validation a verification was done. There is great promise in the method that adapting the flow chart to these changes was omitted. Some mention will be made of how these steps could have been done, but those changes will be mostly discussed in Chapter 5.

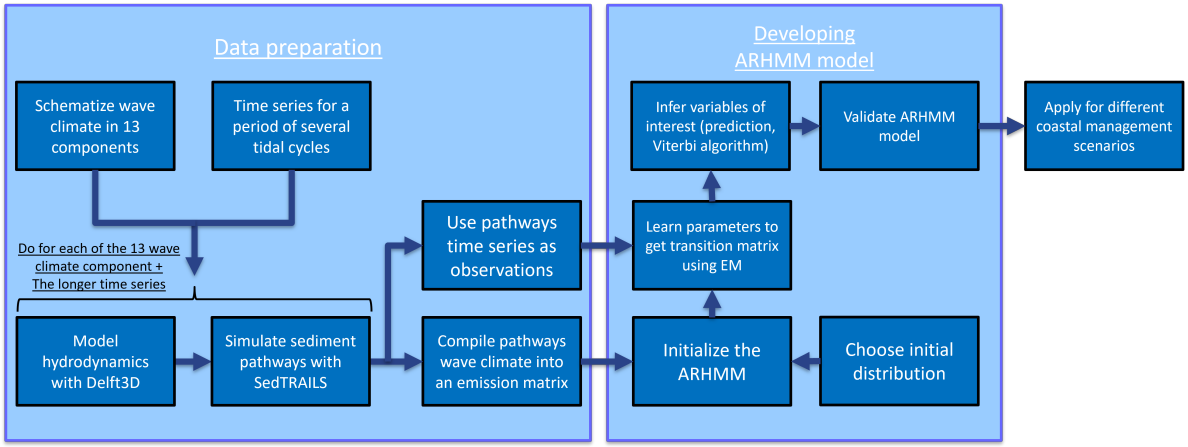


Figure 3.1: Flow chart for data preparation and developing the final model. Due to some limitations in the Python packages used, a distribution was chosen instead of the learning step and the validation was only done in a qualitative way instead of a quantitative way.

3.1. Schematizing the wave climate and data choice

For the first step data was needed for the model. What data was available depended on the chosen model area. The Ameland tidal inlet shown in figure 3.2 was chosen, which is one of the tidal inlets in the Netherlands connecting the Wadden Sea with the North Sea. This tidal inlet has been extensively studied as part of the SEAWAD and Kustgenese 2.0 projects (van Prooijen et al., 2020), which is why there was already a Delft3D model available that just had to be adapted for the different hydrodynamic conditions.

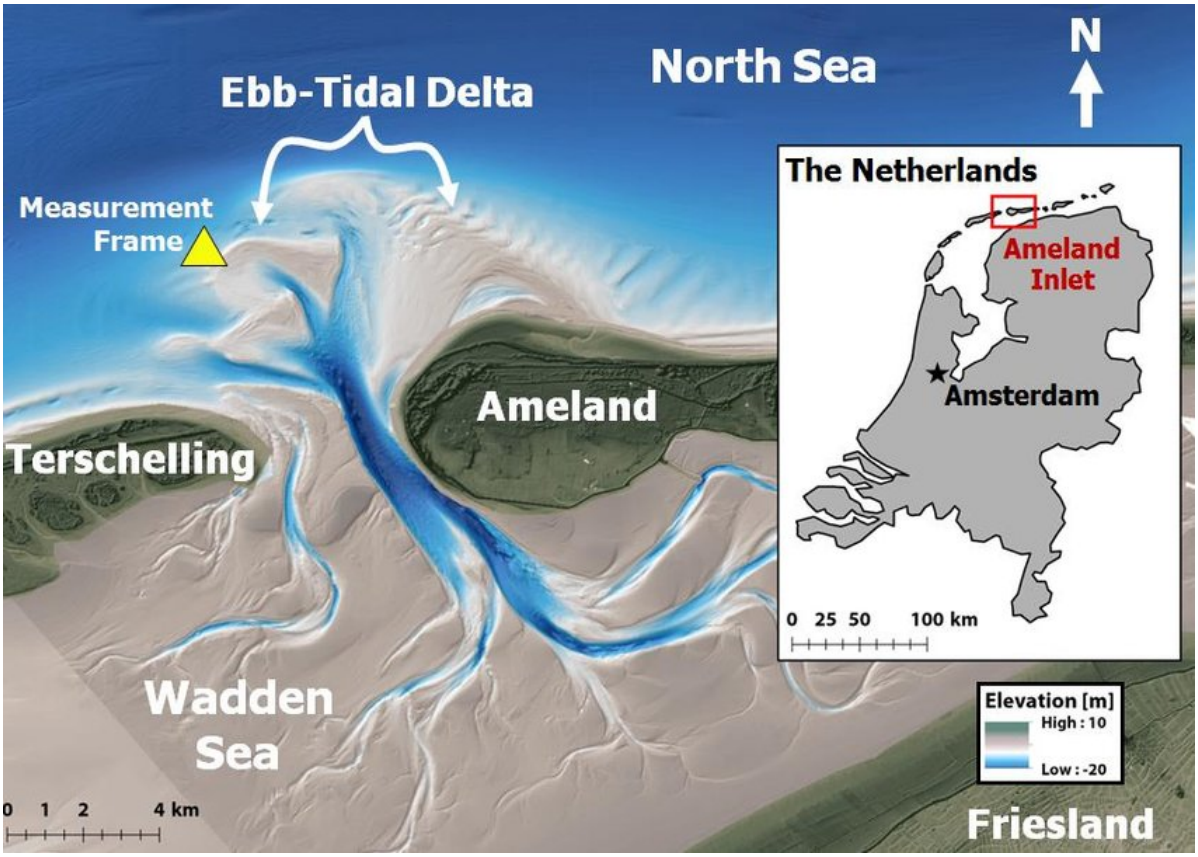


Figure 3.2: Site overview Ameland inlet. (Pearson et al., 2019)

Probabilistic models do not have descriptions of the physical processes, which is why to make this type of model we have to take into account all the different conditions the system can be in. The available software to compute sediment pathways SedTRAILS uses hydrodynamic data, which is why the hydrodynamic model Delft3D was used to first make a representation of the hydrodynamic conditions the system can be in before transforming these conditions to sediment pathways.

There is a lot of experience in describing the hydrodynamics an areas can be in for designing hydraulic structures like breakwaters and oil platforms. For this reason many methods already exist to describe the different hydrodynamic conditions an area can be in, which is done by determining the *wave climate* of an area. Like a normal climate, a wave climate describes the weather conditions over a long period of time, which is why it is based on very long time series from measurement stations. Processing so much data to do any calculations that need the wave climate would not be computationally feasible, which is why generally an *input-reduction technique* is used to make a *wave climate schematization*.

A good wave climate schematization describes the wave climate to the same effect as the big dataset does, but with less data. There are many ways to do a wave climate schematization (Antolínez et al., 2016; Benedet et al., 2016). For this thesis the wave climate schematization shown in figure 3.3 was chosen, that was already made for the Ameland tidal inlet area. This wave climate schematization considers the significant wave height ($Hm0$), the peak wave period (Tp), the wave direction (Dir), the setup, the wind velocity (U_{wind}), the wind direction (Dir_{wind}) and gives each wave condition a percentage weight. To create this wave climate schematization De Fockert (2008) used the data from measurement station Schiermonnikoog Noord (SON) over 1989-1999 and reduced 126 wave conditions to 12 using a morphological wave climate as described by Latteux (1995). The leftover 1.16% of this wave climate represents the cases of $H_s < 0.25m$, which is added as a 13^{th} wave condition without any waves. The method to create the morphological wave climate can be summarized as that the original wave conditions were iteratively reduced one by one, taking into account the morphological effect stays the same by redistributing the weights of the wave conditions with each step until the 12 conditions were left.

Wave condition [#]	Hm0 [m]	Tp [s]	Dir [°N]	Setup [m]	U_{wind} [m/s]	Dir_{wind} [°N]	Weight [%]
002	0.49	4.84	22.77	-0.13	4.65	88.04	17.58
009	0.52	4.87	292.97	0.03	5.08	224.09	11.35
020	0.99	5.06	264.07	0.10	9.17	215.45	13.21
024	0.99	6.29	322.68	-0.01	4.88	267.05	24.50
030	1.49	5.73	53.63	-0.38	9.60	80.80	8.22
051	1.98	7.13	338.01	0.03	6.99	332.86	6.67
052	1.98	6.98	351.21	-0.05	7.52	5.99	4.35
060	2.47	6.94	278.78	0.57	13.81	252.61	2.37
061	2.47	7.26	293.40	0.48	11.14	262.90	7.35
087	3.45	8.59	336.61	0.33	11.10	336.12	2.02
102	4.47	9.44	307.05	1.06	15.00	284.07	0.99
118	5.88	11.26	324.60	1.35	14.60	315.92	0.23

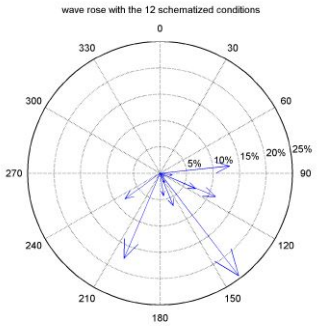


Figure 3.3: Wave climate data for Ameland inlet at Schiermonnikoog noord buoy (1989-1999) as from De Fockert (2008). On the left a table with the wave climate components, on the right a wave rose diagram with the different wave directions of the wave climate components

3.2. Data transformation

Having a wave climate schematization Delft3D and SedTRAILS were used to compute the sediment pathways data necessary for the probabilistic model. This was done by first running Delft3D for each of the 13 wave conditions described by the wave climate schematization. These conditions the model was run for a full tidal cycle of 24 hours and 50 minutes with also some extra time for spin-up effects of the model. Choosing that time frame corresponds with the most apparent pattern in the system, being the tide. The same period of a full tidal cycle was also chosen by De Fockert (2008) for determining the morphological wave climate. The second step the data from each run was used to do a SedTRAILS run giving the sediment pathways data as the (x,y) positions of each sediment particle populated into the model for each time step in the time period of a full tidal cycle. The details of both models will be discussed in the next subsections.

3.2.1. Delft3D

To model the hydrodynamics Delft3D version (4.04.01) from Deltares was used (Lesser et al., 2004). As described by the user manual: “Delft3D-FLOW is a multi-dimensional (2D or 3D) hydrodynamic (and transport) simulation program which calculates non-steady flow and transport phenomena that result from tidal and meteorological forcing on a rectilinear or a curvilinear, boundary fitted grid” (Deltares, 2023). The program does this by solving the Navier Stokes equations for an incompressible fluid, under shallow water and the Boussinesq assumptions when given initial conditions and boundary conditions. There are also several other equations included in the program of which we will just use the ones to include the effect of waves.

Waves in Delft3D are computed by coupling SWAN, a third generation wave model (Ris, 1997), with the flow model of Delft3D. The model uses a different grid for the wave and flow simulation, allowing the wave grid to be less fine than the flow grid. In figure 3.4 this can be seen in that the red grid for the waves shows a mesh size (highest resolution of 500 m), in comparison to the green grid representing the flow grid (highest resolution of 60 m).

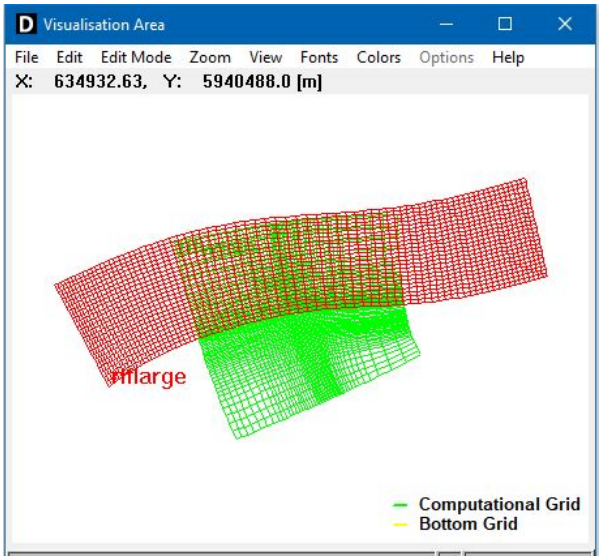


Figure 3.4: Grids for the flow and wave part of the model

Together with the grid the bathymetry is also necessary for the model, which was constructed from interpolation of bathymetry measurements from 1989. This seems a bit outdated, but is not so relevant for proving the concepts discussed in this thesis. Boundary conditions consist of a harmonic boundary condition in the north and Neumann boundary conditions in the east and west. The bathymetry and positions of the boundary conditions for the model can be seen in figure 3.5. Other settings involve an uniform Manning roughness being used with a value of 0.026,

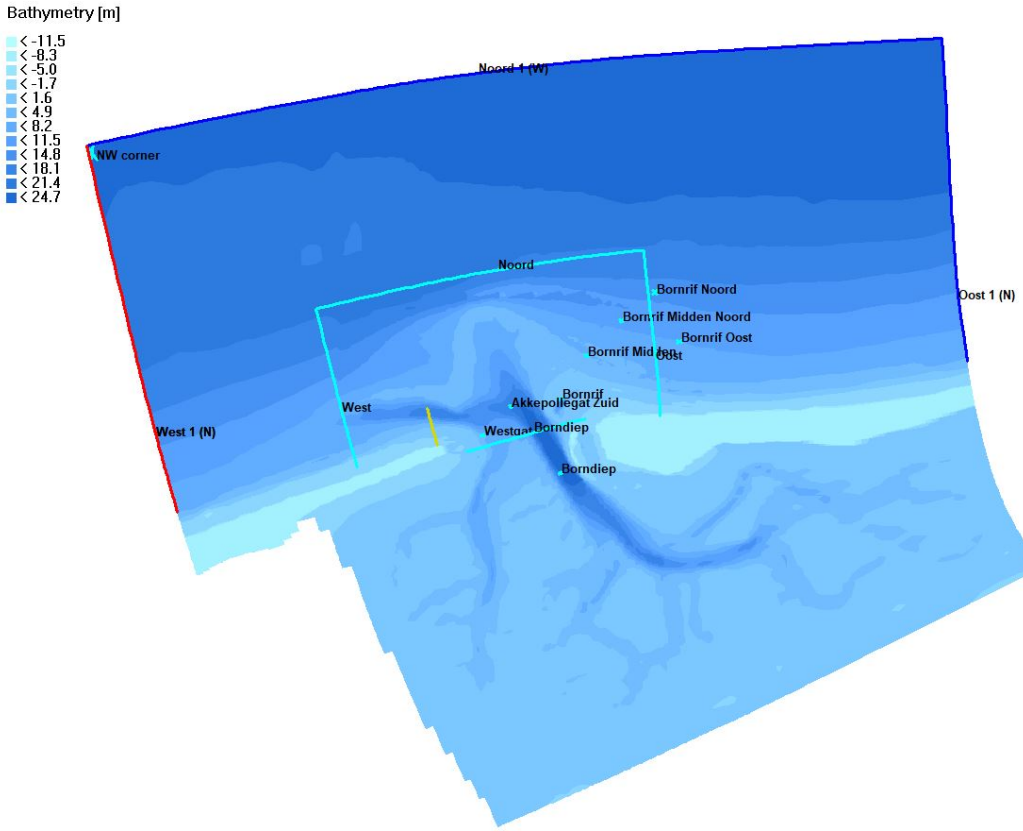


Figure 3.5: Visualization bathymetry of the flow model for the Ameland tidal inlet area. Also including the positions of the boundary conditions of the model on the north, east and west border of the model.

The time frame for the model is set at 4 days duration with a time interval of 1 minute. The output of the model is every 10 minutes. The model was run for all the different wave climate components of De Fockert (2008) wave climate for a time duration of 4 days for spin-up effects. Actual run time was for each run between 20 and 30 minutes. For the next step SedTRAILS only uses a full tidal cycle of data and output for the depth averaged velocity, bed shear stress, maximum bed shear stress and grid cell surface area.

3.2.2. SedTRAILS

To model the sediment pathways the *Lagrangian modelling* software SedTRAILS was used (Pearson et al., 2023). Most coastal sediment transport models use an Eulerian approach that describe fluxes going in and out of a control volume. A Lagrangian approach is different in that the frame of reference moves with the individual particles, taking into account the forces acting on the particle. This gives the full trajectory of particles in the system.

The Lagrangian approach of SedTRAILS uses equation 3.1 from Soulsby et al. (2011) to advect particles through the system. This equation calculates the mean speed of a mobile grain during a time step U_{gr} by multiplying the current speed averaged over the lowest 1 m of the water column U_c with 3 different factors, which can vary between 0 and 1. Consequently U_{gr} can never be bigger than U_c. The first factor F stands for the freedom factor describing if a particle is trapped or free to move. The second factor P decides the probability that a current working on the sediment particle can actually move it. This is done by an expression involving the Shields parameter. The third factor R is a reduction factor on u_c as once a particle is in motion, the mode of transport (i.e., bedload vs suspended load) depends on how far it is above the threshold of motion.

$$U_{gr} = F \cdot P \cdot R \cdot U_c \tag{3.1}$$

where:

F = freedom factor
P = probability a free grain is moving
R = reduction factor for the speed of a mobile grain compared with U_c
U_c = current speed averaged over the lowest 1 m of the water column

SedTRAILS computes sediment particle pathways in a few steps. In the first step the pre-processing is done by loading the instantaneous flow velocity field, the bed level and the bed shear stress due to waves and currents from the Delft3D data to calculate from those the sediment velocity as described above using Soulsby et al. (2011). Some extra settings that have to be included for calculating the sediment velocities are the sediment size and the initial freedom factor. SedTRAILS is not a mixed sediment model in that it just computes the sediment pathways for one grain size. The sediment size chosen for the model is 400 μm, which is relatively coarse sand for Ameland (Pearson, 2022) but in our case favorable to to minimize the amount of particles leaving the areas that will be defined for the connectivity network. The initial freedom factor was chosen to be 1, which means all particles populated in the model are on the surface free to move.

In the second step the area is divided into 5 areas to create the clusters with centroids to be able to make a connectivity network. These polygons were chosen that area 1, 2 and 3 consist of the ebb-tidal delta and the coastal areas east and west of it and area 4 and 5 consist of the two main channels with corresponding tidal flats in the basin. This distinction between areas is made to separate possible sinks in the system at a coarse scale. The polygon file to create these areas can just be made using Delft3D. Using the centroids and cluster boundaries each area is filled with sources at a resolution of $500m$ from which particles can be released at any amount and any interval. These settings were set to release only one particle at the start of the run. Figure 3.6 shows the settings so far already create the initial state of the system. It can be seen that the amount of sources is limited as a finer grid of sources would limit the ability for visual interpretation of the system after advection.

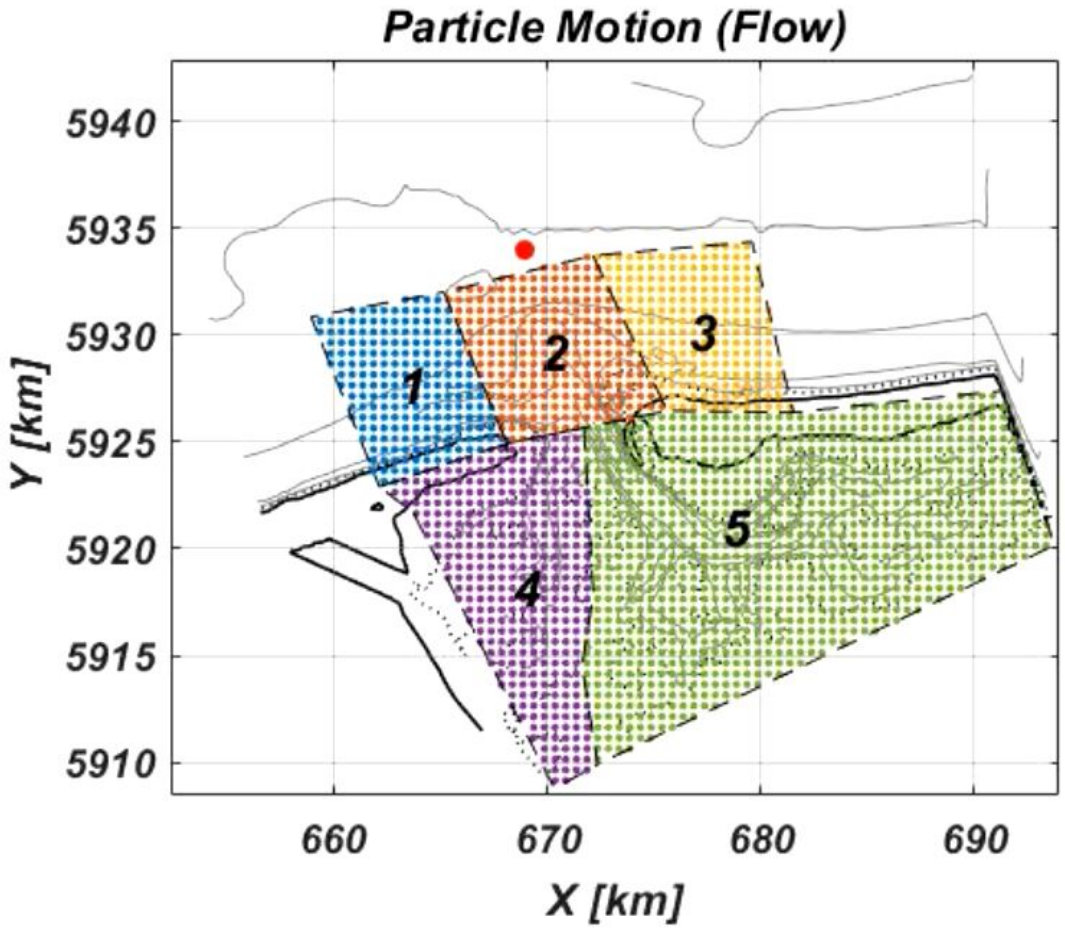


Figure 3.6: SedTRAILS visualization of the Ameland tidal inlet area showing the system divided into 5 areas using polygons. Each area shows the initial distribution before advection being populated with particles for a resolution of $500m$ (Pearson et al., 2023). The red dot north of area 2 corresponds to a measurement point used for later output for showing the current velocity over the tidal cycle.

The third step the sediment particles are advected using the sediment velocities computed in the first step. The data obtained are time series of the (x,y) positions of each particles released by each source.

In the fourth step this data was used to visualize how the system behaves. As said before these sediment pathways were modeled for each of the 13 components of the wave climate schematization. In figures 3.7 and 3.8 for a very calm condition with $H_s = 0.49m$ and a very energetic condition with $H_s = 5.88m$ the output for step four are shown at the end of the tidal cycle. Explaining all the output: left top visually shows the position of each particle in the system, right top shows the system described as a connectivity network left bottom a time series shows the current velocity at a measurement point north of area 2 and last right bottom shows an adjacency matrix of the system. Both the connectivity network and adjacency matrix show the strength of those transitions relative to the other components. Something that is already very noticeable is that the adjacency matrices of both figures are very similar suggesting that the sediment particles for both cases move in the same way, but the amount of movement is the main difference.

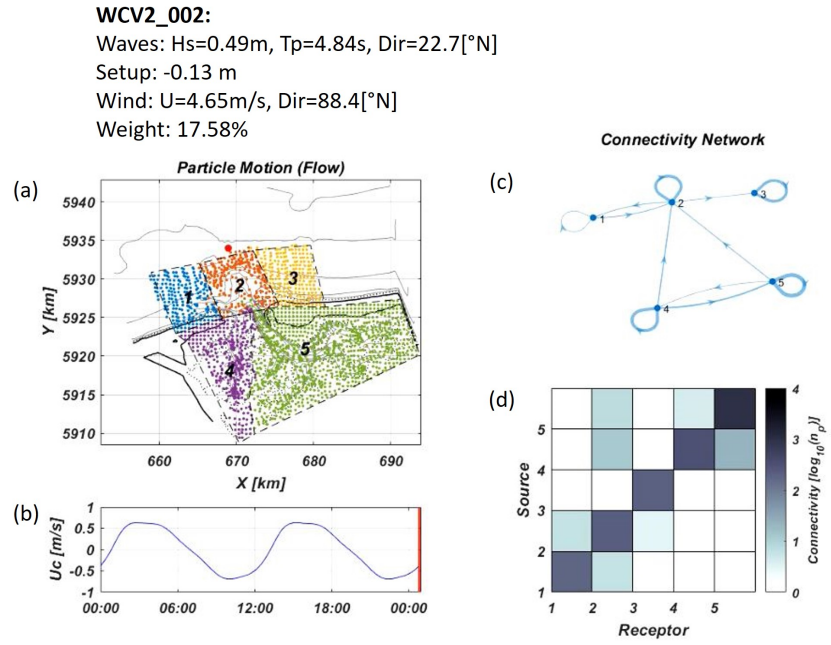


Figure 3.7: SedTRAILS output wave condition 002 after a full tidal cyle of 24 hours and 50 minutes. (a) Shows the position of each sediment particle, (b) shows the current velocity signal at the measurement point north of cell 2, (c) shows the connectivity network and (d) shows the adjacency matrix. (Pearson et al., 2023)

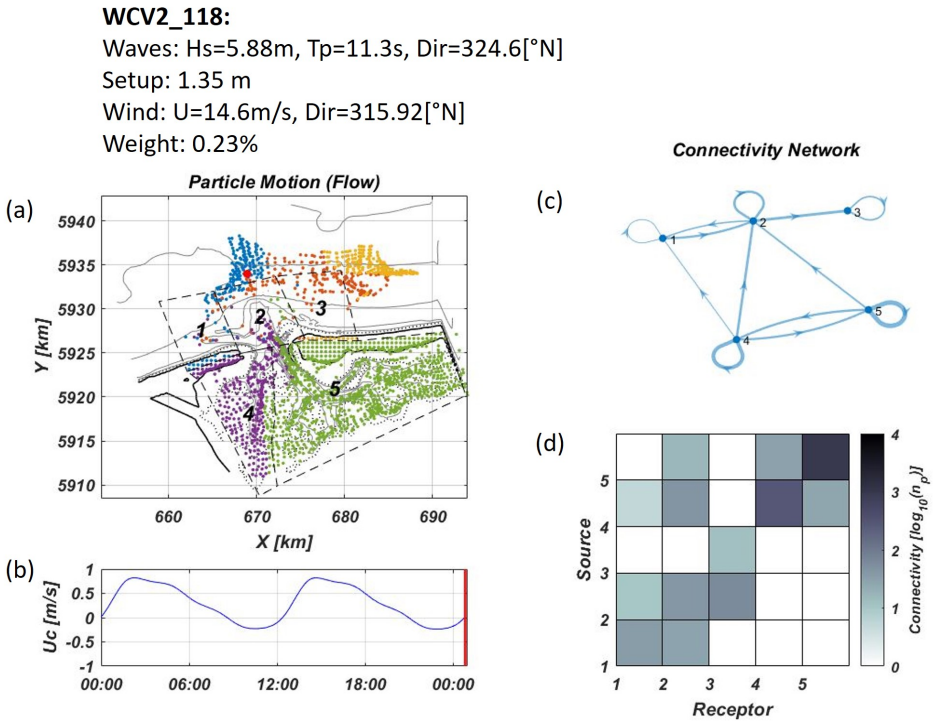


Figure 3.8: SedTRAILS output wave condition 118 after a full tidal cyle of 24 hours and 50 minutes. (a) Shows the position of each sediment particle, (b) shows the current velocity signal at the measurement point north of cell 2, (c) shows the connectivity network and (d) shows the adjacency matrix. (Pearson et al., 2023)

3.3. Developing the probabilistic model

Having the sediment pathways data the next step was to develop a probabilistic model. This process consisted of first creating a *transition matrix* of the full wave climate by making transition matrices of each wave climate component and summing those components multiplied by their weights as described by the wave climate, which we will call *the Markov model*. To get a more advanced model a *Dynamic Bayesian Network* (DBN) model was developed to be able to add covariates to the model. A different approach to add covariates to the Markov model was considered by using a regression model (Ataharul Islam & Chowdhury, 2006), but the DBN approach was chosen as it was considered more flexible. The DBN model will be explained in three sections. First the structure of the model will be explained, which links the SedTRAILS model to the probabilistic model and classifies the DBN as an *Auto-regressive Hidden Markov Model* (ARHMM). The second section will explain the programming steps done and how the data is used to create this type of model. The last section due to time constrains instead of a validation a verification of the DBN model is described. Also some brief mention will be made about how this approach is different from the learning and validation step originally intended, but will be discussed in more detail in Chapter 5.

3.3.1. Transition matrices

The connectivity description of the SedTRAILS data already gave a matrix that looks like a Markov chain. This matrix was called an *adjacency matrix*, which counts the amount of particles from a source in a receptor and takes the log of that amount to scale the axis. A *Transition matrix* is different in that it describes the probability of transfer from a source to the different receptors. To get a transmission matrix each row must only contain probabilities and the sum of each row must equal one. Making a transition matrix from an adjacency matrix was done by first creating a version without the log and storing the sum of each row per time step. Next the same process for the adjacency matrix is done again, but every value in the matrix is divided by the total of the row computed earlier to make probabilities from these values. This method assures that for any point in time we want to obtain the transition matrix, each row will always sum up to one. Using the initial amount of sources per area for the totals would be incorrect as particles outside the defined areas do not have a receptor to be counted in.

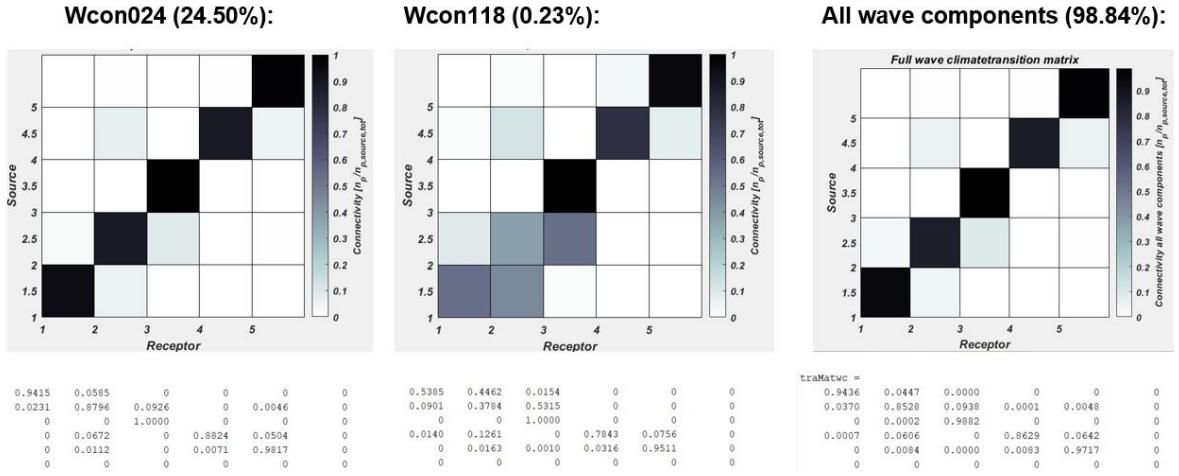


Figure 3.9: Transition matrices for wave climate components 024 and 118 next to the transition matrix representing the sum of wave climate components time their weights (excluding the no waves condition)

For each wave climate component a transition matrix was made and using the weight of each component these can be combined to a single transition matrix describing the system as a Markov chain. In figure 3.9 the transition matrices of wave climate components 024 and 118 can be seen and an early version of the resulting transition matrix that did not yet include the no waves condition. The final transition matrix is what we will call here the *Markov model*. Having an initial distribution of the sediment particles over the areas this matrix can be applied repeatedly to get the sediment distribution at the next time steps. This Markov model does give some information about how the system behaves, but as a limitation this method spreads out extreme components over a long period instead of having it occur just with a low probability.

3.3.2. Model structure

To improve the flexibility and applicability of the model we want to have some of the dependencies in the model be put into separate random variables. This is also in line with the goal of the probabilistic model to have variables that are known to be of stochastic nature to be added directly into the model. For this reason the next step was looking how to add *covariates* to the model. Covariates are extra independent variables that are expected to influence the parameter of interest (Everitt, 2006). It was for example thought of that the wind direction could be an interesting covariate as that may influence our variable of interest (the sediment position in the system) over time.

To look at the model in a probabilistic way it can be noticed in figure 3.10 that the population of each area can be seen as a random variable. The transition probabilities as derived in the previous section are determined by where does that population end up after a chosen time period. This means the probabilities do not only depend on how these particles are advected through the system, but also the size of the population and the shape of the cell in how it positions in the system.

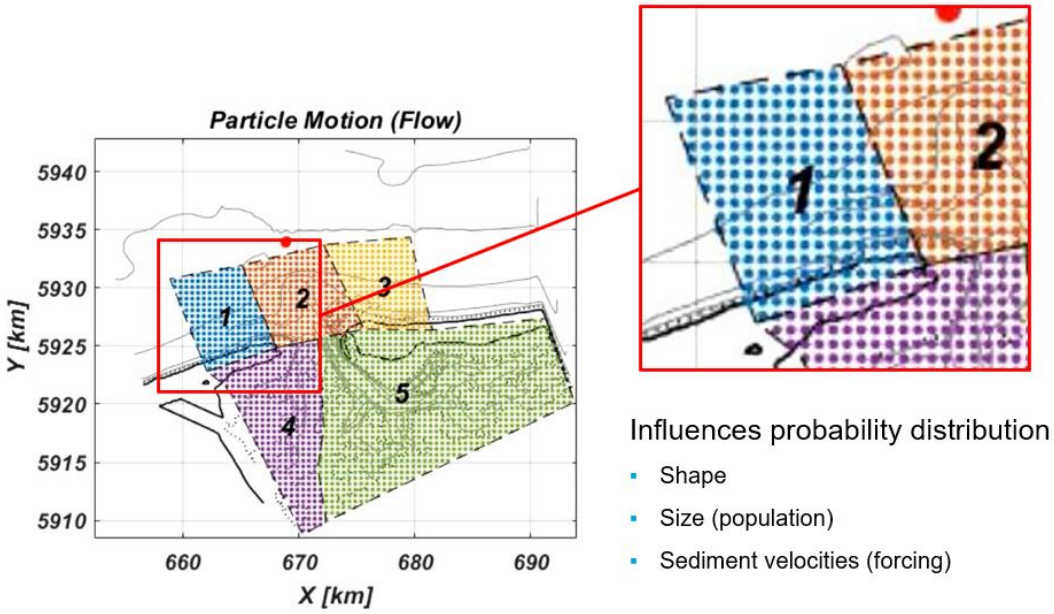


Figure 3.10: Initial position of sediment particles in the system showing the probability distribution of a row in the transition matrix is influenced by the shape and size of the cell together with the forcing working on the sediment particles. (Pearson et al., 2023)

Using this description that the particles of each source area are random variables this can graphically be interpreted as in figure 3.11. This representation of the system corresponds to a DBN for which we can add covariates to the model.

Dynamic Bayesian Network (DBN) models are *probabilistic graphical models* that describe the dependencies between random variables within a system (Murphy, 2002). This is done by describing the system using *graph theory* in terms of *nodes* (also called *vertices*) connected by *edges* (Newman, 2003). The nodes describe the probabilities of variables in the system and the edges describe conditional probabilities. DBN are different from BN in the sense that every time step has random variables, making the networks semi-infinite collections of random variables. Probabilities of interest can be obtained from the model using *inference*. More on the basics of these models have also been discussed in Chapter 2.

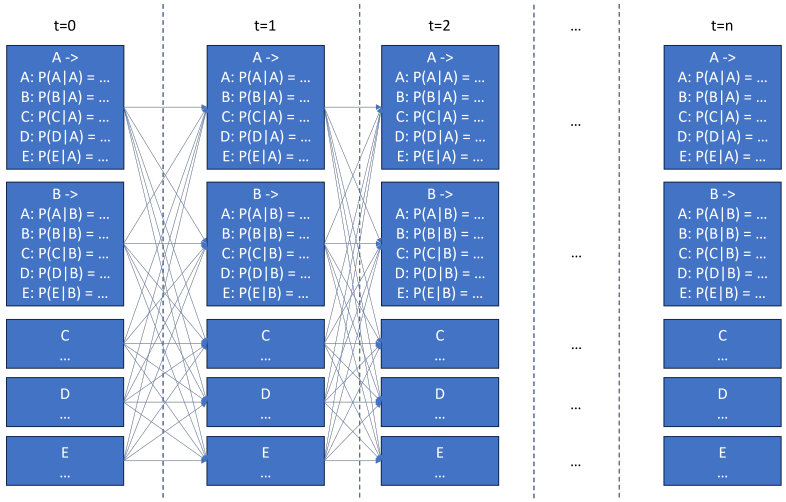


Figure 3.11: First concept of the system visualized as a DBN. In this figure A, B, C, D, E correspond to the areas $1-5$ in Figure 3.10. Each time step (t), a probability distribution for each parameter determines where a particle goes the next time step.

A common structure for DBNs is that the observations of the system are caused by some hidden states in the system. In that case the system is classified as a *State-Space Model* (SSM). DBN like that are extensively described in Murphy (2002), which also mentions that some of the most common types of SSMs are *Hidden Markov Model* (HMM) and *Kalman Filter Model* (KFM). Of these two options the HMM was a good fit in the sense that the hidden states work very well for our system to allow us to also estimate in what sequence the wave climate components follow up each other.

To make the connection of the model to the data, it was noticed that the transition matrix of each wave climate component was available in our approach. This had the advantage that to create our *emission matrix* we can just combine all those transition matrices in an *Auto-regressive Hidden Markov Model* (ARHMM) structure. This model is the same as an HMM, but has the extra link that each observation node in the model is also dependent on a prior observation. Murphy (2002) mentions this being a good quality as this often results in models with higher likelihood.

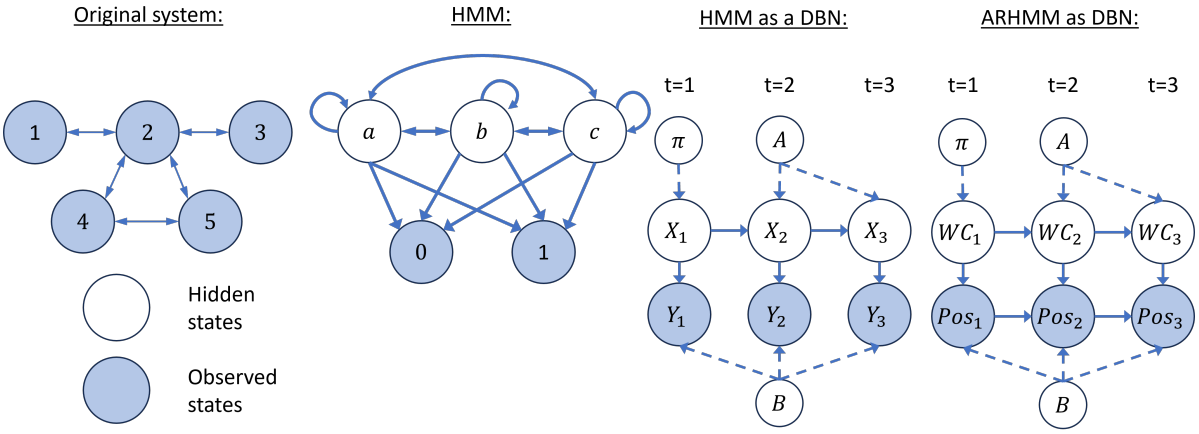


Figure 3.12: Conceptual steps to get from the system as connectivity network to an ARHMM. The first graph shows the system as a connectivity network each node representing an area in the system. The second graph shows we want to describe the system as a HMM to include a covariate as hidden variable in the system. A distinction is made between the states (a,b,c) of the hidden variable and states (0 and 1) of the observed variable. The third graph shows the HMM can be described as a DBN. Both X and Y being random variables having a distribution that can be inferred at each time step. The link is made that the wave climate would be good as hidden variable. In the last graph the link to the data is made that the system can be described as a DBN structured as an ARHMM. In the ARHMM the discrete random variables WC and Pos stand for the wave climate of the system and the position of the sediment in the system.

In figure 3.12 the different conceptual steps that were done so far to get to the ARHMM structure are shown. The abbreviation Pos_t stands for position of a particle in the system (area 1,2,3,4 or 5) at time step t and the abbreviation WC_t stands for the wave climate component $(0, 1, ..., 11, 12)$ at time step t. The data sizes for the ARHMM can be seen in Figure 3.13. The transition matrix for the wave climate is of size (13x13). The emission matrix describes for the ARHMM the conditional probabilities $P(Pos_t|WC_t, Pos_{t-1})$, which has a data size of (12x5x5). These can also be found in Appendix A. The link between WC_1 and Pos_1 was removed to make sure the different amount of particles seeded in the system by SedTRAILS could be used as initial distribution for the verification step. The initial distribution for the wave climate was set to be the weight distribution that was shown in Figure 3.3, but also some experimentation was done by using a distribution sampled from a Poisson distribution with $\lambda = 4$. The value 4 was chosen as it seemed to give a similar shape of the distribution as the mentioned weight distribution. These distributions were also used to create the transition matrix for the wave climate components. This was an alternative that was necessary for omitting the learning step due to time constrains.

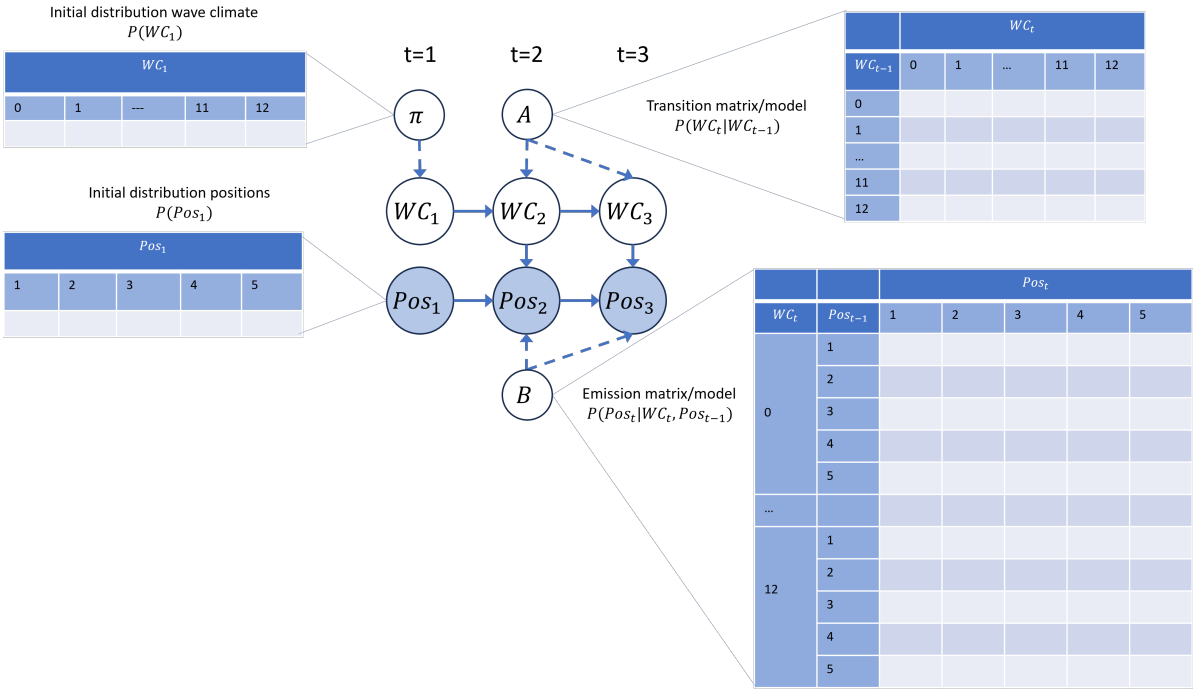


Figure 3.13: The ARHMM showing the data formats for the parts of the model. In the ARHMM the discrete random variables WC and Pos stand for the Wave Climate of the system and the Position of the sediment in the system.

3.3.3. Programming the ARHMM

To program the ARHMM some choices had to be made for the software. Programming the model from scratch was not feasible for this Master thesis in terms of time and some software has already been in development over the past years to make such models in more efficient ways than the writer could do. It was chosen to use a Python package for the programming as that was the easiest to find documentation for.

The package pyAgrum (Gonzales et al., 2017) was chosen which is mostly designed for BNs, but does include a function to model DBNs. Especially for the visualization of the network it was very useful for modelling our ARHMM. Another option that was very promising was the package DYNAMAX (Linderman, 2024), which is a specific Python library for *State-Space Model*s. This package includes some newer techniques favorable to speed up *Machine Learning* computations and Murphy is one of the contributors, but the package is still in development. Applying it for our ARHMM did not seem possible in that this structure did not seem implemented for a discrete emission matrix.

Having a structure for the model as in Figure 3.13 each table has to be filled in. The emission matrix can just be implemented. The exact definitions of the initial distributions are implemented as described in the previous section. The transition matrix that describes how the hidden states (representing the

wave climate components) of the model follow each other up are preferably acquired using a learning technique given data of the observed nodes. Learning techniques compute the system for different values for the parameters and choose the parameters that give the highest likelihood of the system.

The python package that was chosen applies inference using an algorithm called *lazy propagation*. Explained in (Madsen, 2006) this is a hybrid inference algorithm combining the methods of variable elimination and message passing for better computational performance than traditional indirect-computation algorithms. Also the package include the option to do parameter learning using *Expectation Maximization* (EM). Applying this step to the model did not give the expected results that they are only discussed in Chapter 5. Instead independence of wave climate components following each other up was assumed resulting in each row of the transition matrix being the same distribution as the initial distribution.

3.3.4. Verification

To validate the model it would be preferred to obtain sediment pathways data from measurements and split that into a learning and testing dataset. The transition matrix in that case could be learned from the learning set and an estimate from that data could be tested using the testing dataset.

As the learning step was omitted a verification was done to still compare estimates of the model to the results of a SedTRAILS run. The ARHMM model estimates in steps of 24 hours, which is why to compare both outputs a time series of at least days is necessary. SedTRAILS assumes the system is morphostatic, which is why it also does not make sense to choose a very long time period. This also favors the fact that long time series are more computationally expensive. For this reason a simulation period of 1 week using different versions of the probabilistic model was compared with a run of SedTRAILS using 1 week of data. A month run would have also been interesting, but was omitted due to time constrains.

Time series of the wave climate parameters were found at the Rijkswaterstaat (RWS) site for waterinfo (Rijkswaterstaat, 2024). The data at measurement stations Terschelling Noord and Schiermonnikoog Noord were chosen. These were the significant waveheight ($Hm0$), peak wave period ($Tm02$), the wave direction, the setup, the wind direction and the wind velocity. There was no data for the width of the energy distribution, which is why just a value of 10 was chosen. Originally intending to do a month run the period of 01-03-2024 to 01-04-2024 was chosen, which for the week run only extends to 09-03-2024. An extra day for the week run was for the effect of spin-up in Delft3D. It would have been preferred to choose a month with more extreme conditions in the data, but there was only very recent wind data available that it does not contain any data before January 2024. This data was used to create a wavecon file for Delft3D for changing conditions every hour. The results of this run were put into SedTRAILS, which were used to compare the other runs in Chapter 4.

4

Results

As was described in Chapter 3, Delft3D was used to model the hydrodynamics of the area for the 13 hydrodynamic conditions from the wave climate schematization. This wave climate was used to compute the sediment pathways for each of the wave climate component using SedTRAILS. Each wave climate component a transition matrix was made that describes for each source the probabilities to be in any of the cells after the given time period. As a first approach, these components were just summed up using the weights of the wave climate schematization to create a transition matrix representing the effect of the full wave climate, which we call the *Markov model*. For the second approach the system was modelled as a *Dynamic Bayesian Network* (DBN). The structure of the DBN model is that of an *Auto-regressive Hidden Markov Model* (ARHMM) in which the observed states are the sediment positions in time and the latent states describe the follow up of wave climate components in time. The transition matrices of the wave climate components in this model were used to create the *emission matrix*. The other parts necessary for the model were a transition matrix describing the follow-up of wave climate components and a choice for the initial distribution of the system. The transition matrix is first considered independent of the prior wave climate making each row the same distribution. Using this assumption the Poisson distribution and a distribution from the weights of the De Fockert (2008) wave climate schematization were used. Another method to determine the distribution for the wave climate is by using a training dataset and learn the distribution using *Expectation Maximization* (EM). The initial distribution is assumed to just be the weight distribution from the wave climate schematization.

4.1. Markov model

The first version of the model is the transition matrix on the right side in Figure 4.1, which was the sum of the transition matrices of each wave climate component multiplied by their weight of occurrence. This transition matrix estimates the probabilities of where a sediment particle at time t can end up at time $t + dt$. The transition matrix was made such that dt equals a period of one tidal cycle of 24 hours and 50 minutes. One of the first questions is whether the full tidal cycle of two high and low waters is needed, or if just half a tidal cycle with one high and low water would be sufficient?

	post				
pos0	1	2	3	4	5
1	0.9982	0.0018	0.0000	0.0000	0.0000
2	0.1029	0.8608	0.0338	0.0000	0.0024
3	0.0000	0.0059	0.9940	0.0000	0.0001
4	0.0000	0.0597	0.0000	0.8821	0.0582
5	0.0000	0.0092	0.0000	0.0054	0.9853

Half a tidal cycle of 12h 25 min

	post				
pos0	1	2	3	4	5
1	0.9551	0.0449	0.0000	0.0000	0.0000
2	0.0372	0.8640	0.0939	0.0001	0.0048
3	0.0000	0.0003	0.9997	0.0000	0.0000
4	0.0007	0.0608	0.0000	0.8738	0.0647
5	0.0000	0.0086	0.0000	0.0084	0.9830

Full tidal cycle of 24h 50 min

Figure 4.1: Transition matrices for half and a full tidal cycle. Pos0 is the source area of a particle and post is the receptor area of a particle after one time step. Transitions between the two is described in terms of probability that each row sums up to a probability of 1. Transition matrix was visualized using pyAgrum package.

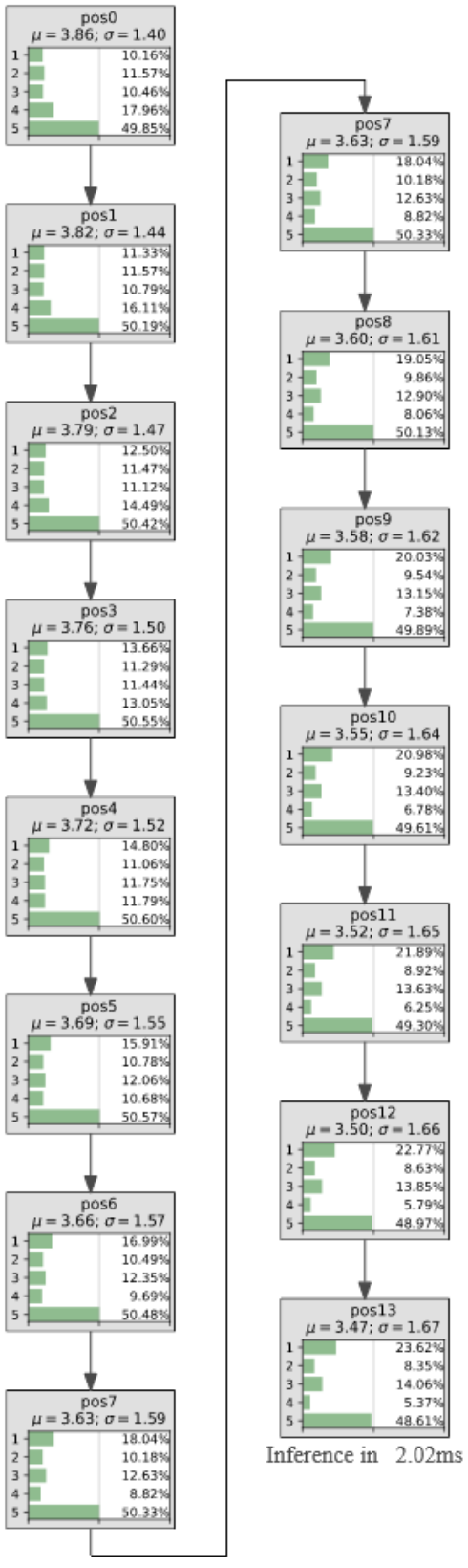
On the left side in Figure 4.1 the transition matrix for half a tidal cycle is shown. To compare both matrices for the same period the left transition matrix has to be applied twice for each time the right transition matrix is applied. This results in that when using half a tidal cycle sediment is more conserved in area 1 and less conserved in areas 2 and 3 than when using a full tidal cycle.

To see the full effect both transition matrices were used to estimate the sediment distribution after a simulation period of one week. To do this the model has to be *unrolled* into individual time slices such that both transition matrices are repeatedly applied to get a simulation period of one week. In Figure 4.2 this can be seen by having the model using full tidal cycles be unrolled into 8 time slices and the model using half tidal cycles be unrolled into 15 time slices.

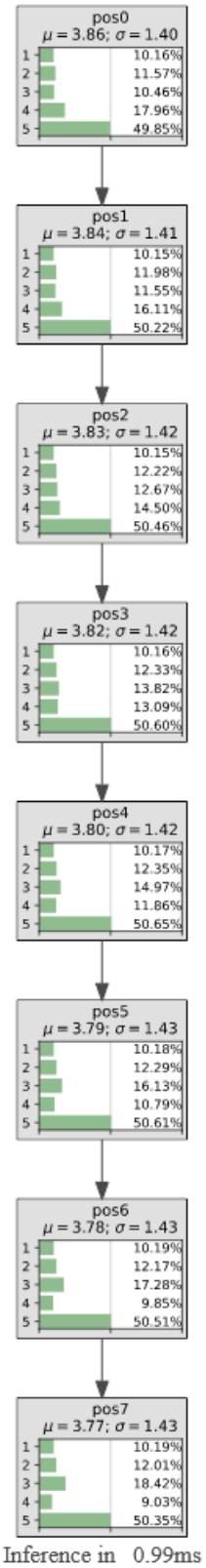
Figure 4.2: Structure of the model when applying the Markov model to estimate sediment positions after one week. Each step representing a full tidal cycle of 24 hours and 50 minutes the transition matrix is applied to estimate the distribution of sediment on the next time step. Unrolled version of the Markov model was visualized using pyAgrum package.

Applying inference by using Lazy Propagation the marginal probabilities in the system for both options were computed in Figure 4.3. It can be seen that both methods start with the same sediment distribution, which was defined using how each area in the SedTRAILS model got seeded with a number of particles at the initial condition. Each time step the effects of the transition matrices accumulates until the last time step, which will mainly be used to compare the methods.

A big difference that can be seen is that when using half a tidal cycle for the transition matrix, the probability of sediment particles being in area 1 increases from the original 10.16% to 23.62%. For the case using a full tidal cycle it increases to 10.19%. It would be expected that this area without the influx of sediment from the longshore current would reduce over time. This makes the full tidal cycle case better, but still not very accurate.



(a) Half a tidal cycle per time step



(b) Full tidal cycle per time step

Figure 4.3: Inference of all the marginal distributions of the Markov model comparing using half a tidal cycle as time step with a full tidal cycle as time step for one week of simulation time. Each variable: "pos" stands for the position of particles in the system as a discrete random variable with state space [1,2,3,4,5] representing the different areas. The number added to each variable stands for the time step. For each computed marginal distribution in the system also the mean μ and standard deviation σ is shown. Pos7 was shown twice in (a) to make comparision with pos8 easier.

4.2. SedTRAILS reference run

To get a better idea how realistic these results are Rijkswaterstaat (RWS) data from 02-03-2024 to 09-03-2024 was used for a wavecon file in Delft3D to simulate the hydrodynamics and run this way a week simulation time of the system in SedTRAILS. The resulting state of the system after one week is shown in Figure 4.4. The begin and ending of the run were also used to quantify the system by making a transition matrix for this time period shown in Figure 4.5.

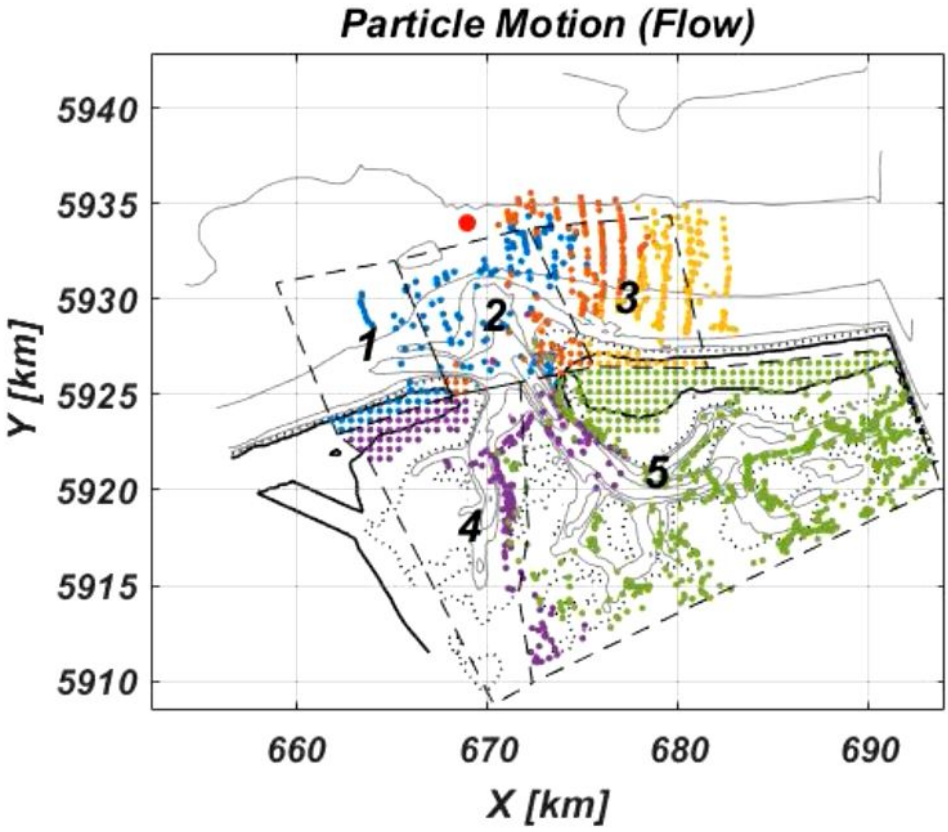


Figure 4.4: SedTRAILS (Pearson et al., 2023) end positions of the particles for a run using RWS data from 02-03-2024 t0 09-03-2024.

It can be seen that outside the basin sediment supplied by wave-driven *longshore transport* at the updrift coast in area 1 would have to bypass the tide-dominated tidal inlet coast to reach the downdrift coast at area 3. Part of the sediment bypasses the ebb-tidal delta in area 2 by following the outer border to reach area 3. The sediment is not caught in the tidal motion of the ebb-channel. This bypassing mechanism was also called *ebb-delta periphery bypassing* in the introduction. Another part of the sediment is pulled into tidal inlet by marginal flood channel at the transition between areas 2 and 4 and ends up in the main ebb channel between areas 2 and 5. From Figure 4.5 it was seen not much of this sediment ends up in area 5 though. Most sediment is expected to flow with the ebb current through the channel to a more seawards point of the ebb-tidal delta. From there a similar motion as for the other bypassing mechanism causes the sediment to reach the downdrift coast at the right corner of area 2 or traverse into area 3. This bypassing mechanism was called *flow bypassing* in the introduction.

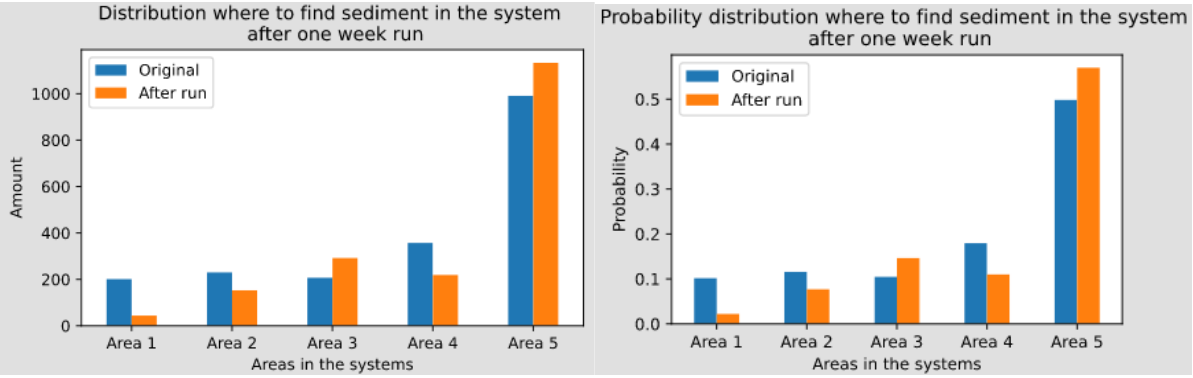
Inside the basin currents coming into the basin by the flood channel and leaving through the main ebb channel at area 5 causes sediment to transport from area 4 to area 5. This seems to be a lot of the sediment that ends up in the main ebb channel. In Figure 4.5 it can be seen that some of the sediment from area 4 also ends up in area 2. Considering sediment of a lot of areas goes through main ebb-channel there is surprisingly little exchange between inside and outside the basin. In (Wang et al., 2018) this can be explained by the tidal basin being *accommodation limited* that the net import is expected to be minimal. Summing up the sediment in and outside the basin also confirms this by only finding a little increase of the sediment inside the basin. Area 5 there is not much change that sediment just seems

to focus more into the channels of this part of the basin. The model has a closed boundary that some expected sediment that would want to leave the domain is prevented from doing that.

Figure 4.5 shows that a bit more than 40% of the sediment from area 4 is transported to area 5, which was seen mostly ending up in the main ebb channel. It may be wondered why area 4 would not quickly lose a lot of sediment over time as it does not have the supply of sediment from longshore transport area 1 has. This could be explained by the assumed morphostatic situation. The system works like a sediment sharing system that if there is a lack of sediment somewhere other parts will quickly supply to aim for some equilibrium (Elias et al., 2019). The bathymetry does not change here that adaptations of the system would not be seen. Another explanation could be that this is a way of sediment sorting. Away from tidal inlets decreasing sizes of sediment are found for the Wadden Sea (Wang et al., 2018). Transport patterns for finer sediments are expected to be different that it could just be the chosen medium sand of $400\mu m$ for the model that just leaves 4. Another effect that may explain why area 4 is not emptying out is wind-generated residual flow (van Weerdenburg et al., 2021). A lot of the sediment from area 4 that does end up in the main ebb channel seems to pile up at the downdrift coast side. Accretion of this sediment could be related to the familiar drumstick shape of barrier islands, which is seen here for Ameland.

	post				
pos0	**1**	**2**	**3**	**4**	**5**
1	0.2316	0.5368	0.2158	0.0000	0.0158
2	0.0000	0.2391	0.7065	0.0000	0.0543
3	0.0000	0.0000	1.0000	0.0000	0.0000
4	0.0000	0.0140	0.0000	0.5714	0.4146
5	0.0000	0.0020	0.0000	0.0152	0.9828

Figure 4.5: Transition matrix using results of 7 tidal cycles of simulation time representing one week. Pos0 is the source area of a particle and post is the receptor area of a particle after 7 tidal cycles. Transitions between the two is described in terms of probability that each row sums up to a probability of 1. Transition matrix was visualized using pyAgrum package.



(a) Distribution of the amount of sediment over the system at the start and end of simulating the system for a duration of onethe week.

(b) The probability distribution of sediment over the system at the start and end of simulating the system for a duration of one week.

Figure 4.6: Distribution of sediment in the system in terms of amount and probability. Loses were not included. Area 3 is expected to be underestimated as looking at Figure 4.4 a lot of the lost sediment borders that area.

Comparing the results from the reference run with the version of the Markov model in Table 4.2, quite some big differences can be seen. Both Markov models estimate the amount of sediment in area 1 would increase from 10.16% to 23.62% or 10.19%. The reference run makes physically more sense by showing the amount of sediment would decrease to 2.21%. Based on this alone the Markov model using a full tidal would be the better model. For area 2 the opposite is found that the model using half a tidal cycle decreases from 11.57% to 8.35%, which is comparable to the 7.70% of the reference run.

The model using a full tidal cycle the opposite is found that there is an increase to 12.01%. For area 3 the model using half a tidal cycle increases from 10.46% to 14.69%, which is similar to the 14.06% of the reference run. The model using a full tidal cycle estimates a higher amount of 18.42%. This seems high, but due to particles travelling out of the system the SedTRAILS reference run distribution does not sum up to 100%. Counting those loses to area 3 the reference run is expected to be closer to that value of 18.42%. Area 5 both Markov models show little changes that for half a tidal cycle the amount decreases from 49.85% to 48.61% and the full tidal cycle it increases to 50.35%. The reference run predicts a larger increase to 55.99%. the full tidal cycle version of the model does a better job again in that it does not decrease the amount of sediment in this area like the half tidal cycle version of the model does.

Overall the results are promising considering the transition probabilities are based on the sediment being uniformly distributed. This is different from the SedTRAILS run, which after a few tidal cycle does not have that same unofrm distribution of sediment over the system.

Table 4.1: Table comparing the resulting probability distributions of where to find sediment in the system using half and a full tidal cycle for the transition matrix with the initial distribution and the reference run of simulating the system for one week using SedTRAILS and RWS data

Sediment distribution over the system (%)					
	Area 1	Area 2	Area 3	Area 4	Area 5
Initial distribution	10.16	11.57	10.46	17.96	49.85
SedTRAILS week reference run	2.21	7.70	14.69	11.02	55.99
Markov model half tidal cycle	23.62	8.35	14.06	5.37	48.61
Markov model full tidal cycle	10.19	12.01	18.42	09.03	50.35

4.3. Autoregressive hidden Markov model

The Markov model describes each time step as a sum of the weights of all wave climate components. A disadvantage of this method is that the model does not have the ability to predict the difference between a storm and normal conditions. The second version of the model is a DBN model structures as an ARHMM to include the effect of a covariate which is the wave climate. The emission matrix was made from the transition matrices of the individual components of the wave climate. This brings the question what the transition matrix for the model should be?

As a first option we will try choosing a transition matrix using the assumption that wave climate components following each other up are independent of each other. This means that each row of the transition matrix has the same probability distribution. Two approaches were chosen to define the transition matrix this way. First approach the weight distribution of De Fockert (2008) wave climate from Figure 3.3 can be used. As a second approach it can be expected that the distribution has low probabilities for extremely energetic conditions, but extremely low energetic conditions of the system also have a low probability. This distribution is not expected to be symmetrical, which is why sampling from a Poisson distribution with $\lambda = 4$ seems another good option to define the transition matrix. Plotting both distributions in Figure 4.7 confirms that these two options are similar in shape.

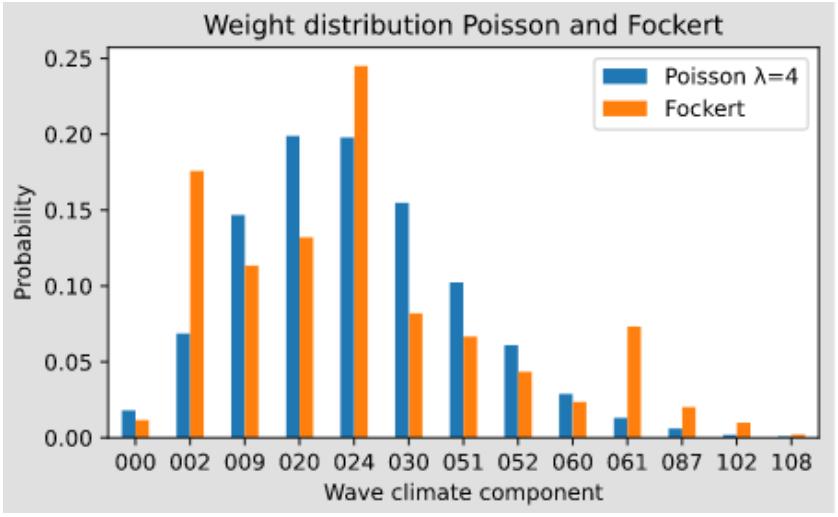


Figure 4.7: Comparison of the distribution of weights as by the wave climate schematization of De Fockert (2008) vs a distribution of weights by a Poisson distribution with $\lambda = 4$.

Like with the Markov model the two options will be tested by doing a simulation of one week again. The unrolled version of the ARHMM can be seen in Figure 4.8.

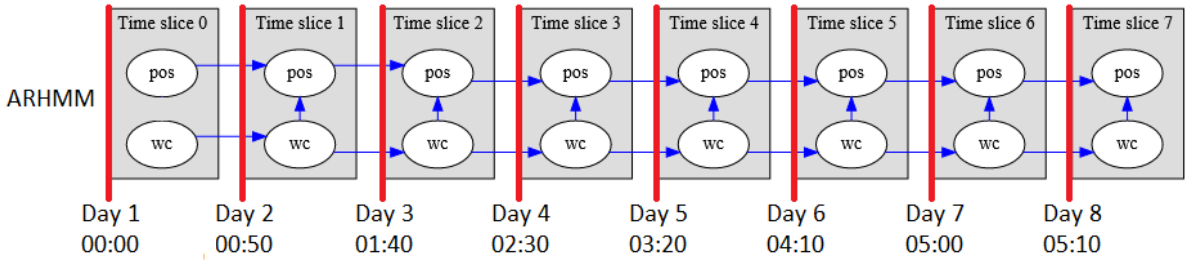


Figure 4.8: Structure of the DBN model as ARHMM to estimate sediment positions (pos) and corresponding order of wave climate component (wc) after one week. Each time step representing a full tidal cycle of 24 hours and 50 minutes described by the transition matrix between wave climate components, the emission matrix for the sediment position and initial conditions of the system. Unrolled version of the ARHMM was visualized using pyAgrum package

Applying inference by using Lazy Propagation the marginal probabilities in the system for both options were computed for the ARHMM in Figure 4.9. Something that can be immediately noticed is these distributions are very similar in shape, but slightly different values. The emission matrix and initial conditions of the sediment distribution over the system were the same that this can only be caused

by the differently chosen transition matrix. This is a good sign for the search for accurate solutions, but will be compared more quantitatively.

Adding the final distributions to table 4.1 gives table 4.2. It can be seen that the method for the transition matrix using the weight distribution of Figure 3.3 results in the same system as the Markov model using full tidal cycle. This shows that to include the wave climate as a covariate to the model, a transition matrix is needed that does not assume independence between the follow-up of wave climate component.

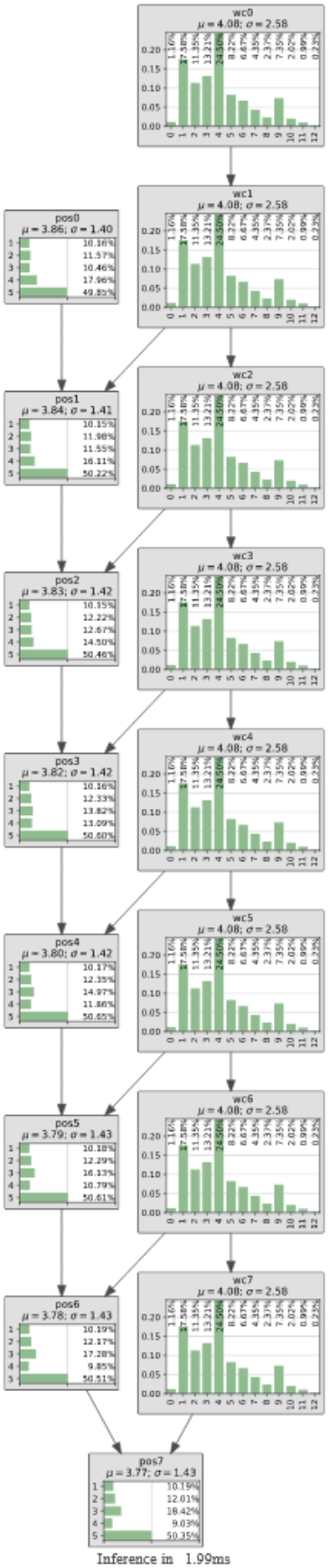
That leaves for the Poisson method only the question if it estimates this specific reference run better than the other method? This is possible as the reference run is just a week without much extremes in comparison to a full wave climate approach that must include extreme components. It can be seen that this description estimates 15.13% ends up at the downdrift coast in area 3. This is closer to the 14.69% of the reference run than the 18.42% of the other method. Like with the Markov models including the loses of the SedTRAILS reference run to the downdrift coast in area 3 is expected to still make the other method better.

The lesser amount of sediment ending up at the downstream coast in area 3 for the Poisson method causes all other areas to get more sediment than when using the Fockert approach. For areas 3,4 and 5 the basin and downdrift coast this gives a better estimate, but for areas 1 and 2 the updrift coast and ebb-tidal delta this makes the estimate worse. The reasoning for this could be that the tail of the distribution that describes the high energy components is smaller for the Poisson distribution than for the Fockert causing less sediment to be transported towards the downdrift coast in area 3.

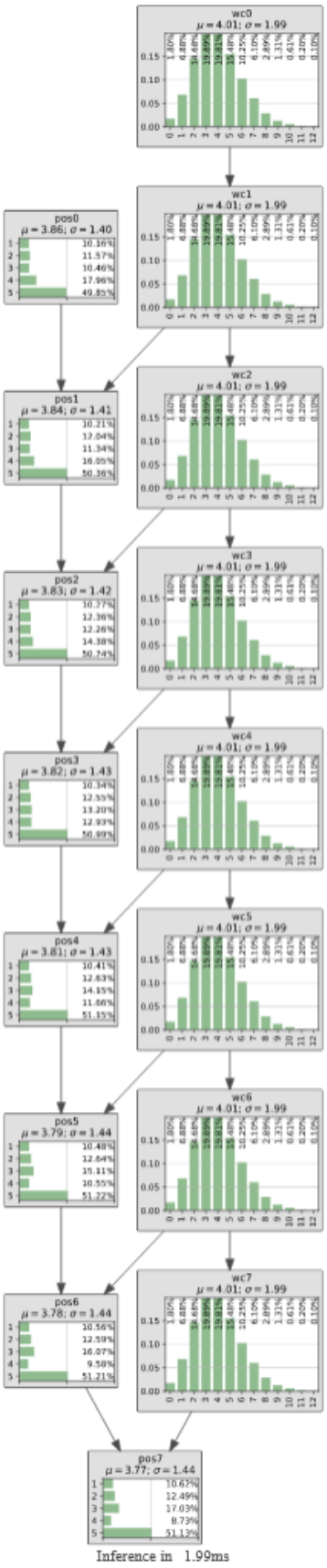
Beside the accuracy of the model also the computational time is a point of interest. The inference done on all the probabilistic models discussed so far only take milliseconds. This is significantly faster than for SedTRAILS, for which computing a run is a matter of minutes. This computational speed difference can be explained by that the model does not calculate the underlying processes saving computational speed. This was also the reasoning that was used in the aim of developing the model.

Table 4.2: Table comparing the resulting probability distributions of where to find sediment in the system after one week simulation time for the De Fockert (2008) weight distribution and a Poisson distribution for the transition matrix with the initial distribution and the reference run using SedTRAILS.

Sediment distribution over the system (%)					
	Area 1	Area 2	Area 3	Area 4	Area 5
Initial distribution	10.16	11.57	10.46	17.96	49.85
SedTRAILS week reference run	2.21	7.70	14.69	11.02	56.99
Markov model half tidal cycle	23.62	8.35	14.06	5.37	48.61
Markov model full tidal cycle	10.19	12.01	18.42	9.03	50.35
ARHMM Fockert	10.19	12.01	18.42	9.03	50.35
ARHMM Poisson	10.57	13.52	15.13	9.22	51.57



(a) Inference using Fockert wave climate as transition matrix



(b) Inference using Poisson distribution with $\lambda = 4$ as transition matrix

Figure 4.9: Inference of all the marginal distributions of the ARHMM comparing using a transition matrix that uses the Fockert wave climate weight distribution and using a Poisson distribution with $\lambda = 4$ as transition matrix. Each variable: "pos" stands for the position of particles in the system as a discrete random variable with state space [1,2,3,4,5] representing the different areas and each variable: "wc" stands for the wave climate components described by a discrete random variable with state space [0,...,12] representing each wave climate component. The number in each variable represents represents the time step in the system. Pos0 stands for the initial distribution, which was added as soft evidence. For each computed marginal distribution in the system also the mean μ and standard deviation σ is shown.

4.4. Adding evidence

Evidence can be applied to the model to estimate how wave climate and sediment positions influence each other. In this way a link can be made with physical scenarios to see if the model shows behaviour expected of the system and give some examples how the model could possibly be used.

4.4.1. Schematized storm simulation

First we consider the scenario of normal wave conditions turning into a storm and calming down again. The evidence for this case will be different initial positions of the sediment particle together with the wave climate sequence: $3-4-8-7-10-5-3$. Both the Fockert and Poisson approach gave the same results with this evidence, which is why in Figure 4.10 all runs used the Fockert approach for the transition matrix of the wave climate.

The upper plot shows transitions between wave climate components, which were set as evidence. For this reason every time step a wave climate components is set to have a probability of 1. In between it shows linear transitions between those wave climate components, but this is just a visualisation. There is no information about the transitions between different time steps that the physical meaning could be completely different. A higher time resolution with different emission matrices would be needed to describe the behaviour on a smaller time scale.

The results for the sediment particles are seen to be different for every starting position. Sediment from the updrift coast (area 1) has a high probability to stay in that area, but has a small chance to go to the ebb-tidal delta (area 2) or continue to the downdrift coast (area 3). Sediment from the ebb-tidal delta (area 2) has a high probability to go to the downdrift coast (area 3). It also has a probability to stay in ebb-tidal delta (area 2) or go back to the updrift coast (area 1) and a very small probability to go to main ebb channel (area 5). Sediment from the downdrift coast (area 3) does not re-enter the system and just shows staying in this area as the model can not show sediment leaving the system. Sediment from west side of the basin (area 4) spreads a lot over the different areas. It is most likely to stay in the same part of the basin (area 4), but may go to the ebb-tidal delta (area 2) or the east side of the basin (area 5). It also has some smaller probabilities to end up at the down- and updrift coast (areas 1 and 3). Sediment from the west side of the basin (area 5) has a high probability to stay in that area, but has a small chance to end up in the ebb-tidal delta (areas 2) or the downdrift coast (area 3).

Comparing with the SedTRAILS run in Figure 4.4, most of these results correspond with the physical processes that were described. The only exception is that for the updrift coast (area 1) the model estimates most sediment to stay in the area, but that does not show in the SedTRAILS run.

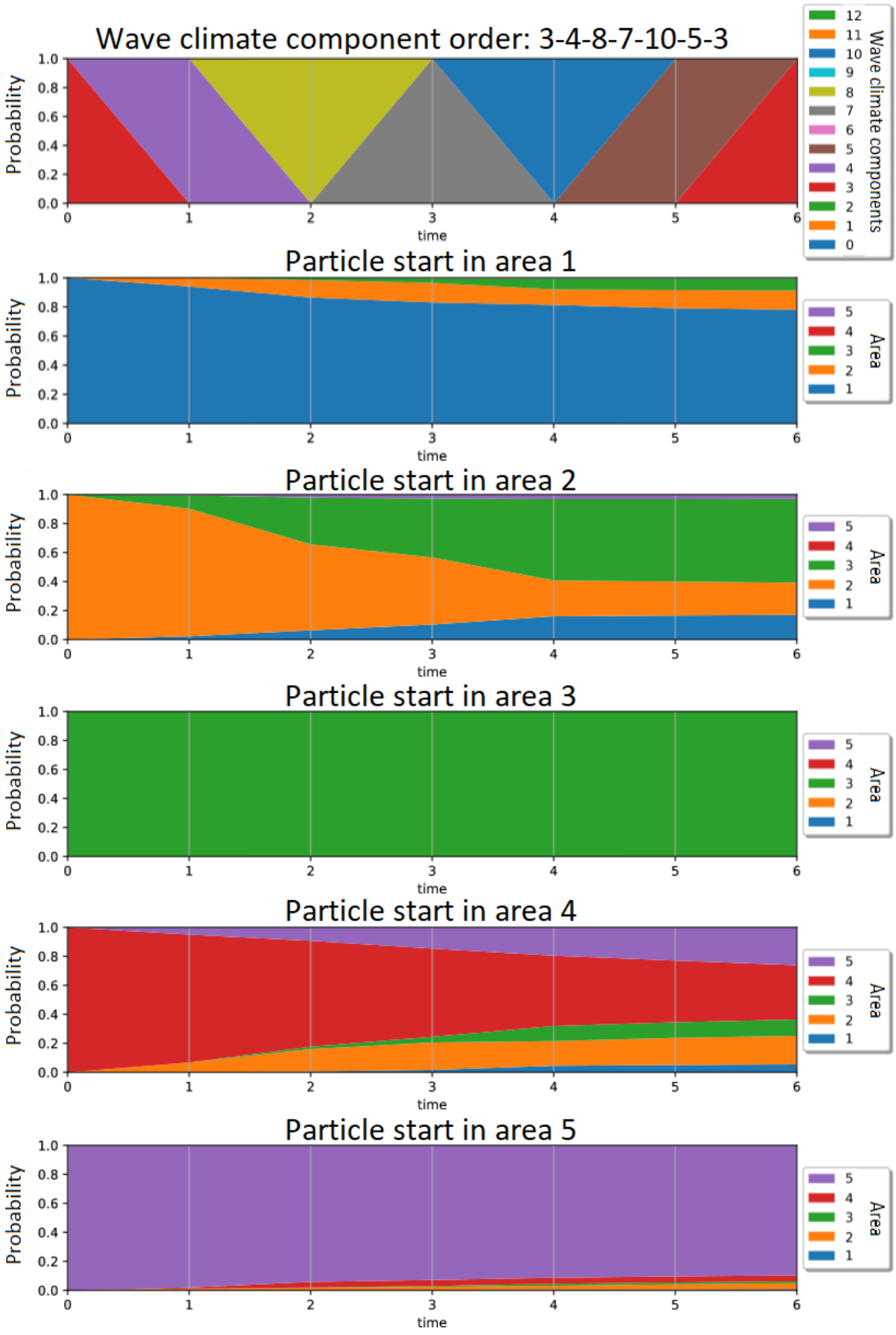
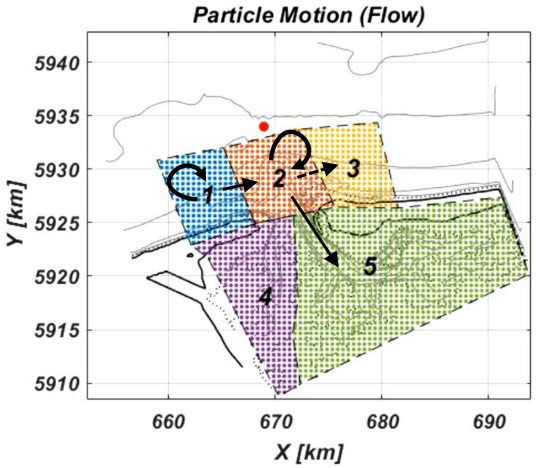


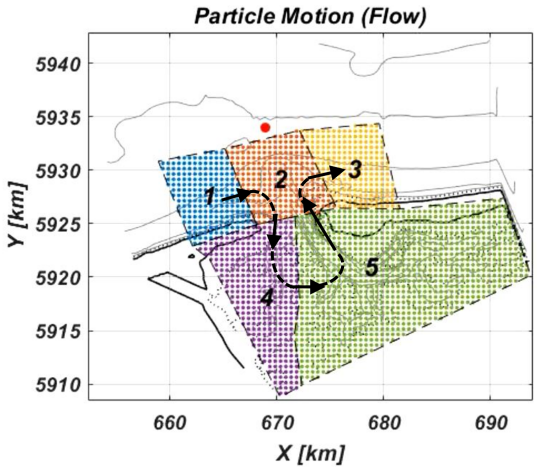
Figure 4.10: Simulated storm of wave climate components order: $3-4-8-7-10-5-3$, for different start positions in the system. For version of the model using the Fockert wave climate distribution for the transition matrix

4.4.2. Pathway occurrences

Another option is setting evidence for the sediment positions and see how the wave conditions that can make that possible are distributed. The sediment pathway scenarios that will be considered are visualized in Figures 4.11a and 4.11b.



(a) Visualization of the sediment pathway for orders 1-1-2-2-5 and 1-1-2-2-3.



(b) Visualization of the sediment pathway for order 1-2-4-5-2-3.

Figure 4.11: Visualizations of different sediment pathways chosen by assigning evidence to the positions nodes in the model.

In Figure 4.12 we first create the pathway scenario of sediment moving with the longshore current and ending up being imported into the basin by setting evidence for the sediment to move between areas in the order: $1-1-2-2-5$. This pathway the sediment particle would enter at the updrift coast (area 1) and "slowly" go to the ebb-tidal delta (area 2) and enter the basin by the main ebb channel (area 5). All the wave climate components have a probability to cause the specific movement of the sediment during a time step in most cases. Some components may have a higher probability during a time step to cause the specific transport from one area to another. For example at $t = 4$ both methods estimate that for the evidence a particle has traveled from the ebb-tidal delta (area 2) to main ebb channel (area 5), the higher wave climate components have a higher probability than for the other transitions in the time series. This would suggest that during those extreme conditions there is a higher probability for the particles to go into the basin. Both methods are similar, but for the Fockert approach wave climate component 9 it says the chance is higher to cause this transition.

At the bottom graph of Figure 4.12 also a different transition is shown where the sediment at $t = 4$ moves to the downdrift coast (area 3) instead. Comparing this to the graph directly above it can be seen that wave climate component 11 is more likely the cause for transport of a sediment particle from area 2 to 3 than from 2 to 5. This is interesting as those transtions can be connected to different bypassing mechanisms.

Remark to be made here that this logic does not mean wave climate component 9 is more likely to import sediment into area 5 than in area 3. As the sediment pathway was given as evidence the actual amount of sediment transport can be less than the transport of a different sediment pathway in opposite direction.

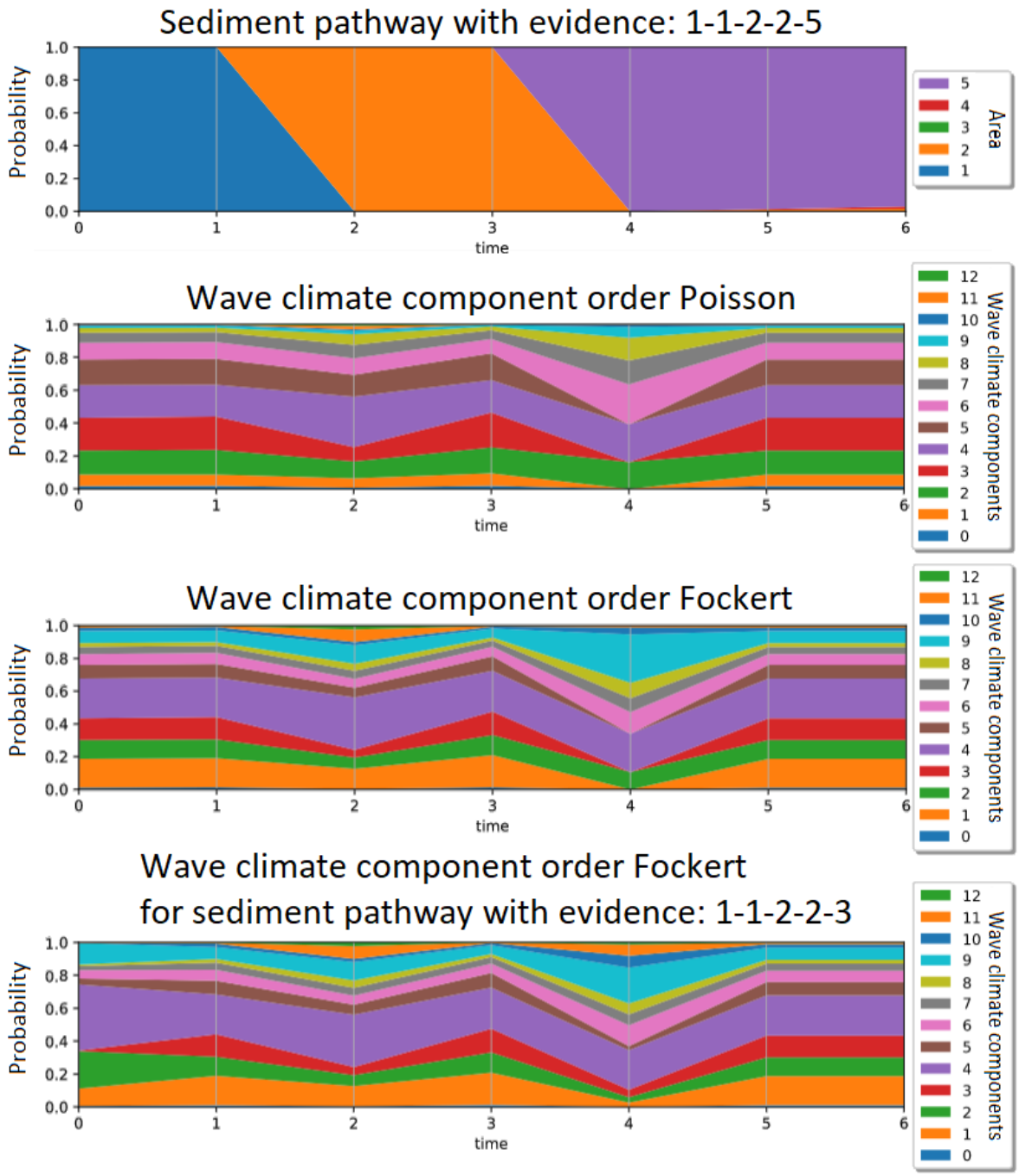


Figure 4.12: Inference result of wave climate component order distribution in time for both methods Fockert and Poisson with evidence of sediment traversing in the order: $1-1-2-2-5$. Bottom plot also shows the result for Fockert method, but for evidence sediment traversing in the order: $1-1-2-2-3$.

In Figure 4.13 the evidence is set for the sediment to move between areas in the order: $1-2-4-5-2-3$. At $t=2$ a special case can be seen that the transition from area 2 to 4 is only included in the part of the emission matrix for wave climate components 0 and 8. This shows that some very low probability transitions that are not be visible in the model are dependent on the wave climate schematization and the amount of particles used to populate the areas in the SedTRAILS runs.

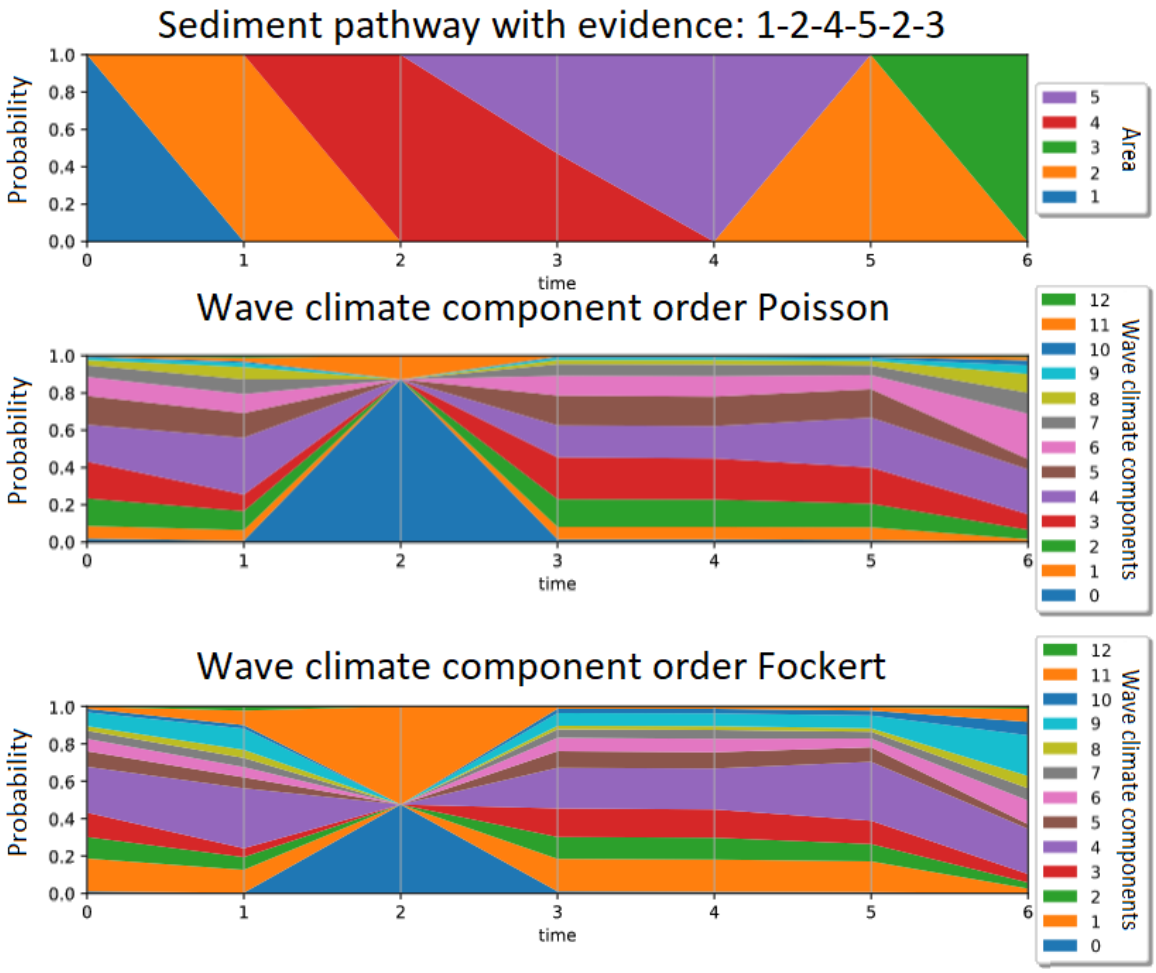


Figure 4.13: Inference result of wave climate component order distribution in time for both methods Fockert and Poisson with evidence of sediment traversing in the order: $1-2-4-5-2-3$.

Overall the physical meaning of the model looks promising that it does describe some of the aspects from how SedTRAILS can describe sediment pathways, but some limits are seen that will be discussed in Chapter 5.

5 Discussion

In Chapter 4 several results were found for the ARHMM and the Markov model. The first results showed that in the Markov model using a full tidal cycle for the transition matrix gives better results, than when using only half a tidal cycle. Comparing these results to a reference run with a simulation time of one week both versions of the model show a mjority of the expected physical processes in the system. An exception on this was persistence of sediment staying at the updrift coast (area 1), which was not seen in the reference run.

In the next part the system was described by an ARHMM that was to add the effect of the wave climate as covariate. For the ARHMM two different transition matrices were tested for the follow up of wave climate components, which were based on the weights in the wave climate schematization and by sampling from a Poisson distribution with $\lambda = 4$. The transition matrix for the wave climate components was assumed to be independent that there were no dependencies between wave climate components following each other up, which resulted in a matrix with the same distribution in each row. This resulted in the system being equivalent to the Markov model showing that to add the effect of the wave climate as covariate the dependencies between wave climate components are needed for the model. Results from the Poisson distribution variant also showed less sediment ending up at the downdrift coast (area 3) leaving more sediment in other parts of the system. This could be explained by the tail of Poisson distribution being smaller than that of the Fockert weight distribution. After that evidence was added to the model to simulate some physical scenarios of the system, which showed again the persistence of sediment to stay at the updrift coast (area 1). Another thing it showed is that some transitions in the model are very rare they do not show up for some of the wave climate components.

5.1. Learning and validation

To get a complete model it would still be necessary to do the learning step and a better validation method like mentioned in the flow chart of Figure 3.1 in Chapter 3 that were omitted due to time-constraints. These steps were intended to be done by using using the week time series and also a month time series to create a SedTRAILS dataset that would be processed to describe for all the particles in the system which area they are each time step for a chosen time resolution. This synthetic database would have to be split into a training and testing dataset. The training dataset would be applied in the model as observations to learn the transition matrix of the model using EM.

An attempt was already made to test the learning step by sampling a database from the model using the transition matrices of the Poisson method, removing the data of the wave climate and applying EM to a system that does not have the chosen transition matrices. The results were unsatisfactory in that in the learned case at the right side of Figure 5.1 the inferred distributions for the wave climate do not look like the Poisson distributions at the left side. The method still looks promising as the test did not schematize the system correctly in terms of structure and initial distributions. Also the used python package pyAgrum (Gonzales et al., 2017) seems more optimized for learning BN, than the latent variable structure of a HMM as DBN. There could be more optimized methods to train this type of models.

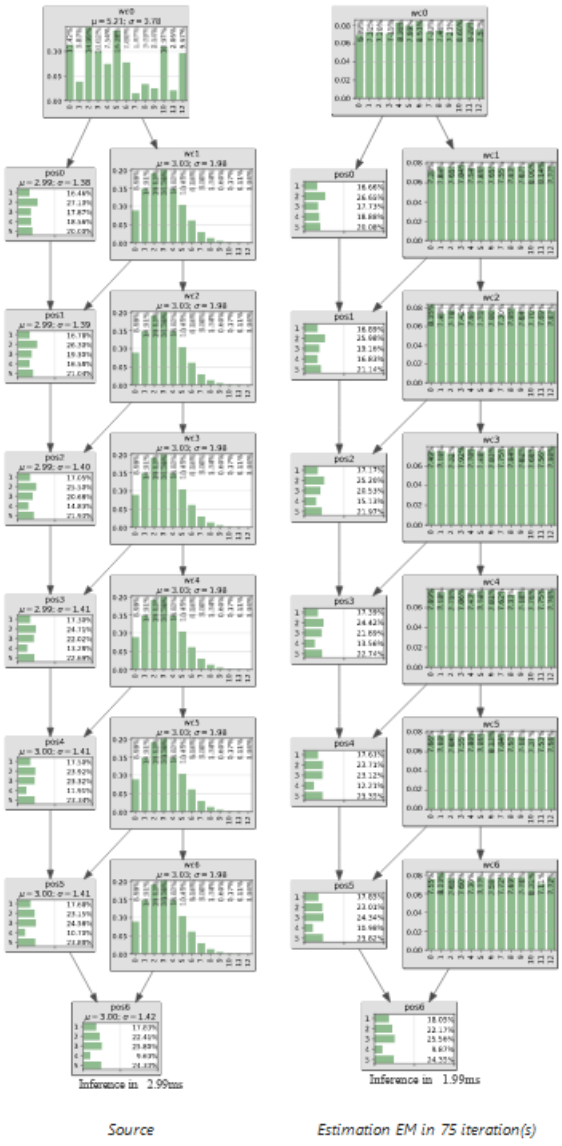


Figure 5.1: Original inference of the system using the Poisson method and inference of a learning attempt using *Expectation Maximization* (EM). The learning attempt using EM was done by sampling from the system with Poisson distribution for the transition matrix, removing the wave climate data and trying to re-learn that Poisson distribution as transition matrix.

The python package DYNAMAX for probabilistic *State-Space Model* (SSM) (Linderman, 2024) seems more optimized for these type of computations. The package is still in development that the emission matrix could not yet be implemented for our ARHMM. The package looks promising though as quite some techniques were used to optimize computational speed for learning and inference of SSM. An example of this is that the package is made using JAX (Bradbury et al., 2018), which is a python library for accelerator-oriented array computation and program transformation. Beside the computational speed advantage JAX gives when doing a lot of array computations, it also uses a different *pseudorandom number generator* (PRNG) (Jax, 2024). There are many variations of PRNG, but Python and many applications use one called Mersenne Twister. This PRNG is known to have some flaws (Matsumoto et al., 2006) that it may be wondered if better PRNG should be chosen (Vigna, 2019) and how that would affect *Machine Learning* (ML) applications.

Probabilistic machine learning methods are in full development at the moment that more packages can be expected to become available in the coming years that learning and validation becomes easier for state-space models.

5.2. Time slice choice

One of the first results was about the choice of the time period each step of the model represents. The sediment pathways data from SedTRAILS makes it possible to create a transition matrix for any moment in the time series. This choice will have a big impact on how the model behaves when it repeatedly applies these probabilities to model the next steps. The choice of one full tidal cycle as time step gave quite good results that could be explained by the known physical processes in the system. The downdrift coast (area 3) was a clear sink in the system increasing from 10.46% to 18.42%. The west side of the basin (area 4) is being a source losing sediment from 17.96% to 9.03% clearly exporting sediment out of that area. The west side of the basin (area 5) is receiving sediment in the main ebb channel, but generally does not change that much only increasing from 49.85% to 50.35%. The ebb-tidal delta (area 2) only increases a little from 11.57% to 12.01%. These results do not get to the exact amount as in the reference run, but they are close enough that it would be a matter of improving the accuracy of the model.

The main exception was the persistence of particles to stay at the updrift coast (area 1). It could be wondered if this was just a low likelihood scenario happening in the reference run, but more likely the transport from the updrift coast (area 1) is not that well represented in one full tidal cycle. During the reference run sediment particles seems to speed up the longer the run goes being more dynamically active. During bypassing the sediment changes direction going into the ebb channel or go along the periphery. The movement in the first tidal cycle could be very limited that these processes have not started yet. The movement between the updrift coast (area 1) and the ebb-tidal delta (area 2) in that case would just behave like oscillate around that border instead of bypassing the tidal inlet.

Choosing a time slice bigger than one full tidal cycle could describe the start of bypassing better, but this also gives the disadvantage that the time resolution of the model will decrease and the results for other areas may get worse. As a solution sediment particles may not just be dependent on their prior state, but also on some more prior states. This would mean the Markov property would not fully be valid that a higher order Markov model would be necessary to increase the accuracy of the model. This would change the ARHMM structure into the structure of Figure 5.2. Something to be noticed is that in this change of the system the wave climate seem to get the role of influence to determine the amount of movement or energy of the system and the autoregressive part of the positional data the direction of the movement.

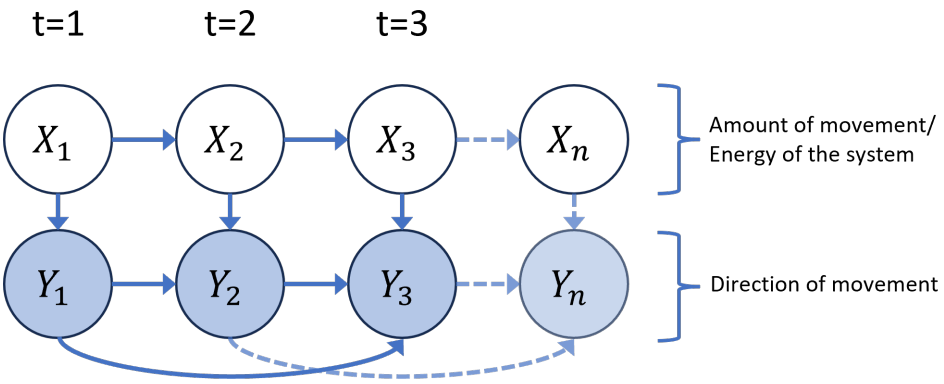


Figure 5.2: Expanding the ARHMM by adding second order Markov dependency between the sediment positions in the system. WC stands for wave climate and pos represents the position of a particle in the system.

In terms of data this would mean the emission matrix would need another dimension to account for each possible position at the prior time step. This extra dimension could be added by dividing the data by source area and create for each source a transition matrix with the second half of the data like shown in Figure 5.3. It would still have to be done for every wave climate component, but as advantage it does not mean any extra runs of Delft3D and SedTRAILS given those runs can represent 2 time steps.

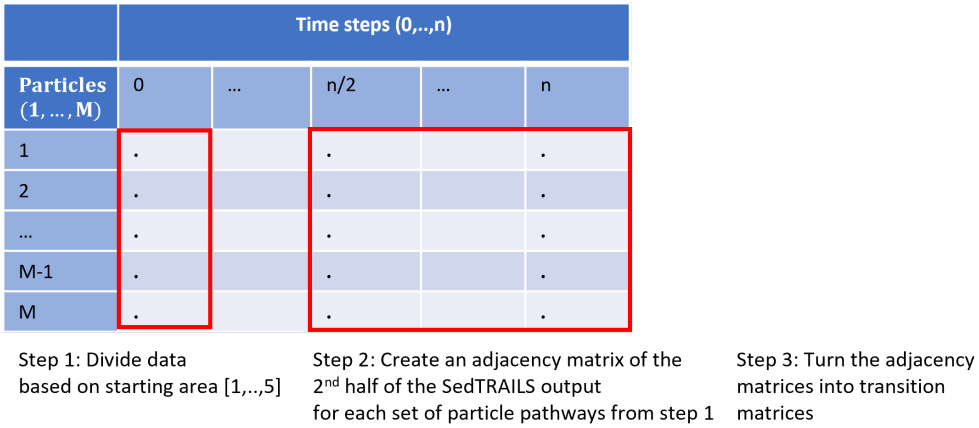


Figure 5.3: Method to create the emission matrix for an ARHMM with 2nd order dependency from SedTRAILS data consisting of M particle pathways with n time steps when described as area positions. The resulting matrix from step 3 is a 5x5x5 matrix, which to get the emission matrix this must be done for each of the wave climate components to get the final 13x5x5x5 matrix.

This method could also be used to model the effect of border areas, which would be important when trying to reduce the time step to get a higher time resolution of the model. For example in Figure 5.4 it can be seen that within a tidal cycle particles near a border between areas are expected to drift at that border making more transitions than for example a particle that would be in the middle of an area. The extra autoregressive component can function here as a measure of if the particle is close to border to another area or not. If it just transitioned between areas it is close to the area border, which can be described by a transition matrix with high probabilities of transition, than when it did not transition between areas the last time step, which could be described by a transition matrix with high probabilities of staying in the same area. To really model with higher time resolution each different physical motion within the tidal cycle would need a different emission matrix. It would be like the same follow up of emission matrices for each tidal cycle. In Figure 5.4 it can also be seen that a schematization of more than 8 emission matrices would be sufficient that doubling the amount of matrices to 16 would not add too much info to the model.

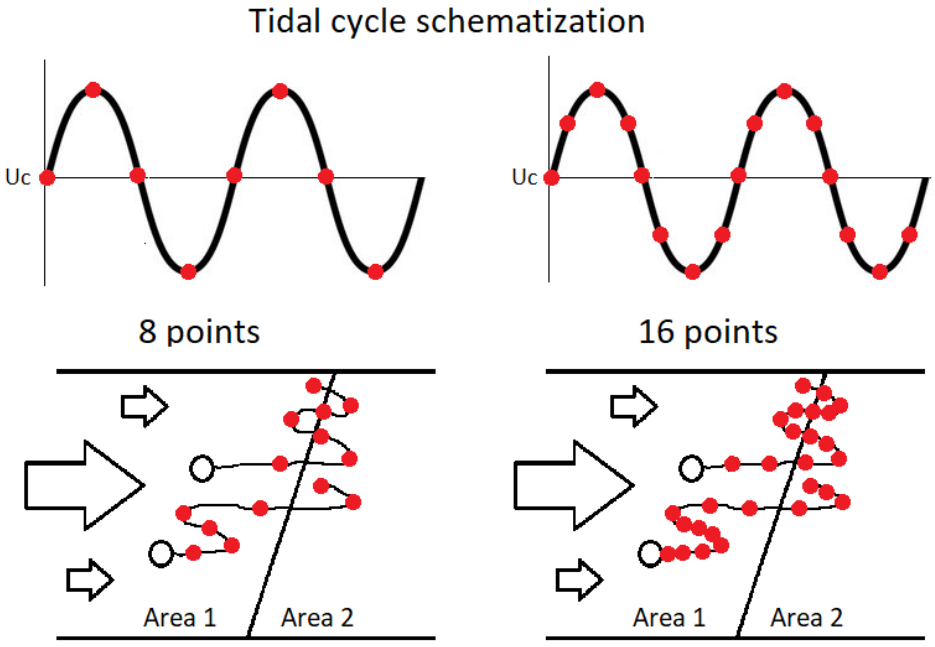


Figure 5.4: Example showing conceptual sketch of a typical sediment pathway during a tidal cycle and that the choice of the time resolution higher than 8 time slices per tidal cycle is expected to only have minimal influence on the transition probabilities when creating a higher order ARHMM

These ideas do add a lot more different matrices to the model, which may end up increasing computation time. Quite a lot of matrices may be empty though that when using the autoregressive idea with a small time step some transitions are simply not possible. For example a particle could not directly travel from the updrift coast (area 1) to main ebb channel (area 5), so that matrix would just be empty.

5.3. Area schematization

So far the model has only been discussed having 5 areas to describe the system. The main advantages of this choice are that the model is easy to interpretate and matrices for the model have stayed quite small favouring the computational cost to create the model. Disadvantages are that the current system only gives a limited spatial resolution of where particles can end up in the system and that some particles left the system affecting the computation of transition probabilities.

For the reference run the effect of particle losses was also seen in Tables 4.1 and 4.2 that the probabilities describing how sediment is distributed over the system did only sum up to 0.9261. In Figure 5.5 the particles for each source area are counted at the start and end of the reference run are counted showing most of these loses are from downdrift coast (area 3). As from Figure 4.4 it was seen that most of those lost particles end up leaving the system at the edges of area 3. Adding these particles to area 3 would increase the probability of finding sediment in area 3 to from 0.1469 to 0.2208. This makes a significant change as the Fockert approach now does a better job estimating the probability of finding sediment in area 3 than the Poisson approach.

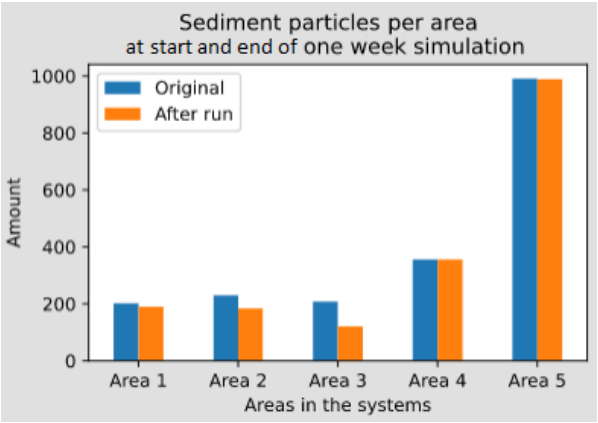


Figure 5.5: Sediment particles of each area inside the system at the start and end of the week reference run using SedTRAILS. The difference can be interpreted as loses of the model. Most of these lost particles can be found having traveled outside the borders of area 3.

To take this effect better into account simply increasing the polygon area of area 3 would not work, as that would give more particle sources close to the edge of the model that can leave the system again. A solution could be to have an empty cell at the outside that is not populated with particles to measure these particle loses. As this area would not have any source particles the model would not have any data for transition probabilities resulting that this area would not need a row in the transition matrix. This would also mean that any particle motion going into this area would get ”stranded” into that area as there are no transition probabilities going out of it, which has the flaw the model would not be able to take into account particles that re-enter the system.

Another idea that could be played with is to not uniformly distribute the sediment particles over an area when populating the polygon. This is a very sensitive thing to change as it will directly affect transition probabilities of the transition matrices. There is already the effect of some particles on land influencing transition probabilities that it may be wondered if the defined areas to be seeded should all be expected to have similar physics working on them. The advantage that non-uniform seeding could give is that if the polygons are initially less densely populated at the edges of the system, it would decrease losses of the system. A better implementation of that would be to just enlarge area 3 in the model, but keep the seeding the same. That way the uniform distribution of particles over the area can remain and particles remain in the system as long as the enlargement of area 3 is big enough. Besides accounting for losses, the initial distribution of particles over the polygon could also give the

opportunity to add info to the system about sediment availability in the system. How useful this can be is a bit unclear as for the model development we are changing a SedTRAILS step, which already has the freedom factor in Soulsby description for the grain velocities to determine sediment availability.

To improve the spatial resolution of the model the amount of cells in the system would have to be increased. Each extra cell would add another row and column to each transition matrix quickly increasing the computational cost of the model, that it could end up with larger computation times than SedTRAILS. To solve this problem it must be noticed the current system uses discrete random variables to describe positions in the system. With a large amount of states this random variable could take it would be better to schematize positions in the system as continuous random variables allowing use of different mathematical methods.

A particularly interesting example is in Murphy (2023), where an ARHMM was created using an emission matrix defined by a multivariate normal distribution and having 5 hidden states. Instead of areas, the emissions are described as (x,y) positions giving the advantage the spatial resolution of the model is equal to the grid. As a disadvantage they have to define functions for the emission matrices describing the physics of the system and learn the parameters for these functions, where our model could just determine the emission matrices from the data described using connectivity. Both models show similarity in that the 5 hidden states in the example describe different physical states of the system that can follow each other up. This is similar to how different wave climate components in our ARHMM can follow each other up. It was mentioned this type of model is called a *regime switching Markov model*, which were said to be widely used in econometrics (Hamilton, 1990).

5.4. Covariates

The initial intent of the DBN approach was to add *covariates* to the model. For our ARHMM the effect of the wave climate as covariate working on the sediment pathways did not show up very well, because of limitations in creating a transition matrix for wave climate components following each other up. Still the approach looks promising in that if learning can be correctly applied not just the transition matrix for the wave climate can be determined, but also adding other covariates by adding more nodes to the model could be considered.

One covariate that could be useful to add is a variable that represents the direction like the wind direction. This variable is already included in the wave climate as part of the Delft3D runs, but having it as a separate influence may account for a lot of variability in the system in that storm conditions with wind blowing in one direction would be very different from storm conditions with wind blowing from the opposite direction.

Another covariate that would be useful for further use of the model is that of human decisions, such as where to place a nourishment in the system. In this case the management would be included in the model by having choices in the model and different rewards for transitions (Howard, 1960). Including decision making into Markov models is also called a *Markov decision processes* documented first by Bellman (1957). A more recent from the medical field can be found in (Alagoz et al., 2010).

Independent of which covariates are chosen be added to the model the ARHMM structure works as a good basis to expand from. In that matter the biggest advantage of choosing an ARHMM structure to describe our system is that it is a robust description lowering the bar to also apply it to other systems.

5.5. Wave climate schematization

The wave climate schematization lies at the foundation for creating an ARHMM to estimate the sediment pathways for our system. De Fockert (2008) wave climate schematization was used, but how would a different wave climate schematization affect the model?

The first version of the model just consisted of a *Markov model* using weights of the wave climate schematization to determine the effect of wave climate components as a sum of all the components. The predictions are highly dependent on the wave climate schematization and had the disadvantage that the low probabilities of high energy events would be split over time. You could know that a storm is predicted in the next few days, but the model would not be able to specifically apply these components of the wave climate to estimate longer sediment pathways than during normal weather.

The ARHMM structure is less dependent on the wave climate schematization as it does not need to use the weights of the wave climate schematization when the transition matrix for the wave climate is determined using parameter learning. Instead the follow up of wave climate components would be learned from the sediment pathway data of a longer time series. This method is a *probabilistic machine learning* method. A disadvantage of this method is that wave climates typically need very long time series to be accurate. Even if it would be computationally possible create such long time series with SedTRAILS, the result would be incorrect as to compute sediment pathways SedTRAILS assumes a morphostatic situation which is only valid for smaller time scales. It does give the advantage though that the wave climate is not just learned in terms of occurrence, but that some dependency is learned how wave climate components follow each other up. This concept of the importance of wave climate conditions following each other up was also found under the terminology *wave chronology* in (Aragón et al., 2023; Malliouri et al., 2023; Southgate, 1995).

To circumvent the issue of the time series it would be easier to learn the transition matrix for the wave climate components separately using wave climate time series as the sediment itself does not cause the wave climate. For the validation to be correct the model would still needs to check how accurate estimates of sediment positions are by comparing them to computations using a time series. An advantage here is that the model does not need to be more accurate than SedTRAILS so a validation of a time period of 1 week or 1 month would probably sufficient. To test the capabilities of the model to estimate accurately for other series of wave climate components multiple runs of 1 week or 1 month could be done by using different time series. This would be a bit like sampling from the wave climate to test the accuracy.

5.6. Sediment

For the SedTRAILS runs a sediment size of 400 μm was chosen. It may be wondered how well the model would behave for different sediment sizes and if a mix of sediment sizes would also be possible to be modelled to end up being able to compute complete sediment fluxes.

Tidal flats mainly consist of very fine sand (90%; grain size $150 - 200\mu m$) and fine-grained muddy sediments (10%) and the ebb-tidal deltas primarily consist of of fine sand ($100 - 400\mu m$) (Wang et al., 2018). This means the sediment size chosen for the model is quite at the larger end of the spectrum. It would be recommended to do more sensitivity tests with different grain sizes. For smaller grain sizes, the sediment pathways are expected to be longer as particles suspended higher into the water column by wave forces will have a longer settling time due to more distance they would need to settle and lower settling velocity because of lower mass. In Figure 1.4 these differences in sediment pathway for different sizes of sediment was also shown. For the model this would mean for a smaller grain size more sediment is expected to end up out of the defined areas resulting in more data losses to define the ARHMM. This is particularly the case for doing the SedTRAILS runs for the high energy components that upscaling of the system might be necessary.

SedTRAILS uses the grain velocity description described in Soulsby et al. (2011) to compute the sediment pathways caused by bedload and suspended sediment transport, but does not yet model the transport of fine sediment ($< 80\mu m$), limiting the availability to model finer sediment sizes for now. If there is sediment pathway data for the different wave climate components, creating an ARHMM should be possible though. A big challenge for fine-sediment modelling is adding the effect of *flocculation*. Adding this effect to sediment transport models in general is still a research topic fully in development (Chassagne & Safar, 2020; Spearman & Roberts, 2000). Not much can be said yet about application of these techniques to Lagrangian sediment transport models like SedTRAILS. The Lagrangian description of sediment transport using sediment pathways does seem favourable that for example the size change of sediment particles due to flocculation could be included that way. For the ARHMM this could potentially be included by adding an extra covariate to the model. Another challenge for fine-sediment modelling are *memory effects*. This is the effect that fine-sediment does not directly respond to changes in currents. Having discussed the possibility of adding more autoregressive components to the model in the section about the time slice choice the ARHMM seems quite well equipped to include memory effects.

The data obtained from SedTRAILS is specific for the chosen sediment size in the pre-processing step. There is not really a necessity to expand the model to more grain sizes as this would make it much more difficult to interpretate the complexity of sediment pathways in the system. If it would be desired to fully compute the quantities of sediment transport, the model would have to account for all the grain sizes in the system. There are some existing sediment transport models that take mixed sediment sizes already into account, of which the most well-known example is the Hirano active layer model (Hirano, 1971) which is implemented in Delft3D. This option in Delft3D was also used In Pearson et al. (2020) to apply to SedTRAILS. Another example of a model that accounts for multiple grain sizes is the aeolian transport model AeoLIS (Hoonhout & de Vries, 2016; van Ijzendoorn et al., 2023). Schmelter et al. (2015) also developed a multi-fraction Bayesian sediment transport model, which could mean adding the effect of mixed sediments is also possible for the ARHMM.

5.7. New opportunities

Several parts of the model have now been discussed showing the way the ARHMM is structured acts as a good basis to develop a probabilistic model that can accurately estimate sediment transport pathways. The methodology to create the ARHMM shows that Lagrangian models like SedTRAILS are an essential step to create the emission matrix of these types of models. To determine the transition matrix for the follow-up of wave climate components a probabilistic machine learning method was suggested, which will contribute to the development of applying *Machine Learning* (ML) methods to the field of sediment transport.

In recent years rapid advancements have been made in the application of ML methods to fluid mechanics Brunton et al. (2020) that it is only be a matter of time for the field of sediment transport to follow. Some developments have already been made, that a good summary of examples can be found in Goldstein et al. (2019). A main concern of ML methods is that these methods can end up in a black-box approach in which we can not recognise the processes that are leading to the results of the model. In the search to create faster models using probabilistic methods this will inevitably lead to a trade-off that processes of low influence have to be omitted to speed up computational speed. The ARHMM in this thesis may have limited accuracy to estimate sediment transport pathways for now, but computation times for inference are almost instant and the only covariate in the system can be physically interpreted as the effect of the wave climate that there is a lot of opportunity to further develop the model and possibly apply this framework also to other sediment transport cases.

6

Conclusion

In this thesis a methodology was developed and tested to create a probabilistic model of sediment pathways with the aim of decreasing computation times. This was done using the following research questions:

RQ1: *What type of probabilistic model could be made for sediment pathways by using Markov chains?*

The currently available models for sediment transport in tidal inlets were first explored to show what knowledge gap probabilistic models can fill in this field of research. To create a probabilistic model Markov chains were used, which were explained as probabilistic graphical models.

A *State-Space Model* (SSM) is a specific probabilistic graphical models that is characterized by having latent variables. All SSM can be described as a *Dynamic Bayesian Network* (DBN) which is a *Bayesian Network* (BN) with time dependence. A specific SSM is the *Hidden Markov Model* (HMM), which models independent observed random variables in a system that are dependent on a series of latent random variables and each other. When the observed states and dependencies in the system are known all marginal probabilities in the system can be determined using *inference*. When some dependencies in the system are not known that distribution can be determined using a *parameter learning*. The explanation of these different models gave a good basis for understanding the final model used, which was a *Auto-regressive Hidden Markov Model* (ARHMM). This type of model is the same as the HMM, but with an extra autoregressive link between the observations. This extra dependency fits a dynamic system like for sediment transport pathways very well as the position of a particle is not just dependent on the underlying forcing, but also the prior position of that particle.

RQ2: *What processing steps are necessary to create the probabilistic model?*

To create the model, 13 wave conditions from a wave climate schematization were used to first model the hydrodynamics of the system using Delft3D, and after that to model sediment pathways using the Lagrangian model SedTRAILS. The obtained sediment pathways were used to create a transition matrix for each run describing the probability distribution of each source where sediment will end up after one tidal cycle. These results were used to create a Markov model using the weights of the wave climate distribution and an ARHMM that describes the wave climate as latent states and the sediment positions as observed states.

RQ3: *How accurate are results of the probabilistic model in comparison to SedTRAILS results?*

The Markov model was tested by comparing the description of the transition matrix using a full tidal cycle and using only half a tidal cycle with a reference run of SedTRAILS using a week of wave data from Rijkswaterstaat (RWS). The method using a full tidal cycle was seen to give better results than the one using only half a tidal cycle. This was clearer when loses of sediment particles out of the system were accounted for.

For the ARHMM, learning the transition matrix for the follow-up of wave climate components was omitted due to time constraints, such that a more simplified matrix was tested with the same distribution for each row. Two options for this distribution were tested, which were the weight distribution of the wave climate and a distribution sampled from a Poisson distribution with parameter $\lambda = 4$. The results were again compared to the reference run of one week, showing that for the weight distribution of the wave climate the same result as that of the Markov model was found. From this was concluded that the transition matrix should not be the same distribution for every row to really add the wave climate as a covariate. The Poisson distribution had a smaller tail in the distribution than the wave climate distribution, which could be connected to the result that less sediment out of all areas ended up in area 3, acting as a sink to the system.

Many ways to improve the model were discussed. A better learning and validation would be necessary to quantify the validity of the model better. For learning a longer time series could be done or as another option the approach could be changed by learning from wave climate data instead. Adding autoregressive components to the model could improve the accuracy of the model. Improved area schematization options of adding a loss cell or changing the source distribution could account for losses when modelling the system. Adding more covariates could improve directional effects in the system that energy and direction are not directly linked as in the used wave climate schematization. Another covariate that could be added is by including management decisions into the model. This could be like the position were to put a nourishment directly linked to the costs. The transition matrix for the follow-up of wave climate components could possibly be learned separately. And last the ARHMM shows some potential to include challenging fine-sediment effects of flocculation and memory. For these last effects SedTRAILS would first have to be further developed to get the necessary sediment pathway data.

RQ4: *How fast can the probabilistic model compute results compared to SedTRAILS?*

Many mentions were made of the computational cost of the probabilistic model to aim for faster computation times than the original Lagrangian model SedTRAILS. In the results it was seen that inference on the created Markov model and ARHMM only took milliseconds. This is significantly faster than SedTRAILS, which takes minutes to compute the sediment pathways for the same system.

The faster computational speed of the Markov model and ARHMM was explained by the fact they do not have to compute the underlying physical processes of the system. Some of the recent existing probabilistic approaches to sediment transport (Kroon, 2024; Schmelter et al., 2012; Schmelter et al., 2011; Schmelter & Stevens, 2013; Schmelter et al., 2015) use a process-based formulation and implement a stochastic description of the variables. Those models have a better guarantee to be accurate, but will have to neglect less relevant processes if they want to save computational speed. The approach in this thesis is the opposite that the developed model makes estimates based on pattern recognition it has of sediment pathway data of the system described using connectivity. This difference in approach is like a bottom up approach versus a top down approach. This makes the creating of a probabilistic model a trade-off between accuracy and computational speed.

It is expected that more covariates or autoregressive dependencies would have to be added to improve accuracy of the model. This will inevitably increase computational time, but depending on the implementation it could still be orders of magnitude faster than SedTRAILS. For a more detailed approach of analysing the computational cost of the model could be analysed in terms of the P vs NP problem. This is one of the millennium problems that is central to the field of computer science.

Based on the above conclusions, we can answer the main research question:

How can we use sediment pathways data of complex tidal systems to create a probabilistic model that can make fast and accurate estimates of those pathways using Markov chains?

Sediment pathways data created using all components of a full wave climate can be used with a connectivity description of the system to create the emission matrix of an *Auto-regressive Hidden Markov Model* (ARHMM). The accuracy of this model is still limited, but has great computational speed in the range of milliseconds. This is a very promising outlook in that further developments could become more accurate at estimating sediment pathways coastal systems.

As recommendations for further research the points in the discussion are things that can be experimented with to improve the accuracy of estimates created using the model. Also for more research into the used ARHMM structure *regime switching Markov chains* could be looked into. This could deepen understanding of the model and uncover possible limitations.

References

Alagoz, O., Hsu, H., Schaefer, A. J., & Roberts, M. S. (2010). Markov decision processes: A tool for sequential decision making under uncertainty. *Medical Decision Making*, *30*(4), 474–483. https://doi.org/10.1177/0272989x09353194

Althoen, S. C., King, L., & Schilling, K. (1993). - how long is a game of snakes and ladders? *77*(478), 76. https://doi.org/https://doi.org/10.2307/3619261

Antolínez, J. A. A., Méndez, F. J., Camus, P., Vitousek, S., González, E. M., Ruggiero, P., & Barnard, P. (2016). A multiscale climate emulator for long-term morphodynamics (muscle-morpho). *Journal of Geophysical Research: Oceans*, *121*(1), 775–791. https://doi.org/https://doi.org/10.1002/2015JC011107

Aragón, M., Martín-Llanes, G., Zarzuelo, C., López-Ruiz, A., & Ortega-Sánchez, M. (2023). Wave schematization for coastal morphodynamics: The role of wave chronology. *Proceedings of the IAHR World Congress*, 501–508. https://doi.org/10.3850/978-90-833476-1-5_iahr40wc-p0628-cd

Ataharul Islam, M., & Chowdhury, R. I. (2006). A higher order markov model for analyzing covariate dependence. *Applied Mathematical Modelling*, *30*(6), 477–488. https://doi.org/https://doi.org/10.1016/j.apm.2005.05.006

Bacher, C., Filgueira, R., & Guyondet, T. (2016). Probabilistic approach of water residence time and connectivity using markov chains with application to tidal embayments. *Journal of Marine Systems*, *153*, 25–41. https://doi.org/10.1016/j.jmarsys.2015.09.002

Bellman, R. (1957). A markovian decision process. *Indiana University Mathematics Journal*, *6*(4), 679–684.

Benedet, L., Dobrochinski, J. P. F., Walstra, D. J. R., Klein, A. H. F., & Ranasinghe, R. (2016). A morphological modeling study to compare different methods of wave climate schematization and evaluate strategies to reduce erosion losses from a beach nourishment project. *Coastal Engineering*, *112*, 69–86. https://doi.org/10.1016/j.coastaleng.2016.02.005

Bosboom, J., & Stive, M. (2021). *Coastal dynamics*. TU Delft Open. https://doi.org/10.5074/T.2021.001

Bradbury, J., Frostig, R., Hawkins, P., Johnson, M. J., Leary, C., Maclaurin, D., Necula, G., Paszke, A., VanderPlas, J., & Wanderman-Milne, S. (2018). Jax: Composable transformations of python+ numpy programs.

Brand, E., Ramaekers, G., & Lodder, Q. (2022). Dutch experience with sand nourishments for dynamic coastline conservation - an operational overview. *Ocean & Coastal Management*, *217*. https://doi.org/10.1016/j.ocecoaman.2021.106008

Brunton, S. L., Noack, B. R., & Koumoutsakos, P. (2020). Machine learning for fluid mechanics. In S. H. Davis & P. Moin (Eds.), *Annual review of fluid mechanics, vol 52* (pp. 477–508). https://doi.org/10.1146/annurev-fluid-010719-060214

Bruun, P. (1954). *Coast erosion and the development of beach profiles* (Vol. 44). US Beach Erosion Board.

Bruun, P. (1962). Sea-level rise as a cause of shore erosion. *Journal of the Waterways and Harbors division*, *88*(1), 117–130.

Bruun, P., & Gerritsen, F. (1959). Natural by-passing of sand at coastal inlets. *Journal of the Waterways and Harbors Division*, *85*(4), 75–107. https://doi.org/10.1061/JWHEAU.0000152

Chassagne, C., & Safar, Z. (2020). Modelling flocculation: Towards an integration in large-scale sediment transport models. *Marine Geology*, *430*. https://doi.org/10.1016/j.margeo.2020.106361

Cook, S. (2000). The p versus np problem. *Clay Mathematics Institute*, *2*(6), 3.

Davis, R. A., & Hayes, M. O. (1984). What is a wave-dominated coast? *Marine Geology*, *60*(1), 313–329. https://doi.org/https://doi.org/10.1016/0025-3227(84)90155-5

De Fockert, A. (2008). *Impact of relative sea level rise on the amelander inlet morphology* (Thesis). http://resolver.tudelft.nl/uuid:5fd29cd3-8673-4319-89b1-a758703a8d36

De Vriend, H. J. (1991). Mathematical-modeling and large-scale coastal behavior .1. physical processes. *Journal of Hydraulic Research*, *29*(6), 727–740. https://doi.org/10.1080/00221689109498955

Deltares. (2023). *Delft3d-flow user manual* (Report). https://oss.deltares.nl/web/delft3d/manuals

Ekanayake, S. P., & Zois, D. S. (2022). Sequential bayesian network structure learning. *56th Asilomar Conference on Signals, Systems, and Computers*, 76–80. https://doi.org/10.1109/ieeeconf56349.2022.10052016

Elias, E. P. L., & Hansen, J. E. (2013). Understanding processes controlling sediment transports at the mouth of a highly energetic inlet system (san francisco bay, ca). *Marine Geology*, *345*, 207–220. https://doi.org/10.1016/j.margeo.2012.07.003

Elias, E. P. L., & van der Spek, A. J. F. (2017). Dynamic preservation of texel inlet, the netherlands: Understanding the interaction of an ebb-tidal delta with its adjacent coast. *Netherlands Journal of Geosciences-Geologie En Mijnbouw*, *96*(4), 293–317. https://doi.org/10.1017/njg.2017.34

Elias, E. P. L., Van der Spek, A. J. F., Pearson, S. G., & Cleveringa, J. (2019). Understanding sediment bypassing processes through analysis of high-frequency observations of ameland inlet, the netherlands. *Marine Geology*, *415*, 105956. https://doi.org/https://doi.org/10.1016/j.margeo.2019.06.001

Elias, E. P. L., & van der Spek, A. J. F. (2006). Long-term morphodynamic evolution of texel inlet and its ebb-tidal delta (the netherlands). *Marine Geology*, *225*(1), 5–21. https://doi.org/https://doi.org/10.1016/j.margeo.2005.09.008

Escoffier, F. F. (1940). The stability of tidal inlets. *Shore and beach*, *8*(4), 114–115.

Everitt, B. (2006). The cambridge dictionary of statistics. http://site.ebrary.com/id/10150287

FitzGerald, D. M., Hubbard, D. K., & Nummedal, D. (1978). Shoreline changes associated with tidal inlets along the south carolina coast. *Coastal Zone'78*, 1973–1994.

Glas, P. C. G. (2023). Nationaal deltaprogramma 2024.

Goldstein, E. B., Coco, G., & Plant, N. G. (2019). A review of machine learning applications to coastal sediment transport and morphodynamics. *Earth-Science Reviews*, *194*, 97–108. https://doi.org/https://doi.org/10.1016/j.earscirev.2019.04.022

Gonzales, C., Torti, L., & Wuillemin, P. H. (2017). Agrum: A graphical universal model framework. *30th International Conference on Industrial Engineering and Other Applications of Applied Intelligent Systems (IEA/AIE)*, *10351*, 171–177. https://doi.org/10.1007/978-3-319-60045-1_20

Hamilton, J. D. (1990). Analysis of time-series subject to changes in regime. *Journal of Econometrics*, *45*(1-2), 39–70. https://doi.org/10.1016/0304-4076(90)90093-9

Herrling, G., & Winter, C. (2018). Tidal inlet sediment bypassing at mixed-energy barrier islands. *Coastal Engineering*, *140*, 342–354. https://doi.org/10.1016/j.coastaleng.2018.08.008

Hirano, M. (1971). River-bed degradation with armoring. *Proceedings of the Japan Society of Civil Engineers*, 55–65.

Hirschberg, P. A., Abrams, E., Bleistein, A., Bua, W., Delle Monache, L., Dulong, T. W., Gaynor, J. E., Glahn, B., Hamill, T. M., Hansen, J. A., Hilderbrand, D. C., Hoffman, R. N., Morrow, B. H., Philips, B., Sokich, J., & Stuart, N. (2011). A weather and climate enterprise strategic implementation plan for generating and communicating forecast uncertainty information. *Bulletin of the American Meteorological Society*, *92*(12), 1651–1666. https://doi.org/10.1175/bams-d-11-00073.1

Hoagland, S. W. H., Jeffries, C. R., Irish, J. L., Weiss, R., Mandli, K., Vitousek, S., Johnson, C. M., & Cialone, M. A. (2023). Advances in morphodynamic modeling of coastal barriers: A review. *Journal of Waterway Port Coastal and Ocean Engineering*, *149*(5). https://doi.org/10.1061/jwped5.Wweng-1825

Hoonhout, B. M., & de Vries, S. (2016). A process-based model for aeolian sediment transport and spatiotemporal varying sediment availability. *Journal of Geophysical Research-Earth Surface*, *121*(8), 1555–1575. https://doi.org/10.1002/2015jf003692

Howard, R. (1960). *Dynamic programming and markov processes*. Technology Press of Massachusetts Institute of Technology. https://books.google.nl/books?id=fXJEAAAAIAAJ

Huang, C., & Darwiche, A. (1996). Inference in belief networks: A procedural guide. *International Journal of Approximate Reasoning*, *15*(3), 225–263. https://doi.org/10.1016/s0888-613x(96)00069-2

Ibe, O. C. (2013). 3 - introduction to markov processes. In O. C. Ibe (Ed.), *Markov processes for stochastic modeling (second edition)* (pp. 49–57). Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-12-407795-9.00003-7

Jax. (2024). Jax - the sharp bits. https://jax.readthedocs.io/en/latest/notebooks/Common_Gotchas_in_JAX.html#rngs-and-state

Jongejan, R., Ranasinghe, R., Wainwright, D., Callaghan, D. P., & Reyns, J. (2016). -%20https://www.sciencedirect.com/science/article/pii/S0964569116300060

Kroon, J. (2024). *Uncertainty in predictions of large-scale sandy interventions*. Delft University of Technology. https://doi.org/https://doi.org/10.4233/uuid:6b572d9e-481a-4be6-815b-dc42647e6feb

Latteux, B. (1995). Techniques for long-term morphological simulation under tidal action. *Marine Geology*, *126*(1-4), 129–141. https://doi.org/10.1016/0025-3227(95)00069-b

Lesser, G. R., Roelvink, J. A., van Kester, J., & Stelling, G. S. (2004). Development and validation of a three-dimensional morphological model. *Coastal Engineering*, *51*(8-9), 883–915. https://doi.org/10.1016/j.coastaleng.2004.07.014

Linderman, S. W. (2024). Dynamax. https://github.com/probml/dynamax

Lodder, Q., & Slinger, J. (2022). The 'research for policy' cycle in dutch coastal flood risk management: The coastal genesis 2 research programme. *Ocean & Coastal Management*, *219*. https://doi.org/10.1016/j.ocecoaman.2022.106066

Madsen, A. L. (2006). Variations over the message computation algorithm of lazy propagation. *Ieee Transactions on Systems Man and Cybernetics Part B-Cybernetics*, *36*(3), 636–648. https://doi.org/10.1109/tsmcb.2005.862488

Malliouri, D. I., Petrakis, S., Vandarakis, D., Moraitis, V., Goulas, T., Hatiris, G. A., Drakopoulou, P., & Kapsimalis, V. (2023). A chronology-based wave input reduction technique for simulations of long-term coastal morphological changes: An application to the beach of mastichari, kos island, greece. *Water (Switzerland)*, *15*(3). https://doi.org/10.3390/w15030389

Matsumoto, M., Saito, M., Haramoto, H., & Nishimura, T. (2006). Pseudorandom number generation: Impossibility and compromise. *Journal of Universal Computer Science*, *12*(6), 672–690. %3CGo%20to%20ISI%3E://WOS:000239790000006

Mulhern, J. S., Johnson, C. L., & Martin, J. M. (2017). Is barrier island morphology a function of tidal and wave regime? *Marine Geology*, *387*, 74–84. https://doi.org/10.1016/j.margeo.2017.02.016

Murphy, K. P. (2002). *Dynamic bayesian networks: Representation, inference and learning* (Thesis).

Murphy, K. P. (2023). *Probabilistic machine learning: Advanced topics*. MIT Press. http://probml.github.io/book2

Newman, M. E. J. (2003). The structure and function of complex networks. *Siam Review*, *45*(2), 167–256. https://doi.org/10.1137/s003614450342480

O'Brien, M. P. (1931). Estuary tidal prisms related to entrance areas. *Civil Engineering*, *1*, 738–739.

O'Brien, M. P. (1969). Equilibrium flow areas of inlets on sandy coasts. *Journal of the Waterways and Harbors Division*, *95*(1), 43–52. https://doi.org/doi:10.1061/JWHEAU.0000622

Pearl, J. (1988). Chapter 3 - markov and bayesian networks: Two graphical representations of probabilistic knowledge. In J. Pearl (Ed.), *Probabilistic reasoning in intelligent systems* (pp. 77–141). Morgan Kaufmann. https://doi.org/https://doi.org/10.1016/B978-0-08-051489-5.50009-6

Pearson, S. G. (2022). *Sediment pathways on ebb-tidal deltas: New tools and techniques for analysis* (Thesis). https://doi.org/https://doi.org/10.4233/uuid:c2fe811c-dc2e-4e1f-bb0c-dc43f11cd1eb

Pearson, S. G., Prooijen, B., de Wit, F., Holzhauer, H., Looff, A. P., & Wang, Z. B. (2019). *Observations of suspended particle size distribution on an energetic ebb-tidal delta*. https://doi.org/10.1142/9789811204487_0172

Pearson, S. G., Prooijen, B. C., Elias, E. P. L., Vitousek, S., & Wang, Z. B. (2020). Sediment connectivity: A framework for analyzing coastal sediment transport pathways. *Journal of Geophysical Research-Earth Surface*, *125*(10), 25. https://doi.org/10.1029/2020jf005595

Pearson, S. G., Reniers, A., & van Prooijen, B. (2023). Unveiling the hidden skeleton of coastal sediment pathways with lagrangian coherent structures. *Coastal Sediments 2023, New Orleans, Louisiana.*

Ranasinghe, R. (2020). On the need for a new generation of coastal change models for the 21st century. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-58376-x

Ranasinghe, R., Duong, T. M., Uhlenbrook, S., Roelvink, D., & Stive, M. (2013). Climate-change impact assessment for inlet-interrupted coastlines. *Nature Climate Change*, *3*(1), 83–87. https://doi.org/10.1038/nclimate1664

Ridderinkhof, W., de Swart, H. E., van der Vegt, M., & Hoekstra, P. (2016). Modeling the growth and migration of sandy shoals on ebb-tidal deltas. *Journal of Geophysical Research: Earth Surface*, *121*(7), 1351–1372. https://doi.org/https://doi.org/10.1002/2016JF003823

Rijkswaterstaat. (2024). Waterinfo. https://waterinfo.rws.nl/#/nav/index

Ris, R. C. (1997). Spectral modelling of wind waves in coastal areas. *Communications on Hydraulic and Geotechnical Engineering - Delft University of Technology*, *97-4*. https://www.scopus.com/inward/record.uri?eid=2-s2.0-0030880373%5C&partnerID=40%5C&md5=d3934d2498bbc627897ab4cc3c465d81

Schmelter, M. L., Erwin, S. O., & Wilcock, P. R. (2012). Accounting for uncertainty in cumulative sediment transport using bayesian statistics. *Geomorphology*, *175*, 1–13. https://doi.org/10.1016/j.geomorph.2012.06.012

Schmelter, M. L., Hooten, M. B., & Stevens, D. K. (2011). Bayesian sediment transport model for unisize bed load. *Water Resources Research*, *47*. https://doi.org/10.1029/2011wr010754

Schmelter, M. L., & Stevens, D. K. (2013). Traditional and bayesian statistical models in fluvial sediment transport. *Journal of Hydraulic Engineering*, *139*(3), 336–340. https://doi.org/10.1061/(asce)hy.1943-7900.0000672

Schmelter, M. L., Wilcock, P., Hooten, M., & Stevens, D. K. (2015). Multi-fraction bayesian sediment transport model. *Journal of Marine Science and Engineering*, *3*(3), 1066–1092. https://doi.org/10.3390/jmse3031066

Schuttelaars, H. M., & de Swart, H. E. (2000). Multiple morphodynamic equilibria in tidal embayments. *Journal of Geophysical Research-Oceans*, *105*(C10), 24105–24118. https://doi.org/10.1029/2000jc900110

Soulsby, R. L., Mead, C. T., Wild, B. R., & Wood, M. J. (2011). Lagrangian model for simulating the dispersal of sand-sized particles in coastal waters. *Journal of Waterway Port Coastal and Ocean Engineering-Asce*, *137*(3), 123–131. https://doi.org/10.1061/(asce)ww.1943-5460.0000074

Southgate, H. N. (1995). The effects of wave chronology on medium and long term coastal morphology. *Coastal Engineering*, *26*(3-4), 251–270. https://doi.org/10.1016/0378-3839(95)00028-3

Spearman, J. R., & Roberts, W. (2000). Comparison of flocculation models for applied sediment transport modelling. *6th International Conference on Cohesive Sediment Transport (INTERCOH 2000)*, *5*, 277–293. %3CGo%20to%20ISI%3E://WOS:000177780800019

Stive, M., de Schipper, M., Luijendijk, A., Ranasinghe, R., de Vries, J. V., & Aarninkhof, S. (2013). The sand engine: A solution for the dutch delta in the 21st century? *35th World Congress of the International-Association-for-Hydro-Environment-Engineering-and-Research (IAHR)*, 252–260. %3CGo%20to%20ISI%3E://WOS:000343761500024

van Ijzendoorn, C. O., Hallin, C., Reniers, A., & de Vries, S. (2023). Modeling multi-fraction coastal aeolian sediment transport with horizontal and vertical grain-size variability. *Journal of Geophysical Research-Earth Surface*, *128*(7). https://doi.org/10.1029/2023jf007155

van Prooijen, B. C., Tissier, M. F. S., de Wit, F. P., Pearson, S. G., Brakenhoff, L. B., van Maarseveen, M. C. G., van der Vegt, M., Mol, J. W., Kok, F., Holzhauer, H., van der Werf, J. J., Vermaas, T., Gawehn, M., Grasmeijer, B., Elias, E. P. L., Tonnon, P. K., Santinelli, G., Antolínez, J. A. A., de Vet, P. L. M., … de Looff, H. (2020). Measurements of hydrodynamics, sediment, morphology and benthos on ameland ebb-tidal delta and lower shoreface. *Earth System Science Data*, *12*(4), 2775–2786. https://doi.org/10.5194/essd-12-2775-2020

van Weerdenburg, R., Pearson, S. G., van Prooijen, B., Laan, S., Elias, E., Tonnon, P. K., & Wang, Z. B. (2021). Field measurements and numerical modelling of wind-driven exchange flows in a tidal inlet system in the dutch wadden sea. *Ocean & Coastal Management*, *215*, 105941. https://doi.org/https://doi.org/10.1016/j.ocecoaman.2021.105941

Vigna, S. (2019). It is high time we let go of the mersenne twister. *arXiv preprint arXiv:1910.06437*.

Wainwright, D. J., Ranasinghe, R., Callaghan, D. P., Woodroffe, C. D., Jongejan, R., Dougherty, A. J., Rogers, K., & Cowell, P. J. (2015). -%20https://www.sciencedirect.com/science/article/pii/S0378383914002130

Walton Jr, T. L., & Adams, W. D. (1976). Capacity of inlet outer bars to store sand. In *Coastal engineering 1976* (pp. 1919–1937).

Wang, Z. B., Townend, I., & Stive, M. (2020). Aggregated morphodynamic modelling of tidal inlets and estuaries. *Water Science and Engineering*, *13*(1), 1–13. https://doi.org/10.1016/j.wse.2020.03.004

Wang, Z. B., Elias, E. P. L., van der Spek, A. J. F., & Lodder, Q. J. (2018). Sediment budget and morphological development of the dutch wadden sea: Impact of accelerated sea-level rise and subsidence until 2100. *Netherlands Journal of Geosciences*, *97*(3), 183–214. https://doi.org/10.1017/njg.2018.8

Wright, L. D., & Short, A. D. (1984). Morphodynamic variability of surf zones and beaches - a synthesis. *Marine Geology*, *56*(1-4), 93–118. https://doi.org/10.1016/0025-3227(84)90008-2

Zhang, K., Douglas, B. C., & Leatherman, S. P. (2004). Global warming and coastal erosion. *Climatic Change*, *64*(1), 41–58. https://doi.org/10.1023/B:CLIM.0000024690.32682.48

A

Emission matrix and transition matrices

This appendix shows the created matrices for the *Auto-regressive Hidden Markov Model* (ARHMM) that were too big to be in the results section. First the transition matrices are shown and after that the emission matrix.

A.1. Transition matrices

Both transition matrices assume independence of wave climate components following each other up resulting in the same distribution for each row.

A.1.1. De Fockert transition matrix

	wc1												
wc0	0	1	2	3	4	5	6	7	8	9	10	11	12
0	0.0116	0.1758	0.1135	0.1321	0.2450	0.0822	0.0667	0.0435	0.0237	0.0735	0.0202	0.0099	0.0023
1	0.0116	0.1758	0.1135	0.1321	0.2450	0.0822	0.0667	0.0435	0.0237	0.0735	0.0202	0.0099	0.0023
2	0.0116	0.1758	0.1135	0.1321	0.2450	0.0822	0.0667	0.0435	0.0237	0.0735	0.0202	0.0099	0.0023
3	0.0116	0.1758	0.1135	0.1321	0.2450	0.0822	0.0667	0.0435	0.0237	0.0735	0.0202	0.0099	0.0023
4	0.0116	0.1758	0.1135	0.1321	0.2450	0.0822	0.0667	0.0435	0.0237	0.0735	0.0202	0.0099	0.0023
5	0.0116	0.1758	0.1135	0.1321	0.2450	0.0822	0.0667	0.0435	0.0237	0.0735	0.0202	0.0099	0.0023
6	0.0116	0.1758	0.1135	0.1321	0.2450	0.0822	0.0667	0.0435	0.0237	0.0735	0.0202	0.0099	0.0023
7	0.0116	0.1758	0.1135	0.1321	0.2450	0.0822	0.0667	0.0435	0.0237	0.0735	0.0202	0.0099	0.0023
8	0.0116	0.1758	0.1135	0.1321	0.2450	0.0822	0.0667	0.0435	0.0237	0.0735	0.0202	0.0099	0.0023
9	0.0116	0.1758	0.1135	0.1321	0.2450	0.0822	0.0667	0.0435	0.0237	0.0735	0.0202	0.0099	0.0023
10	0.0116	0.1758	0.1135	0.1321	0.2450	0.0822	0.0667	0.0435	0.0237	0.0735	0.0202	0.0099	0.0023
11	0.0116	0.1758	0.1135	0.1321	0.2450	0.0822	0.0667	0.0435	0.0237	0.0735	0.0202	0.0099	0.0023
12	0.0116	0.1758	0.1135	0.1321	0.2450	0.0822	0.0667	0.0435	0.0237	0.0735	0.0202	0.0099	0.0023

Figure A.1: The 13x13 transition matrix using the weight distribution from De Fockert (2008) also found in Figure 3.3.

A.1.2. Poisson transition matrix

	wc1												
wc0	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
0	0.0180	0.0688	0.1468	0.1989	0.1981	0.1548	0.1025	0.0610	0.0289	0.0131	0.0061	0.0020	0.0010
1	0.0180	0.0688	0.1468	0.1989	0.1981	0.1548	0.1025	0.0610	0.0289	0.0131	0.0061	0.0020	0.0010
2	0.0180	0.0688	0.1468	0.1989	0.1981	0.1548	0.1025	0.0610	0.0289	0.0131	0.0061	0.0020	0.0010
3	0.0180	0.0688	0.1468	0.1989	0.1981	0.1548	0.1025	0.0610	0.0289	0.0131	0.0061	0.0020	0.0010
4	0.0180	0.0688	0.1468	0.1989	0.1981	0.1548	0.1025	0.0610	0.0289	0.0131	0.0061	0.0020	0.0010
5	0.0180	0.0688	0.1468	0.1989	0.1981	0.1548	0.1025	0.0610	0.0289	0.0131	0.0061	0.0020	0.0010
6	0.0180	0.0688	0.1468	0.1989	0.1981	0.1548	0.1025	0.0610	0.0289	0.0131	0.0061	0.0020	0.0010
7	0.0180	0.0688	0.1468	0.1989	0.1981	0.1548	0.1025	0.0610	0.0289	0.0131	0.0061	0.0020	0.0010
8	0.0180	0.0688	0.1468	0.1989	0.1981	0.1548	0.1025	0.0610	0.0289	0.0131	0.0061	0.0020	0.0010
9	0.0180	0.0688	0.1468	0.1989	0.1981	0.1548	0.1025	0.0610	0.0289	0.0131	0.0061	0.0020	0.0010
10	0.0180	0.0688	0.1468	0.1989	0.1981	0.1548	0.1025	0.0610	0.0289	0.0131	0.0061	0.0020	0.0010
11	0.0180	0.0688	0.1468	0.1989	0.1981	0.1548	0.1025	0.0610	0.0289	0.0131	0.0061	0.0020	0.0010
12	0.0180	0.0688	0.1468	0.1989	0.1981	0.1548	0.1025	0.0610	0.0289	0.0131	0.0061	0.0020	0.0010

Figure A.2: The 13x13 transition matrix using a distribution sampled from a Poisson distribution with $\lambda = 4$.

A.2. Emission matrix

The 13x5x5 emission matrix determined by running Delft3D for the area of Ameland inlet for different wave climate components of the wave climate schematization by De Fockert (2008) and using these hydrodynamic to compute the sediment pathways using SedTRAILS and turning those into transition matrices.

		pos3				
wc3	pos2	1	2	3	4	5
0	1	0.9848	0.0152	0.0000	0.0000	0.0000
	2	0.0133	0.9735	0.0088	0.0044	0.0000
	3	0.0000	0.0050	0.9950	0.0000	0.0000
	4	0.0000	0.0196	0.0000	0.9356	0.0448
	5	0.0000	0.0050	0.0000	0.0040	0.9909
1	1	0.9686	0.0314	0.0000	0.0000	0.0000
	2	0.0268	0.9598	0.0134	0.0000	0.0000
	3	0.0000	0.0000	1.0000	0.0000	0.0000
	4	0.0000	0.0336	0.0000	0.9048	0.0616
	5	0.0000	0.0081	0.0000	0.0040	0.9879
2	1	0.9738	0.0262	0.0000	0.0000	0.0000
	2	0.0268	0.9420	0.0268	0.0000	0.0045
	3	0.0000	0.0000	1.0000	0.0000	0.0000
	4	0.0000	0.0280	0.0000	0.9048	0.0672
	5	0.0000	0.0071	0.0000	0.0030	0.9899
3	1	0.9834	0.0166	0.0000	0.0000	0.0000
	2	0.0274	0.9406	0.0320	0.0000	0.0000
	3	0.0000	0.0000	1.0000	0.0000	0.0000
	4	0.0000	0.0393	0.0000	0.8764	0.0843
	5	0.0000	0.0081	0.0000	0.0040	0.9879

Figure A.3: Emission matrix of the ARHMM. Part (1/3).

4	1	0.9415	0.0585	0.0000	0.0000	0.0000
	2	0.0231	0.8796	0.0926	0.0000	0.0046
	3	0.0000	0.0000	1.0000	0.0000	0.0000
	4	0.0000	0.0672	0.0000	0.8824	0.0504
	5	0.0000	0.0112	0.0000	0.0071	0.9817
5	1	0.9677	0.0323	0.0000	0.0000	0.0000
	2	0.0411	0.9315	0.0274	0.0000	0.0000
	3	0.0000	0.0000	1.0000	0.0000	0.0000
	4	0.0000	0.0532	0.0000	0.8739	0.0728
	5	0.0000	0.0081	0.0000	0.0050	0.9869
6	1	0.9628	0.0372	0.0000	0.0000	0.0000
	2	0.0664	0.7441	0.1801	0.0000	0.0095
	3	0.0000	0.0000	1.0000	0.0000	0.0000
	4	0.0000	0.0927	0.0000	0.8315	0.0758
	5	0.0000	0.0061	0.0000	0.0101	0.9838
7	1	0.9503	0.0497	0.0000	0.0000	0.0000
	2	0.0711	0.7773	0.1422	0.0000	0.0095
	3	0.0000	0.0000	1.0000	0.0000	0.0000
	4	0.0000	0.0927	0.0000	0.8343	0.0730
	5	0.0000	0.0071	0.0000	0.0101	0.9828

Figure A.4: Emission matrix of the ARHMM. Part (2/3).

8	1	0.9157	0.0843	0.0000	0.0000	0.0000
	2	0.0483	0.6715	0.2609	0.0000	0.0193
	3	0.0000	0.0000	1.0000	0.0000	0.0000
	4	0.0029	0.1236	0.0000	0.8247	0.0489
	5	0.0000	0.0071	0.0000	0.0343	0.9586
9	1	0.9314	0.0686	0.0000	0.0000	0.0000
	2	0.0631	0.6408	0.2767	0.0000	0.0194
	3	0.0000	0.0000	1.0000	0.0000	0.0000
	4	0.0057	0.1133	0.0000	0.8244	0.0567
	5	0.0000	0.0081	0.0000	0.0253	0.9666
10	1	0.9583	0.0417	0.0000	0.0000	0.0000
	2	0.1330	0.5222	0.3350	0.0000	0.0099
	3	0.0000	0.0000	1.0000	0.0000	0.0000
	4	0.0056	0.1155	0.0000	0.7944	0.0845
	5	0.0000	0.0111	0.0000	0.0091	0.9798
11	1	0.6489	0.3511	0.0000	0.0000	0.0000
	2	0.0400	0.2914	0.6571	0.0057	0.0057
	3	0.0000	0.0196	0.9804	0.0000	0.0000
	4	0.0089	0.1361	0.0000	0.7722	0.0828
	5	0.0000	0.0153	0.0020	0.0529	0.9298
12	1	0.5385	0.4462	0.0154	0.0000	0.0000
	2	0.0901	0.3784	0.5315	0.0000	0.0000
	3	0.0000	0.0000	1.0000	0.0000	0.0000
	4	0.0140	0.1261	0.0000	0.7843	0.0756
	5	0.0000	0.0163	0.0010	0.0316	0.9511

Figure A.5: Emission matrix of the ARHMM. Part (3/3).