Erasmus University Rotterdam

Master Thesis

Econometrics & Management Science

Improving Ship Motion Forecasts by Combining Measurement Data with Physical Model Forecasts

Author:

Pieter de Wet*

External Supervisor:

Mark Paalvast

Supervisor:

Andreas Pick

Second Reader:

Richard Paap

October 16, 2018

ERASMUS UNIVERSITEIT ROTTERDAM

NEXT LEVEL FORECASTING

ADVANCED OFFSHORE TECHNOLOGY

**Student ID: 427863*

Abstract

This paper compares models to improve ship motion forecasting performance on 4, 8 and 12 hours ahead forecasts by combining time series measurement data with the currently used physical model forecasts. The first is a simple component model which serves as a basic benchmark. The second is a local level model where the physical model forecast is included as a component in the state equation of which we consider two variants: one where the weighting coefficients for the components in the state equation are estimated upfront and kept fixed and one where these coefficients are re-estimated during the forecasting exercise. The third is a dynamic mixture model where the physical model is included in the same way as in the local level model, but this specification allows for level shifts in order to react to peak responses. In both the second and the third model the parameters are estimated by the Bayesian technique of Gibbs sampling. The performance is measured by the average root mean square error over the forecasting exercise. The simple component model does not improve the performance compared to the physical model forecasts. The local level model with re-estimated coefficients leads to improvements in the 4 and 8 hour forecasting horizons but fails to do so for 12 hours ahead. The strongest forecasting improvements of 14-44% are found using the local level model with pre-estimated coefficients. This model performs better than both the dynamic mixture model with fixed and with varying coefficients, which find improvements of 11-41% and 11-36% respectively. The performances of the models hold when tested on multiple subsets of data, which indicates that finding the improvement is not a coincidence. The local level model is also more suitable for implementation due to advantages in computing time, sensitivity to prior parameters and the data length needed to evaluate the model than the dynamic mixture models. Therefore it is the recommended method to use as basis when considering implementation in the on-board software to improve the ship motion forecasting performance.

Keywords: Time series forecasting; State space model; Gibbs sampling; Kalman filter.

Contents

1 Introduction

Offshore operations are performed by complex vessels, tailor made to do a specific task such as pipe laying, cable laying, or lifting a wind turbine. The vessels are very costly in operation, ranging from 50k to 1 mln euros a day. This means that maximizing the share of time in which actual work can be done, known as workability, is key, as every hour that can be saved during operation translates into thousands of euros. Workability is decreased due to ship motions and operating at maximum potential asks for delicate planning and execution to account for the motions. Watching the various environmental loadings, fine-tuning vessel-headings and even having meteorologists on board is part of this practice. To really be able to react on the environment, and therefore increase the workability, it is crucial to have excellent expectations upfront. This is where ship motion forecasts come into play.

A novel system has been designed by MO4, which not only generates forecasts but enables the captain to visually evaluate the forecasts based on different meteorological data suppliers. The MO4 system translates the spectral energy densities supplied by the meteorologists into ship motions. It does so by using physical models based on hydrodynamic theory, which is known to be very accurate due to extensive model scale tests and theory development. The problem is, the theory works on *known incoming loadings* where in this setting we only have the meteorological forecasts which turn out to fluctuate in quality and accuracy from time to time and also between one supplier and another. Therefore the accuracy of the MO4 forecasts are strongly dependent on the precision of the meteorological forecasts. Improving the metocean forecasts is not an objective of this research, as meteorology is a specialization of its own. Instead, we take on the view that we have an unobserved state of which the measurements are realizations through a state space process, and the metocean forecasts may play a role in estimating the state of the underlying system. In this way, the measured data is integrated into the forecasting process. This link was missing before.

In this research we investigate whether it is possible to improve the near future MO4 forecasts (4, 8 and 12 hours ahead) by combining the physical model forecasts made by MO4 with time series data of ship motions. For this purpose we design multiple models with varying complexity and different characteristics and compare them by their forecasting capabilities and suitability for implementation.

Multiple period ahead out of sample forecasts are generated for each time period in the sample data until the end of the data set. To compare the models the average root mean square error over the forecasting exercise is calculated to express the forecasting performance. To determine the suitability of each model, they are judged on their computing time, their robustness to different data and on the amount of data they need to run properly. This is because the software has to run on-board and should be interactive. It has to work when the vessel operates in various conditions leading to different data and it should work as quick as possible after the operation starts with minimum time for data collection.

The first model we use is a simple component model based on the measurements and the significant wave height forecast. Its advantages lie in its modeling simplicity and computing time.

The second model is based on a local level model where the physical model forecast data is included through a component of the state equation. It is split into two sub models: one where the weighting coefficients for the components in the state equation are estimated beforehand using data of an earlier operating period and then kept fixed during the forecasting exercise. In the other model these coefficients are re-estimated during the forecasting exercise. Re-estimating the coefficients leads to more flexible weighting of the physical model and the time series data, where fixing the coefficients leads to more consistent forecasting results. Also, when using fixed coefficients the model is better able to cope with small data sets, while the model fails to find proper results when re-estimating the coefficients when the sample gets small.

In the third model we add the possibility for the state to shift in level, as this can improve the capability of the model to account for heading or weather changes which result in strongly different vessel responses. The shifts in level are modeled and estimated following the efficient Bayesian inference approach for dynamic mixture models given by Gerlach, Carter, and Kohn (2000). This model needs 20 times more computing time than the first variant of the local level model and almost 40 times more than the second with fixed coefficients. The dynamic mixture model is also sensitive to prior settings and fails to work properly on small sample sets. Fixing the weighting coefficients in the dynamic mixture model increases the speed of estimating the model but the general disadvantages of estimating the level shifts, discussed above, remain.

The simple component model does not improve the forecasts. The local level model with re-estimation of the weighting coefficients finds improvements in the shorter horizons of 5-21% but does not improve the situation when forecasting for 12 hours ahead in this sample set. The local level model used with fixed weighting coefficients gives the best forecasting results with improvements compared to the physical model of 14-44%. The dynamic mixture model with re-estimated weighting coefficients finds improvements of 11-36% and 11-41% with fixed coefficients. The results depend on the forecasting horizon and the physical model used, where in general stronger improvements are found for the shorter forecasting horizons. The results hold when the models are tested on multiple different data subsets.

In section 2 the data at the attention of this research is discussed. The choice of the models and their details are discussed in section 3. The estimation procedure of each of the models is treated in section 4. The way we evaluate the models, and how we determine the performance is discussed in section 5. Then the results of all models are given in section 6 followed by a discussion of the results where we focus on the characteristics of the models in section 7. Finally we conclude the research and give recommendations for further research in section 8.

2 Data

The strength of this research is that it aims to integrate multiple sources of data to get to more accurate forecasts of ship motions. In this section the characteristics of the different data sources will be discussed. We will look at the the measurement data of the ship motions, the metocean supply data and the forecasts resulting from the physical model. The ship motions are measured constantly, while the metocean supply data is given for certain coordinates, which is the location of operation. Therefore only the coinciding part of the data, where the ship is in the operating zone, is relevant, we call this the operating window in this report.

2.1 The Measurement Data

The ships have measuring instruments installed which produce various types of measurement data. Ship motion has 6 degrees of freedom, as it can translate in 3 directions (heave, surge and sway) and it can rotate in 3 directions (yaw, roll and pitch). See figure 1. The motions are measured in terms of displacement, velocities and accelerations.

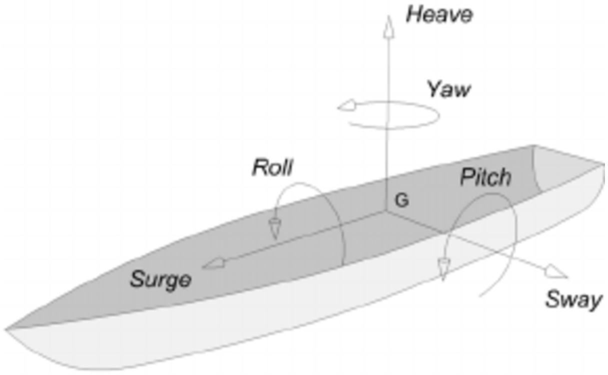

Figure 1: Degrees of freedom of ship motions
Source: Varela and Soares (2011)

We are interested in the limits for certain operations of the ship, for instance for a cable laying operation. For this it is necessary to know the motions of the specific point on the ship where the cable travels over the chute and critical stresses arise which could cause damage in the cable. In many cases the limits are determined using the vertical velocity, as from there one can derive the displacements as well as the accelerations. For this research we will therefore look at the vertical velocity, also known as the heave-velocity.

The actual heave velocity is measured at high frequency (20Hz), however to evaluate longer term ship motions a 15 minute standard deviation is calculated and reported. This indicates the 'volatility' of the movement and therefore indicates the risk level, to bring it to econometric terms. We will use this heave-velocity standard deviation (σ_{v_z}) in the models. This data can be seen in Figure 3 for one operating window, which we refer to as operating window 2, with 435 observations.

In the graph we see that the time series fluctuates strongly around various levels, in this sense it reflects the sea state, or the energy in the sea. The sea state does not have fixed levels, but it can stay close to a certain level for a longer period. The motions are also dependent on the heading of the ship, which for instance becomes visible in the sharp breaks in the level between Jan 21 and Jan 22: If a ship changes heading, this strongly influences the heave motions, this influence can be seen in Figure 29 in appendix A.3.

If we disregard the physical influences on the data and view it in a time series context, it seems to behave very unpredictable: it changes level often, and goes up and down seemingly without coherence. In this sense it has the appearance of a random walk series,

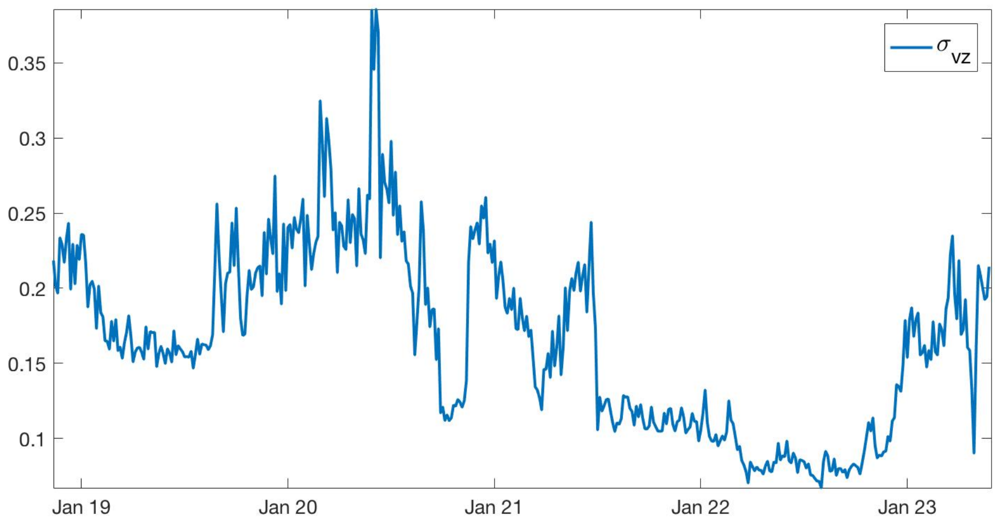

Figure 3: Heave velocity standard deviation for one operating window

and we investigate the possibility of a unit root. An Augmented Dicky Fuller test is conducted on the univariate time series of the observations of operating period 2 specified above. For this test we find that a unit root is present with a p-value of 0.1628, tested at the 0.05 significance level. This indicates that we are looking at data that has non-stationary characteristics, this can lead to assumptions for analysis not longer being valid (t-ratios can not be used as normal) and it can lead to high R^2 values without actual correlation in the data.

2.2 The Metocean Data

For decades or longer it has been a challenge to forecast the weather. It is a complex system and it is mainly modeled by institutional research instances such as the World Meteorological Organization. These models are available and are interpreted by various companies. These companies interpolate the larger scale models and are able to supply more accurate weather forecasts for specific regions. They also translate ‘weather’ to the dimensions we are interested in in this field of study, which is the energy in the waves. The energy is a function of the wave height. When looking at the sea there will not be only one wave we should take into account, but it can be viewed as a whole array of different wave heights and periods, which is a discrete model of the continuous real life situation. In this field of study the energy in this system is expressed by the significant wave height (H_s, subscript s stands for significant), accompanied by the zero crossing period (T_z, subscript z stands for zero crossing). The mentioned companies give forecasts for H_s and T_z which are in general used on board ships to determine workability. For the operation we are analyzing in this research, we have three such suppliers of metocean data, of which the forecasts for *Jan 19 - Jan 27*, which we refer to as operating window 2, are plotted in Figure 4.

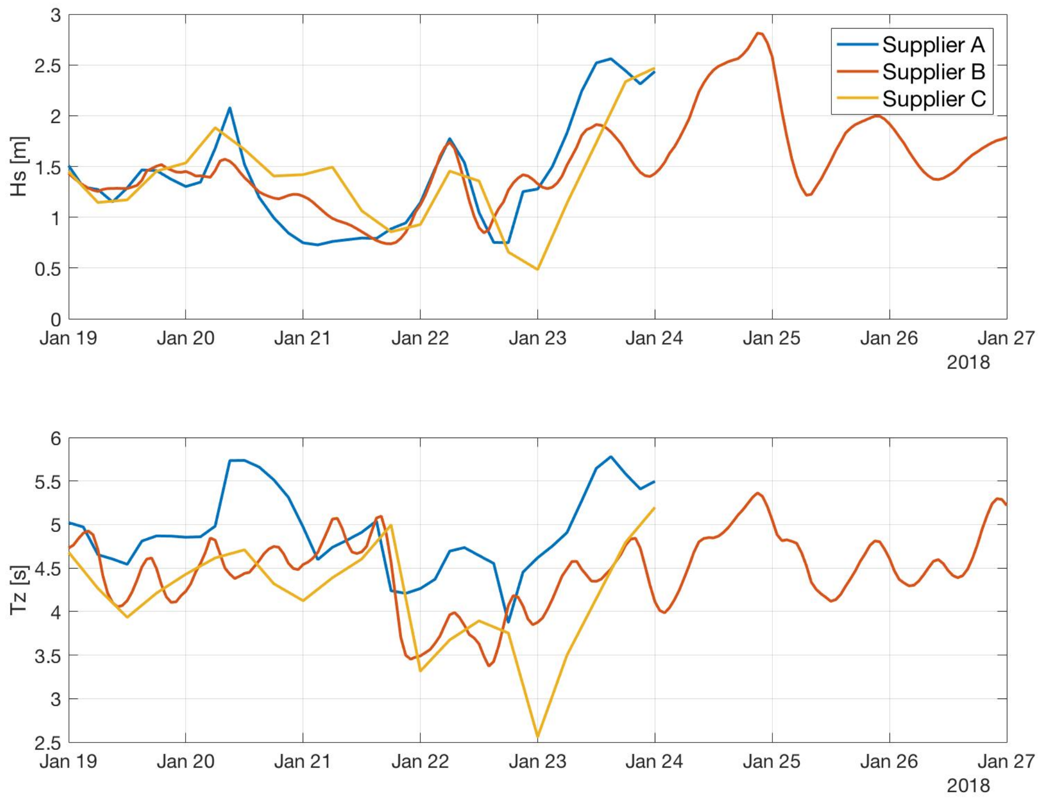

Figure 4: Significant wave height and wave period by 3 suppliers

We can clearly see differences in the characteristics of the forecasts. First of all, they have different forecasting horizons. Secondly they have a different updating frequency, and next to that they have different time interval between values which results in smoother lines and more variability. The details of these differences are given in table 1. Apart from the different characteristics it is also clearly visible that the suppliers do not give the same results. This means when using one or another, one might get a different conclusion for workability. This is the reason that in practice always multiple forecasts are used to make decisions on workability.

Table 1: Metocean data characteristics per supplier

Supplier	Interval	Periods	Horizon
A	3	41	5 days and 3 hours
B	1	193	8 days and 1 hour
C	6	21	5 days and 6 hours

The interval is the time in hours between the data points. The periods are the total amount of periods contained in a received dataset.

The hydrodynamic literature explains a linear relation between the significant wave height and the heave motions (Massie and Journée (2001)), therefore a relation can be expected between the ship motions and the forecasted H_s, depending on the characteristics of the vessel. This is where the physical model of MO4 comes in.

2.3 Forecasts from the Physical Model

As explained above, based on the forecasted H_s we may expect that the ship motions would follow in roughly the same way. In practice it is therefore common to use H_s and T_z to determine very roughly the expected motions. This is done by looking up the responses in tables, then from this it is decided whether workability limits are reached. However, a ship reacts very differently to these wave heights and periods when the heading of the ship change, as we have seen in Figure 3.

The underlying data used to generate the forecasts of H_s and T_z is actually an energy density expectation in 360 degrees. This is called a 2d spectral density, see Figure 5. The MO4 software uses this spectral density to evaluate the expected motions of the ship in each heading direction, and therefore gives a more accurate and usable expectation of motions. For more information on MO4 see appendix A.1. If we generate the forecasts with the MO4 software we can interpolate the result and plot the forecast for this time period of one heading next to the actual measurements. The result for supplier A is shown in Figure 6, where 0 degrees heading is used and we look at the heave velocity standard deviation again (σ_{v_z}). We can already see that this model performs pretty well in this data subset, it looks like a fairly good estimation of the motions.

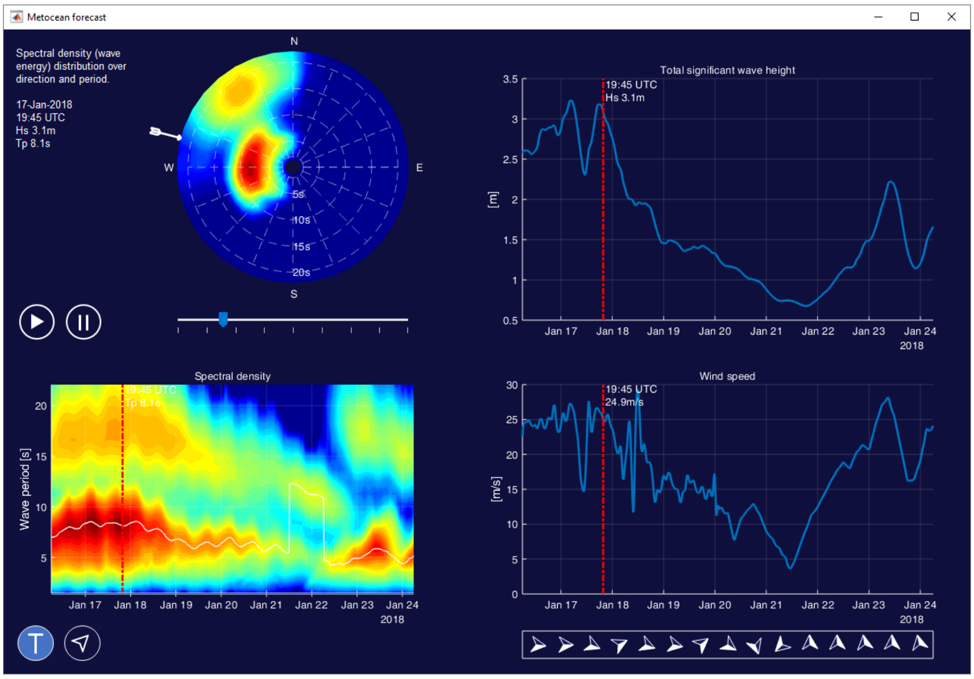

Figure 5: Spectral density (left) versus H_s and T_z representation (right)

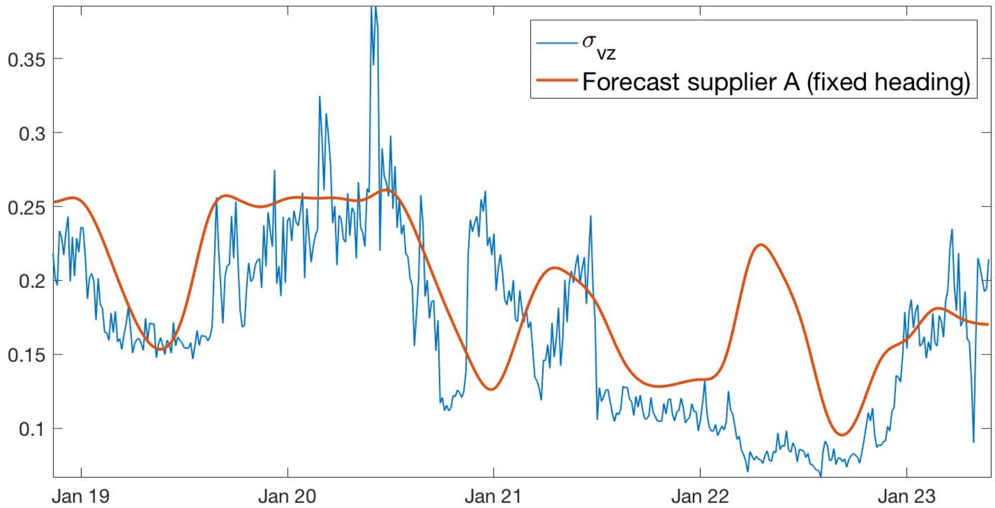

Figure 6: Forecast of supplier A with heading of 0 degrees of operating window 2

On the one hand we have the measured response data, of which the corresponding heading is known afterwards. In the forecasting situation on-board, it is different. The forecast is generated according to a heading input by the captain. For this reason we can not replicate the real situation in the forecasting exercise. The closest we can get is to adjust the physical model forecasts for the measured heading in hindsight. The result for the heave velocity standard deviation (σ_{v_z}) can be seen in Figure 26. The data resulting from this is the forecast if the captain would have inserted the actual measured heading.

We refer to these physical model forecasts with input of supplier A, B and C as $f_{t,A}$, $f_{t,B}$ and $f_{t,C}$ respectively. We should note here that the suppliers have different updating frequencies, which is the time between a new data file is received. In this research we use the most up to date data which is not entirely correct: this results in a more positive scenario when the forecasting horizon is longer than the updating interval. Due to this setup the forecasting performance does not depend on the forecasting horizon for the physical model forecasts. The physical forecasts made by using the data of supplier B and C are also plotted and included in appendix A.2.

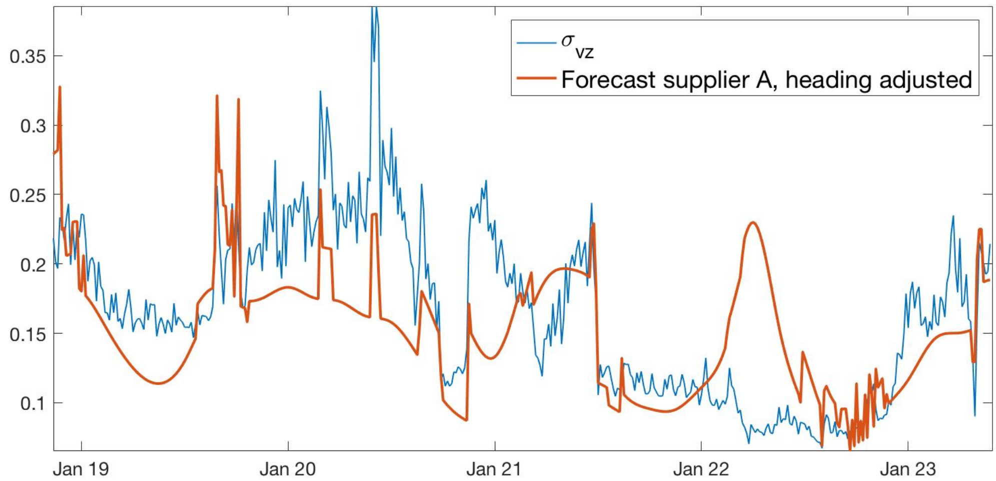

Figure 7: Forecast of supplier A adjusted for the measured headings of operating window 2

2.4 Forecasting Sample

The models are designed to be used in an out of sample forecasting exercises. This will be done by considering the data of operating window 2 which has 435 observations and which is plotted in Figure 6. We use the first 135 observations of the data to establish the model, leaving 300 observations to test the forecasting performance on. The forecast is then made at each time period while moving forward until the end of the data set is reached. The data of operating period 1, which is used for prior investigation has a length of 319 observations and can be found in Figure 30 in appendix A.4. This is data of a period between Jan-11 and Jan-15, which is close to the dates of the data we will be analyzing and therefore it should be relevant data to construct the prior on.

3 Models

3.1 Models in Research

This research compares the forecasting capabilities and suitability of multiple models to combine the physical model forecasts with time series data of ship motions. The following models are assessed for this purpose:

1. **Simple component model:**
   This model includes the expected significant wave height and one lagged variable in a linear model to estimate the vessel motions.

2. **Local level model with physical forecast:**
   Here the time series is modeled by an additive model with a slowly varying component that is then modeled by a random walk plus noise series in a state space equation, as given by Durbin and Koopman (2012). We consider two cases.

   (a) With re-estimation of weighting coefficients

   (b) With fixed and pre-estimated weighting coefficients

3. **Dynamic mixture model with physical forecast:** In this model the state space equation is used, comparable to the local level model above, but now a discrete latent variable is introduced which allows for shifts in the estimated state. An example is taken to the mixture of normals given by Carter and Kohn (1994), and the further use of it by Gerlach et al. (2000). Again we consider two cases:

   (a) With re-estimation of weighting coefficients

   (b) With fixed and pre-estimated weighting coefficients

We consider the last two models both in a case where the weighting coefficients are re-estimated during the procedure and in a case that the coefficients are estimated upfront and kept fixed throughout the forecasting exercise. The coefficients determine how sensitive the model is to the physical model forecasts. In the first case it means that also during the forecasting exercise at each new time period, the coefficients will be re-estimated. Allowing the coefficients to vary like this can lead to improved forecasting performance, as this flexibility allows the model to decrease the weight assigned to the physical forecast when it is giving inaccurate forecasts or to increase the weight when it is more accurate, see Figure 4 in section 2. By allowing this flexibility, the parameters are also estimated on varying data and on less data at some points. This can lead to the parameters to change a lot which may give unexpected estimating results, and possibly decreased performance. Therefore we will also construct these models without this flexibility.

In the second and third model, we have to estimate multiple parameters based on few observations because we want to be able to use the model to forecast straight away when starting the operation, and not only after acquiring a lot of data. Classical statistical approaches, or frequentist approaches, such as maximum likelihood estimation typically struggle to find optima in situations with few observations bu with multiple parameters to estimate. Frequentists restrict the assignment of probabilities to statements that describe the outcome of an experiment that can be repeated (Greenberg (2012)). This method of assigning probabilities suffers from the problem that its definition requires repeating the experiment an infinite number of times which is generally impossible. Another problem of this method is to assign a probability to a statement which cannot be considered as an outcome of a repeated experiment. Better suitable for the goal of this research is to follow the view of subjective probabilities and to implement Bayesian inferencing techniques. The methods used in sections 3.3 and 3.4 follow this point of view. An example is included in appendix A.5 to illustrate the difference in characteristics between the Frequentist point of view and that of the subjective probability, or Bayesian point of view.

3.2 Simple Component Model

The first model we consider will be built from components that are readily available and it will serve as a simple model in order to benchmark the performance of the other more complex models. Considering the data, we have measurements of the ship motions, data of the sea state, as well as the expected sea state in the near future (H_s). This data is acquired without analyzing it through a physical model, but it is a part of each meteorological data input. Combining these data sources into a model with linear components is an idea that seems feasible and accessible in the sense of modeling and computation. For we will not use the physical model, nor will we employ filtering and smoothing techniques.

In section 2 we saw that the data is non-stationary, and seems to behave almost like a random walk series. For this reason we choose to include only one lagged variable of the data in the simple component model. If we look at the data more closely, we can argue that the level of the observations can be determined by factoring the significant wave height (H_s), this is supported by hydrodynamic theory, which explains a linear relation between wave height and ship motion Massie and Journée (2001). An example to show this relation is included in appendix A.6.

Now we set up the model combining the first order lagged variable with the expected H_s in a linear model, given as:

$$\hat{y}_t = c_1 y_{t-1} + c_2 H_{s,t}, \qquad t = 1, ..., n, \quad s = \text{significant height.} \tag{1}$$

Here c_1 and c_2 are the coefficients which will determine the weight of the components of the model during forecasting. If c_1 is higher, more weight is given to the lagged variable, if c_2 is higher, more weight is given to the expected sea state. The coefficients will be estimated through linear regression on the same initialization set as described above. This gives us a simple model that could serve as a base case benchmark to the complex models. The only downside is, H_s is still estimated by the meteorological companies, therefore it is not free from their accuracy performance. Because we have three suppliers of metocean data, this will result in a sub-models for each H_s supplier A, B and C. It should be noted that H_s is not the same data as will be used in the following models, where f_t is used. The difference is explained in section 2.3.

3.3 Local Level Model with Physical Forecast

Apart from the simple component model above, all models in this research are based on state space representations. We will first discuss the use of state space models in time series and discuss why it is a useful technique in this research. In this section we build on a special case of state space models which is the local level model.

3.3.1 State Space Models in Time Series Analysis

The purpose of state space analysis, as stated by Durbin and Koopman (2012), is to infer properties of a system of which the series $(\boldsymbol{x}_1, ..., \boldsymbol{x}_n)$ is unobserved but are associated to a series of observations $(y_1, ..., y_n)$ from which knowledge is gained. The relation between $\boldsymbol{x}_t$

and y_t is then specified by the *state space model.* A great deal of the history of this type of modeling originates from engineering standpoints, where Jazwinski (1970) remarked that for an engineer to control the performance of a system, he should know the *state* of the system. To determine the state, the engineer will mostly have noisy measurements (due to various reasons, such as contamination in measurement devices or unavailability of direct data) from which he should estimate the true system state. An example in engineering is the problem of navigation, where the state consists of position and velocity of the craft given measurements on various variables. This clearly indicates the link to the problem in this research, where we are interested in the state of a vessel, being some variable of its motion, given some measurements.

There are many more practical examples available of state space modeling in engineering as well as other fields of study. These prove that it is a tested technique outside economics. We will extend these researches by focusing on multiple period ahead forecasting, which is not a general application in engineering and has more relation to economic modeling. Also we will include different sources of future data, which are forecasts based on physical models, and combine it with the time series data to get a new, possibly superior, forecast.

A general representation of the state space model, consisting of an observation equation and a state equation, is given by Hamilton (1994) as:

$$\begin{aligned} \boldsymbol{\xi}_{t+1} &= \boldsymbol{F}\boldsymbol{x}_t + \boldsymbol{v}_{t+1}, \\ \boldsymbol{y}_t &= \boldsymbol{A}'\boldsymbol{x}_t + \boldsymbol{H}_t'\boldsymbol{\xi}_t + \boldsymbol{w}_t. \end{aligned} \tag{2}$$

Here $\boldsymbol{\xi}_t$ is the $(r \times 1)$ unobserved state vector. $\boldsymbol{F}$ is an $(r \times r)$ matrix and the vector $\boldsymbol{v}_t$ is i.i.d. $N(\mathbf{0}, \boldsymbol{Q})$. $\boldsymbol{y}_t$ is the $(n \times 1)$ observation vector at time t where $\boldsymbol{H}'$ is an $(n \times r)$ matrix of coefficients, and $\boldsymbol{w}_t$ can be seen as the measurement error and is assumed to be i.i.d. $N(\mathbf{0}, \boldsymbol{R})$. Next to that, Hamilton (1994) includes a $(k \times 1)$ vector of observed variables which are exogenous and are included through the $(n \times r)$ coefficient matrix $\boldsymbol{A}'$.

The Local Level Model Next we can consider the specific case of a state space model which is the local level model

$$\begin{aligned} y_t =& \mu_t + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma_\varepsilon^2), \\ \mu_t =& \mu_{t-1} + \eta_t, \qquad \eta_t \sim N(0, \sigma_\eta^2). \end{aligned} \tag{3}$$

In this model the state μ_t and disturbance ε_t are unobserved and the disturbances ε_t and η_t are independent for all t. The parameters to be estimated in this model are σ_ε^2 and σ_η^2. Additionally the initialization μ_1 needs to be specified to achieve a complete model, this can be done by using $\mu_1 \sim N(\hat{\mu}_{1|0}, \hat{P}_{1|0})$, where $\hat{\mu}_{1|0}$ is the unconditional mean, which is 0, following the specification of (2). $\hat{P}_{1|0}$ is the unconditional variance which can be calculated following Hamilton (1994). Instead, we can also set it to a large number, mimicking a diffuse prior, which is chosen as $\hat{P}_{1|0} = 10^8$. The local level model is actually a random walk plus noise specification: the noise is inserted through the white noise process (NID) and the random walk is specified by the first order autoregressive specification of the level μ_t. When the variance of the observation disturbance is set to 0 we are left with a pure random walk process ($\sigma_\varepsilon^2 = 0$), and when the level disturbance is set to zero ($\sigma_\eta^2 = 0$) we find a white noise process with constant level (Durbin and Koopman (2012)).

The Local Level Model with Physical Forecast The goal of this research is to use forecasts generated by a physical model in combination with time series data to improve the forecasting capability. The physical forecast data, which is already available for at least 5 days ahead at the moment of forecasting, gives us an indication of the expected state but certainly does not give an exact expectation of the actual measured responses, as seen in section 2.3. The state space representation enables us to include these forecasts elegantly because we do not have to take it as a direct component of the observation, but rather as a latent influence on the unobserved state. We adjust the local level model to include the physical forecast data, as explained in section 2.3, through the state equation by including f_t as follows in the state space equation:

$$\begin{aligned} y_t =& \mu_t + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \\ \mu_t =& \begin{bmatrix} \phi_1 & \phi_2 \end{bmatrix} \begin{bmatrix} \mu_{t-1} \\ f_t \end{bmatrix} + \eta_t, \qquad \eta_t \sim N(0, \sigma_\eta^2), \end{aligned} \tag{4}$$

where f_t indicates either one of the physical forecasts. Here $\boldsymbol{F}_t = (\phi_1,\ \phi_2)$ needs to be estimated next to the parameters of the local level model σ_ε^2 and σ_η^2. We collect the unknown parameters in $\boldsymbol{\theta}$ for notation purposes. This brings us to the difference between part (a) and part (b) of this model, that is, with and without re-estimation of the coefficients in $\boldsymbol{F}_t$. In the first one, we proceed in the forecasting exercise by re-estimating the model at each new time period, which includes estimating the weighting coefficients in $\boldsymbol{F}_t$. The re-estimation of these parameters lead to the flexibility we want in order to account for the recent accuracy of the physical forecast. The estimation procedure of both sub models is discussed in section 4.1.

3.4 Dynamic Mixture Model with Physical Forecast

In section 2 we saw that apart from the fluctuations in the observations around some mean level, the time series often sharply increases or decreases, which seem like structural breaks. These incidents can be seen as level shifts of the unobservable state of the system. We want to include this property, as these shifts may be very useful in forecasting the correct vessel motions for the next observations, as there might have been a change in the characteristics of the response due to weather, loading, or heading.

3.4.1 Bayesian Approach to Structural Change

To cope with these shifts we can model the state equation such that it is capable of switching to another level when necessary. For this purpose a discrete latent variable $\boldsymbol{K} = (K_1, ..., K_n)$ can be added to allow for structural change in the model. By letting K_t take on a value of either 0 or a positive constant, it can be multiplied with the error term in the model to allow level shifts, innovation outliers or breaks in parameters. If $K_t = 0$ the level stays in place, but when it takes on a positive value we find a large variance which can lead to a level shift. Carter and Kohn (1994) as well as Shephard (1994) drew K conditional on the states and observations. However, in data with structural breaks, the states will be highly correlated with the break points. The sampling method by McCulloch and Tsay (1993) improved the efficiency by not conditioning on the states but on the error. Next, a method to generate $\boldsymbol{K}$ in $O(n)$ operations was contributed by Gerlach et al. (2000), leading to great improvement in efficiency. This method functions on evaluating probabilities after integrating out the states. As K_t takes a finite number of values, the elements of the probabilities can be computed for each possible value, and K_t can then be found by normalizing the probabilities. We will follow this approach to achieve a dynamic mixture model including physical forecast data. They present the state space model as:

$$
\begin{aligned}
y_t &= g_t + \boldsymbol{h}_t'\boldsymbol{x}_t + \boldsymbol{G}_t\boldsymbol{u}_t, \\
\boldsymbol{x}_t &= \boldsymbol{c}_t + \boldsymbol{F}_t\boldsymbol{x}_{t-1} + \boldsymbol{\Gamma}_t\boldsymbol{u}_t.
\end{aligned} \tag{5}
$$

We write the state space equation again in the form of a local level model and add the physical forecast data f_t and the level shift procedure through K_t as follows

$$
\begin{aligned}
y_t =& \mu_t + \sigma_u e_t, \qquad e_t \sim N(0,1), \\
\mu_t =& \begin{bmatrix} \phi_1 & \phi_2 \end{bmatrix} \begin{bmatrix} \mu_{t-1} \\ f_t \end{bmatrix} + \sigma_v K_t v_t, \qquad v_t \sim N(0,1).
\end{aligned} \tag{6}
$$

Looking at the representation given in (5), we can identify the parameters as $\boldsymbol{x}_t = \mu_t$, $u_t = (e_t, v_t)'$, $\boldsymbol{c}_t = 0$, $g_t = 0$, $h_t = 1$, $\boldsymbol{G}_t = \sigma_u$, $\boldsymbol{F}_t = (\phi_1,\ \phi_2)$ and $\boldsymbol{\Gamma}_t = \sigma_v K_t$. This gives us the unknown model parameters μ_1, $\boldsymbol{F}$, σ_u and σ_v. Next to that, we will need to estimate the level shift (the vector $\boldsymbol{K}$), as this is generally unknown a priori. Further, given $\boldsymbol{K}$, the dynamic model will be assumed to be a linear and Gaussian state-space model (Gerlach et al. (2000)). Throughout this section we define for any variable z_t, that when $s \leq t$ we write $\boldsymbol{z}^{s,t} = (\boldsymbol{z}_s, ..., \boldsymbol{z}_t)$, following from this we have $\boldsymbol{z} = \boldsymbol{z}^{1,n}$.

Prior distribution for K An advantage of this model representation is that one can achieve a model that can capture the idea of a process that is unchanged in some periods whereas in others small but frequent breaks as well as large but infrequent breaks can occur (Giordani and Kohn (2008)). Inspecting the data of this research gives reason to believe we will encounter breaks of various magnitude as well, but we will not define a different prior probability to the magnitudes. Rather, if p is the amount of possibilities in K_t, we will simply split the prior probability equally, leading to the prior specification for $\boldsymbol{K}$ in Table 2.

Table 2: Prior distribution for K_t

K_t	0	1	4
$pr(K_t)$	$1-\pi$	$\pi/2$	$\pi/2$

Looking at the data of operating period 1 (the data of a previous operating period, not the one we are analyzing, see section 2.4 and Figure 30 in appendix A.4) we see that there might be between 10 and 20 level shifts (let us say 15) in 320 observations, which translates into the prior estimate $\pi = \frac{15}{n} \approx 0.05$ for $\boldsymbol{K}$. Here we have followed the assumption that a shift is justified when the observations stay around a level for more than 3 periods. Next to that, we assume that the data of operating period 1 will be informative for the data of operating period 2 which we will perform the forecasting procedure on and therefore consider unknown.

We have added the physical forecast data of time t as a component of the state μ_t into the model, as this information is already available at time t. The weight of this data point is determined by the factor $\frac{\phi_2}{\phi_1+\phi_2}$. The sum of ϕ_1 and ϕ_2 should be approximately 1, because this part of the model is additive and the level is decomposed into a part explained by the previous observation and a part by the physical model forecast of time t. The estimates of ϕ_i and thus the weight of this decomposition could determine the relevance of the physical model at a time period. If the weight given to the physical model is low, it does not have so much importance in forecasting either and we will trust more on the signal of the time series. The way the parameters are estimated is discussed in section 4.2.

4 Model Estimation

The estimation of the simple component model is done by linear regression, and the details are treated together with the results, as no substantial methods are used. Next we discuss the estimation of the local level models, where first case (a) is treated where the coefficients in $\boldsymbol{F}$ are re-estimated in each forecasting step. Case (b) is a simplification where these coefficients are only estimated before the forecasting exercise starts. The same holds for the dynamic mixture models.

4.1 Estimation of the Local Level Models

Carter and Kohn (1994) show us how to estimate parameters in a linear state space model with errors that are a mixture of normals. They use a Gibbs Sampler to carry out Bayesian inference. An important innovation of this paper compared to the general approach of Carlin, Polson, and Stoffer (1992) to Bayesian inference in state space models is that here the states are all generated at once by using the time ordering of the state space model instead of generating the states one by one by means of the Markov properties. To obtain the states they use the Kalman filter, which might not be the fastest filtering algorithm but an alternative suggested by Tierney (1994) using an Metropolis step within the Gibbs sampler to speed up state generation did not lead to practical success. We will use their procedure based on the Kalman steps to obtain state estimates in this model.

4.1.1 Local Level Model Case (a)

The first sub model of the local level models asks for estimation of the parameters in $\boldsymbol{F}_t$, as well as σ_ε^2, σ_η^2 and the starting point μ_1 where, in contrast to case (b), $\boldsymbol{F}$ will be re-estimated in each time step. As previously stated, we will make use of Bayesian inferencing techniques to estimate the parameters. A simulation technique that makes Bayesian analysis of models as these possible are Markov Chain Monte Carlo (MCMC) methods.

MCMC is a simulation method based on Markov theory. It is an iterative procedure where one samples from a Markov Chain of which the limiting distribution equals the posterior distribution. The basis of such an algorithm is the construction of a transition density having an invariant density equal to that of the target distribution, that is, one should find a density which can yield draws that are from the same density as the variables of interest. A general principle of finding such density (the kernel) is the Metropolis-Hastings algorithm. A special case of this is the Gibbs algorithm or Gibbs sampler. It can be used when one wants to sample from some nonstandard joint distribution but where it is possible to sample from every conditional distribution. The general algorithm can be found in appendix A.7. The variation of the Gibbs algorithm used to estimate this model is given next.

Local Level Gibbs Sampling Procedure

1. Start with $m = 0$ and set prior distributions and define starting values in $\boldsymbol{\theta}^{(0)}$
2. Set $m = m + 1$ and run the Kalman filter using $\boldsymbol{\theta}^{(m-1)}$
3. Run the State Smoother to obtain the state parameter estimates (mean and variance).
4. Sample new values $\boldsymbol{\theta}^{(m)}$ from the conditional posterior distributions .
5. Return to 3. until the desirable amount of iterations ($m = M$).

Kalman Filtering The goal of running the Kalman filter and smoother is to get an estimate of the state vector conditional on the data (y) and the unknown parameters (θ). That is, we want to update our knowledge of the system every time new information comes in, which fits into the subjective probability point of view. If we look at the state space representation in (2) as a joint normal distribution of the observations and the state, then we can use the observations to update and change our belief of this distribution as they come in. Then, having formed this distribution, we 'draw' from these conditional distributions to get new estimates of the state. After a certain amount of observations we might find that these parameters converge. This idea mainly works through applying the result of the Normal Lemma, given in appendix A.8, sequentially for time period in the data set.

As before, we define the state as a Gaussian process such that $\mu \sim N(\hat{\mu}, V_t)$. Then the Kalman Filter consists of the following 2 steps for each time period $t = 2, ..., T$ to find the conditional distribution:

- Predicting step

$$\hat{\mu}_{t+1|t} = \boldsymbol{F}\hat{\mu}_{t|t} \tag{7}$$

$$V_{t+1|t} = \boldsymbol{F}V_{t|t}\boldsymbol{F}' + \sigma_\eta^2 \tag{8}$$

- Updating step

$$\hat{\mu}_{t+1|t+1} = \hat{\mu}_{t+1|t} + V_{t+1|t}(V_{t+1|t} + \sigma_\varepsilon^2)^{-1}(y_{t+1} - \hat{\mu}_{t+1|t}) \tag{9}$$

$$V_{t+1|t+1} = V_{t+1|t} - V_{t+1|t}(V_{t+1|t} + \sigma_\varepsilon^2)^{-1}V_{t+1|t} \tag{10}$$

Here $\boldsymbol{F}$, σ_u and σ_v are as defined in (4). What we have found is thus the distribution of our state, conditional on the observed data and on the unknown parameters, which are fixed throughout this process. What we can see in the updating step, that is when an observation comes in, is that the updated variance decreases from the predicted value: an observation gives us more information and consequently the uncertainty shrinks.

State Smoothing Next, we use the state smoothing algorithm, which iterates backwards through the observations. This way it uses the complete sample information to obtain an improved estimate of the state vector ($\hat{\mu}_{t|T}$) for $t = 1, ..., T$. Basically, the law of iterated expectation enables us to first condition on μ_{t+1}, which is actually unobserved, and then we average out all possible μ_{t+1} using the information in the full sample. The most important part of the smoothing equations are the smoothed state estimate $\hat{\mu}_{t|T}$ and its variance ($V_{t|T}$), which are

$$\hat{\mu}_{t|T} = \mu_{t|t} + V_{t|t}F'V_{t|t+1}^{-1}(\hat{\mu}_{t+1|T} - \hat{\mu}_{t+1|t}), \tag{11}$$

$$V_{t|T} = V_{t|t} - V_{t|t}F'V_{t|t+1}^{-1}(V_{t|t+1} - V_{t|T})V_{t|t+1}^{-1}FV_{t|t}. \tag{12}$$

Sampling from the Conditional Posteriors The state estimates resulting from the Kalman filter and smoother are based on the previous values in $\boldsymbol{\theta}$, but they result in updated estimates of the state and the variance. With these updates we can draw new parameters for $\boldsymbol{\theta}$. Returning to our model, having parameters $\boldsymbol{\theta} = (\mu_1, \boldsymbol{F}, \sigma_\varepsilon, \sigma_\eta)$, we still have two Gaussian processes. This means we can follow Greenberg (2012) to identify the distributions that can be used in the algorithm. First, we can identify the distribution for our observations ($y = (y_1, ..., y_n)$) as:

$$y \sim N_n(\mu, \sigma_\varepsilon^2 \boldsymbol{I}_n),$$

followed by the prior distribution for the variance in the observation equation

$$\sigma_\varepsilon^2 \sim IG(\alpha_0/2, \delta_0/2). \tag{13}$$

Then we can find the conditional posterior distribution

$$\sigma_\varepsilon^2 | y \sim IG(\alpha_1/2, \delta_1/2), \tag{14}$$

where the parameters for the gamma distribution follow from

$$\alpha_1 = \alpha_0 + n, \qquad \delta_1 = \delta_0 + (y - \mu)'(y - \mu). \tag{15}$$

Here α_0 and δ_0 are the priors to the distribution. There are multiple ways to arrive at this distribution. One standard approach is given by Petris, Petrone, and Campagnoli (2009) and is based on the fact that the full conditional distribution is proportional to the joint distribution of all the random variables considered. This calculation can be found in A.10.

Continuing with the state equation, where we can define the state estimates ($\mu = (\mu_1, ...\mu_n)$) as:

$$\mu \sim N_n(\boldsymbol{X}\boldsymbol{F}, \sigma_\eta^2 \boldsymbol{I}_n),$$

where we have the prior distributions

$$\boldsymbol{F} \sim N(\boldsymbol{F}_0, \boldsymbol{B}_0), \qquad \sigma_\eta^2 \sim IG(a_0/2, d_0/2). \tag{16}$$

These assumptions result in finding that

$$\boldsymbol{F}|\sigma_\eta, y \sim N(\overline{\boldsymbol{F}}, \boldsymbol{B}_1), \tag{17}$$

where

$$\boldsymbol{B}_1 = [\sigma_\eta^{-2}\boldsymbol{X}'\boldsymbol{X} + \boldsymbol{B}_0^{-1}]^{-1}, \quad \overline{\boldsymbol{F}} = \boldsymbol{B}_1[\sigma_\eta^{-2}\boldsymbol{X}'\mu + \boldsymbol{B}_0^{-1}\boldsymbol{F}_0], \tag{18}$$

and that

$$\sigma_\eta^2|\boldsymbol{F}, y \sim IG(a_1/2, d_1/2), \tag{19}$$

where similar to before we find the distribution parameters

$$a_1 = a_0 + n, \qquad d_1 = d_0 + (\mu - \boldsymbol{X}\boldsymbol{F})'(\mu - \boldsymbol{X}\boldsymbol{F}). \tag{20}$$

Here $\boldsymbol{F}_0$, $\boldsymbol{B}_0$, a_0 and d_0 are the prior parameters to the distributions and $\boldsymbol{X} = \begin{bmatrix} \mu_{t-1} & f_t \end{bmatrix}'$. In this case the above equations mean we can sample the variances from inverse gamma distributions and the parameters in $\boldsymbol{F}$ from a joint normal distribution.

The Gibbs sampler can be applied when all the conditional posterior distributions are standard (i.e. they are known distributions which we can simulate), enabling us to find the posterior distribution of the unknown parameters in $\boldsymbol{\theta}$. This concludes step 4 of the Gibbs defined procedure. The only missing component are the starting values and prior parameters.

After the Markov chain has converged (at iteration $m = m^*$), the simulated values of

$\{\theta_i^{(m)}, m \leq m^*\}$ can be used as a sample from the joint posterior distribution. Next to that, the parameter estimates for each parameter in $\boldsymbol{\theta}$ can be calculated by (21), after defining a burn in sample of B and a desired sample size $B+G$. We will use $B = 100$ and $B+G = 1000$.

$$\hat{\theta}_i = \frac{1}{G-B} \sum_{g=B+1}^{B+G} \theta_i^{(g)} \tag{21}$$

Prior Parameters and Starting Values The prior parameters needed in the method above can be found by analyzing data of the previous operating window. This is the data of an operation between January 10 and January 15, which is not far from the data central to this research and therefore is considered relevant, see Figure 30 in appendix A.4. This follows the argument of Greenberg (2012) that the prior distributions should be selected on best possible previous research or investigation, as due to seasonal differences the data characteristics and sea states encountered there are most likely the most relevant for the next operating window. Also with this choice we do not re-use any data, as would be the case when constructing parameters while looking at the data of operating window 2.

For this purpose we use the non-Bayesian technique of maximum likelihood estimation based on the same method of Kalman filtering and smoothing to acquire the state estimates. The difference is that here the likelihood is evaluated and then by using an optimization algorithm the parameters are found. The problem with this method is that it needs many observations to find optimal parameters, but even with lots of data it may encounter problems because the likelihood can not be properly maximized over multiple parameters. Here it is applied on the full data set of the previous operating period (320 observations) and we use its result to determine the prior parameters. Details of the method can be found in appendix A.9. Using asymmetric starting values for the variances ($\sigma_\varepsilon > \sigma_\eta$ and signal to noise ratio $q = 1/3$), we find the ML parameter estimates in Table 3.

Table 3: ML estimates

ϕ_1	ϕ_2	σ_ε	σ_η	μ_1
0.917	0.088	0.007	0.003	0.073

The prior parameters necessary to get to prior distributions resembling the maximum likelihood results are found as given in Table 4. The distribution of the errors should for instance resemble the gamma distributions used to sample the variances from. The starting values can be drawn from the distributions resulting from these parameters.

Table 4: Prior parameters for the local level model

α_0	δ_0	a_0	d_0	$\boldsymbol{F}_0$	$\boldsymbol{B}_0$
2	1	2	0.1	$\begin{bmatrix} 0.9 & 0.1 \end{bmatrix}$	$\begin{bmatrix} 2 & 0 \\ 0 & 0.5 \end{bmatrix}$

The Gibbs Kernel Theory reads that we should find convergence in the parameters after repeating steps 1 to 5 for sufficiently large m. In practice, convergence is not guaranteed, and the conditions to find convergence to the target are not straightforward to verify. When analyzing the results we will investigate whether we have found convergence of this sampler.

4.1.2 Local Level Model Case (b)

The second sub model of the local level models is a simplification of the one with varying coefficients. In this case, we use the investigation on the previous operating window to establish the values in $\boldsymbol{F}$ and keep them fixed throughout the whole forecasting exercise in the operating window we have at hand. Again, this resembles the case in practice where we would base our model on the most recent available data, and use it directly when starting the new operation to provide improved forecasts based on the measurement data as it comes in. This means that the process of the varying coefficient model is used but now with $\phi_1 = 0.9$ and $\phi_2 = 0.1$ fixed throughout the procedure. The same prior parameters and starting values are used and the parameter estimates can be calculated by (21).

4.2 Estimation of the Dynamic Mixture Models

In the first case of the dynamic mixture models, we estimate all parameters at each time step during the forecasting exercise, where in the second case we use predetermined coefficients throughout the whole forecasting exercise as explained in 3. Concerning the estimation procedure, we start with the first, after which the second is easily explained as it is a simplification of the first case.

4.2.1 Dynamic Mixture Model Case (a)

In this sub model of the dynamic mixture model we estimate all the parameters in each time step when forecasting. In the dynamic mixture model with physical forecast data, as given in (6), we have to estimate the unknown parameters μ_1, $\boldsymbol{F}$, σ_u and σ_v, which we collect in $\boldsymbol{\theta}$. Next to that, we have to estimate the vector $\boldsymbol{K}$. To achieve this, we will again make use of a Gibbs sampling procedure, similar to the one in section 4.1.1. Now we combine it with one of the recursions given by Gerlach et al. (2000) for efficient Bayesian inference for Dynamic Mixture Models to generate $\boldsymbol{K}$ in an efficient way. This leads to the following sampling procedure which results in simultaneous estimation of the states, the parameters and the vector $\boldsymbol{K}$.

Dynamic Mixture Gibbs Sampling Procedure

1. Start with $m = 0$: Set prior distributions and define starting values for the unknown parameters $\boldsymbol{\theta}^{(0)}$ and for the vector $\boldsymbol{K}$.
2. Run *Recursion 1* to obtain a new vector $\boldsymbol{K}$
3. Update the estimates of the state using the new values for $\boldsymbol{K}$.
4. Sample new values $\boldsymbol{\theta}^{(m)}$ from the conditional distributions using the updated state estimates.
5. Accept or reject the draws based on the rejection scheme. If rejected return to 4.
6. Return to 2. and repeat a desirable amount of times.

Sample K and Update the State Estimates In step 2, we run *Recursion 1* given by Gerlach et al. (2000). It assumes the other parameters are given and does not estimate these. The objective of this recursion is to generate $\boldsymbol{K}$ which is done by calculating the following conditional likelihood:

$$\begin{aligned} p(\boldsymbol{K}_t|\boldsymbol{y}, \boldsymbol{K}_{s\neq t}) &\propto p(\boldsymbol{y}|\boldsymbol{K})p(\boldsymbol{K}_t|\boldsymbol{K}_{s\neq t}) \\ &\propto p(\boldsymbol{y}^{t+1,n}|\boldsymbol{y}^{1,t}, \boldsymbol{K})p(y_t|\boldsymbol{y}^{1,t-1}, \boldsymbol{K}^{1,t}) \times p(\boldsymbol{K}_t|\boldsymbol{K}_{s\neq t}). \end{aligned} \tag{22}$$

Given the data and given a vector of $\boldsymbol{K}_{t\neq s}$ (K_t for all other periods ($t \neq s$)), this gives for each value of $\boldsymbol{K}_t$ the likelihood at a time period (t). If we have the likelihood of each possible K_t at a specific time period we can generate K_t: as we can accept it to be the most likely possibility out of $\boldsymbol{K}_t$ for each t.

The term $p(\boldsymbol{K}_t|\boldsymbol{K}_{s\neq t})$ is obtained from the prior. To efficiently calculate the other two terms we follow the mentioned recursion through its lemmas. We give the main working of the Lemmas here such that the recursion can be followed. The full recursion and its Lemmas can be found in A.11, and the proofs can be found in Gerlach et al. (2000). The method is discussed referring to the general state space model as given in (2).

Summary of the Recursion Lemmas

Lemma 1. Let $r_{t+1} = \text{var}(y_{t+1}|\boldsymbol{x}_t, \boldsymbol{K}^{1,t+1})$. Then the following holds:

a. $E(y_{t+1}|\boldsymbol{x}_t, \boldsymbol{K}^{1,t+1})$

$$= g_{t+1} + \boldsymbol{h}'_{t+1}(\boldsymbol{f}_{t+1} + \boldsymbol{F}_{t+1}\boldsymbol{x}_t) \tag{23}$$

$$r_{t+1} = (\boldsymbol{h}'_{t+1}\boldsymbol{\Gamma}_{t+1} + \boldsymbol{G}_{t+1})(\boldsymbol{h}'_{t+1}\boldsymbol{\Gamma}_{t+1} + \boldsymbol{G}_{t+1})', \tag{24}$$

$$E(x_{t+1}|\boldsymbol{x}_t, y_{t+1}, \boldsymbol{K}) = \boldsymbol{a}_{t+1} + \boldsymbol{A}_{t+1}\boldsymbol{x}_t + \boldsymbol{B}_{t+1}y_{t+1},$$

and

$$\text{var}(x_{t+1}|\boldsymbol{x}_t, y_{t+1}, \boldsymbol{K}) = \boldsymbol{C}_{t+1}\boldsymbol{C}'_{t+1},$$

where $\boldsymbol{C}_{t+1}\boldsymbol{C}'_{t+1}$ can be factored to get a matrix of $\boldsymbol{C}_{t+1}$ that is either null or has full rank.

b. One can write

$$\boldsymbol{x}_{t+1} = \boldsymbol{a}_{t+1} + \boldsymbol{A}_{t+1}\boldsymbol{x}_t + \boldsymbol{B}_{t+1}y_{t+1} + \boldsymbol{C}_{t+1}\boldsymbol{z}_{t+1}$$

where $z_{t+1} \sim N(0, I)$ and is independent of $\boldsymbol{x}_t$ and y_{t+1} (conditional on $\boldsymbol{K}$).

Lemma 2. For $t = 1, ..., n-1$ the conditional density $p(\boldsymbol{y}^{t+1,n}|\boldsymbol{x}_t, \boldsymbol{K})$ is independent of $\boldsymbol{K}^{1,t}$ and can be expressed as

$$p(\boldsymbol{y}^{t+1,n}|\boldsymbol{x}_t, \boldsymbol{K}) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{x}'_t\boldsymbol{\Omega}_t\boldsymbol{x}_t - 2\boldsymbol{\mu}'_t\boldsymbol{x}_t)\right\} \tag{25}$$

The terms $\boldsymbol{\Omega}_t$ and $\boldsymbol{\mu}'_t$ are computed recursively as follows:

- Set $\boldsymbol{\Omega}_n = 0$ and $\boldsymbol{\mu}'_n = 0$.
- For $t = n-1, ..., 1$, calculate $\boldsymbol{\Omega}_t$ and $\boldsymbol{\mu}_t$ through recursive formulas, given in the full lemmas in the Appendix.

Let $\boldsymbol{m}_t = E(\boldsymbol{x}_t|\boldsymbol{y}^{1,t}, \boldsymbol{K})$, $\boldsymbol{V}_t = \mathrm{var}(\boldsymbol{x}_t|\boldsymbol{y}^{1,t}, \boldsymbol{K})$, and $R_t = \mathrm{var}(\boldsymbol{y}_t|\boldsymbol{y}^{1,t-1}, \boldsymbol{K})$. The next lemma is the Kalman filter for the state space model in (2).

Lemma 3.

a. $R_t = \boldsymbol{h}_t'\boldsymbol{F}_t\boldsymbol{V}_{t-1}\boldsymbol{h}_t'\boldsymbol{F}_t + (\boldsymbol{h}_t'\boldsymbol{\Gamma}_t + \boldsymbol{G}_t)(\boldsymbol{\Gamma}_t'\boldsymbol{h}_t + \boldsymbol{G}_t')$,

$$\boldsymbol{m}_t = (\boldsymbol{I} - \boldsymbol{J}_t\boldsymbol{h}_t')(\boldsymbol{f}_t + \boldsymbol{F}_t\boldsymbol{m}_{t-1}) + \boldsymbol{J}_t(y_t - g_t),$$

and

$$\boldsymbol{V}_t = \boldsymbol{F}_t\boldsymbol{V}_{t-1}\boldsymbol{F}_t' + \boldsymbol{\Gamma}_t\boldsymbol{\Gamma}_t' - \boldsymbol{J}_t\boldsymbol{J}_t'R_t, \tag{26}$$

where

$$\boldsymbol{J}_t = [\boldsymbol{F}_t\boldsymbol{V}_{t-1}\boldsymbol{F}_t' + \boldsymbol{\Gamma}_t(\boldsymbol{\Gamma}_t'\boldsymbol{h}_t + \boldsymbol{G}_t')]/R_t.$$

b. The conditional density

$$p(y_t|\boldsymbol{y}^{1,t-1}, \boldsymbol{K}^{1,t}) \propto R_t^{-1/2}\exp\left\{-\frac{1}{2R_t}(y_t - g_t - \boldsymbol{h}_t'(\boldsymbol{f}_t + \boldsymbol{F}_t\boldsymbol{m}_{t-1}))^2\right\}.$$

c. We can write $\boldsymbol{V}_t = \boldsymbol{T}_t\boldsymbol{T}_t'$, where the matrix $\boldsymbol{T}_t$ either has full column rank if $\boldsymbol{V}_t \neq 0$ or is null if $\boldsymbol{V}_t = 0$. Conditional on $\boldsymbol{K}$ we express $\boldsymbol{x}_t$ as $\boldsymbol{x}_t = \boldsymbol{m}_t + \boldsymbol{T}_t\boldsymbol{\xi}_t$, where $\boldsymbol{\xi}_t$ is $N(\boldsymbol{0}, \boldsymbol{I})$ and independent of $\boldsymbol{y}^{1,t}$.

The next lemma uses Lemma 3 to efficiently evaluate the factor $p(\boldsymbol{y}^{t+1,n}|\boldsymbol{y}^{1,t}, \boldsymbol{K})$ in (22).

Lemma 4.

$$\begin{aligned} p(\boldsymbol{y}^{t+1,n}|\boldsymbol{y}^{1,t}, \boldsymbol{K}) &= \int p(\boldsymbol{y}^{t+1,n}|\boldsymbol{x}_t, \boldsymbol{K}^{t+1,n})p(\boldsymbol{\xi}_t|\boldsymbol{K}^{1,t})d\boldsymbol{\xi}_t \\ &\propto |\boldsymbol{T}_t'\boldsymbol{\Omega}_t\boldsymbol{T}_t + \boldsymbol{I}|^{-1/2}\exp\left\{-\frac{1}{2}[\boldsymbol{m}_t'\boldsymbol{\Omega}_t\boldsymbol{m}_t - 2\boldsymbol{\mu}_t'\boldsymbol{m}_t\right. \\ &\left. -(\boldsymbol{\mu}_t - \boldsymbol{\Omega}_t\boldsymbol{m}_t)'\boldsymbol{T}_t(\boldsymbol{T}_t\boldsymbol{\Omega}_t\boldsymbol{T}_t + \boldsymbol{I})^{-1}\boldsymbol{T}_t'(\boldsymbol{\mu}_t - \boldsymbol{\Omega}_t\boldsymbol{m}_t)]\right\} \end{aligned} \tag{27}$$

The recursion for generating $\boldsymbol{K}$ in $O(n)$ operations is then given as:

Recursion 1: Sampling $\boldsymbol{K}$ *From Reduced Conditionals*

Step 1. Given the current value for $\boldsymbol{K}$, calculate $\boldsymbol{\Omega}_t$ and $\boldsymbol{\mu}_t$ for $t = n-1, ..., 1$, using the recursions in Lemma 2.

Step 2. Given $E(\boldsymbol{x}_0)$, $\mathrm{var}(\boldsymbol{x}_0)$, perform the following for $t = 1, ..., n$:

a. For each value of $\boldsymbol{K}_t$:

- Obtain $\boldsymbol{R}_t$, $\boldsymbol{m}_t$ and $\boldsymbol{V}_{t-1}$ as in Lemma 3.
- Obtain $p(y_t|\boldsymbol{y}^{1,t-1}, \boldsymbol{K}^{1,t})$ as in Lemma 3, part b and obtain $p(\boldsymbol{y}^{t+1,n}|\boldsymbol{y}^{1,t}, \boldsymbol{K})$ as in Lemma 4.

b. Obtain $p(\boldsymbol{K}_t|\boldsymbol{y}, \boldsymbol{K}_{s\neq t})$ for all values of $\boldsymbol{K}_t$ by normalization. Then generate $\boldsymbol{K}_t$.

c. Update $\boldsymbol{m}_t$ and $\boldsymbol{V}_t$ as in Lemma 3, using generated values of $\boldsymbol{K}_t$.

Sampling from the Conditional Posteriors The recursion above is able to run due to the fact that we took the unknown parameters in $\boldsymbol{\theta}$ as given. However, these need to be estimated. In recursion step 2.c we update $\boldsymbol{m}_t$ and $\boldsymbol{V}_t$ using the newly generated $\boldsymbol{K}_t$. This gives us new estimates of the state parameters as these updates are the estimated level and an estimate of variance (similar to the Kalman filter steps). This corresponds to the Gibbs Sampling procedure of section 4.1.1, where we achieved this by the general Kalman filtering algorithm. Again we can sample from conditional distributions to get new values for the unknown parameters. The posterior distributions are similar to the ones defined in section 4.1.1, as in this step the vector $\boldsymbol{K}$ is considered given. To sample from the posterior distributions one can therefore use (13) to (20) but now with $\sigma_\varepsilon = \sigma_u$ and $\sigma_\eta = \sigma_v$. The parameter estimates can be found by using (21). Next to that, we should define new prior parameters for these distributions.

Rejection Scheme and Prior Parameters Sampling from the distributions above lead to draws of ϕ_1 and ϕ_2 from a joint normal distribution. Therefore they can have a fairly large spread depending on the variance. However, for this model it is necessary to have $\phi_1 + \phi_2$ very close to 1, as otherwise the method fails to find relevant level shifts (level estimates goes towards 0). To achieve this we have two options: one is to set the priors very narrow in order to find exactly the values that give a sum of values close to one. The other is to employ a rejection scheme, in order to reject draws that do not live up to the expectation. Setting very narrow priors is not in line with the flexibility we want to achieve, as we want to allow for a certain weighting of the components by these coefficients which is unknown a priori. We employ a rejection scheme by setting a wide prior variance for the coefficients, and by putting a strong prior on $\phi_1+\phi_2$ such that draws that do not fall within $0.999 \leq \phi_1 + \phi_2 \leq 1.001$ are not taken into account. This results in re-sampling 50 times as often as the original sampling amount, which clearly decreases the efficiency of the model but it leads to the desirable result of parameters that change throughout the forecasting procedure.

The maximum likelihood estimates in Table 3 are used again to give an indication of suitable prior parameters. The prior parameters which will be used for the dynamic mixture model are given in Table 5 and are again based on the maximum likelihood estimates in Table 3 in section 3.3.

Table 5: Prior parameters for the dynamic mixture model

α_0	δ_0	a_0	d_0	$\boldsymbol{F}_0$	$\boldsymbol{B}_0$
2000	0.5	2	0.1	$\begin{bmatrix} 0.9 & 0.1 \end{bmatrix}$	$\begin{bmatrix} 2 & 0 \\ 0 & 0.5 \end{bmatrix}$

The starting values are drawn from the distributions corresponding to the parameters in Table 5. Further, μ_1 is drawn from a normal distribution with mean and variance of the observations of the previous operating window.

4.2.2 Dynamic Mixture Model Case (b)

The dynamic mixture model will also be tested by using the weighting coefficients in the state equation of $\phi_1 = 0.9$ and $\phi_2 = 0.1$ throughout the whole forecasting procedure. This can lead to a benefit as we do not need the rejection scheme of the previous section. The rest of the model is kept equal and so is the estimation procedure.

5 Application

In this research the application of the models lies in forecasting after implementing the model in the existing software. Therefore we will first discuss how the forecasts are made and how the performance is measured. Next we define criteria to determine how suitable each model is for implementation.

5.1 Forecasting

An innovation compared to other state space models is made by including future forecast data into the model, and using this in combination with the time series model to construct a new forecast. Forecasting will be done based on the estimated level. Using (6) for the dynamic mixtures or (4) for the local level models, the 1 step ahead forecast ($h = 1$) can be written as

$$\hat{\mu}_{T+1|T} = E[\mu_{T+1}|I_T] = \begin{bmatrix} \phi_1 & \phi_2 \end{bmatrix} \begin{bmatrix} \hat{\mu}_{T|T} \\ f_{T+1} \end{bmatrix}. \tag{28}$$

Then onwards for $h = 2$ we get:

$$\hat{\mu}_{T+2|T} = \begin{bmatrix} \phi_1 & \phi_2 \end{bmatrix} \begin{bmatrix} \hat{\mu}_{T+1} \\ f_{T+2} \end{bmatrix}, \tag{29}$$

and for $h > 2$ it holds

$$\hat{\mu}_{T+h|T} = \begin{bmatrix} \phi_1 & \phi_2 \end{bmatrix} \begin{bmatrix} \hat{\mu}_{T+h-1} \\ f_{T+h} \end{bmatrix}. \tag{30}$$

Here $\hat{\mu}_{T|T}$ is the smoothed estimate of the signal at $t = T$. The forecast values $\hat{\mu}_{T+h|T}$ are normal distributed variables, as $\hat{\mu}_{T|T}$ is a normal distributed variable and we can view adding data from f_t as adding a constant to it, because this data is fixed. We expect that as we continue to forecast multiple steps ahead from $t = T$ to $t = T + h$, the predicted values will lie closer and closer to the physical forecast data because in each step the signal is influenced by it. Next to that, this effect propagates forward as the next prediction ($T + h$) is based on the previous state estimate ($T + h - 1$) and the new physical data point (f_{T+h}). This is a desirable effect, as the level of the time series itself is less and less indicative of the expected sea state further in the future, while this is exactly where the strength of the physical forecast lies. The weight assigned by ϕ_2 to the physical forecast will determine how quickly the forecast is drawn towards it and away from the time series data.

5.1.1 Forecasting Performance

This research has the objective to improve the current forecasts for ship motions. We focus on multiple period ahead forecasts, because the MO4 forecasting tool is mainly used to determine whether an operation can continue multiple hours or even days upfront. In this research we investigate the forecasting performance of 4, 8 and 12 hours ahead forecasts. With the data sampling frequency this implies 16, 32 and 48 period ahead forecasts because the standard deviation in 15 minute intervals is used (forecasting horizons $h =$

$(16, 32, 48)$). In normal time series analysis, forecasting this many periods ahead would not be likely to give relevant results with this type of data. However, in this research we have a special case where we have the future forecast data of the physical model at hand. The physical forecasts alone enables us to make useful forecasts up to a few days ahead, which of course are of lesser quality than shorter time periods ahead.

In practice, the forecasts will be generated with the frequency of the sampling period (t), which in this case is 15 minutes. This means that starting from T we will have a multiple period ahead forecast (for $T+h$) at each time period. To measure the performance of this moving window forecast, we will calculate the root mean square forecast error for each forecasting horizon separately and for each model by

$$RMSFE_t^{(M)} = \sqrt{\frac{1}{h}\sum_{t=T+1}^{T+h}(y_t - \hat{y}_t^{(M)})^2}, \tag{31}$$

where M stands for the respective model used, and $\hat{y}_{t,M}$ is the respective forecast by each model. At the end of the forecasting exercise, we will have L multiple period ahead forecasts for each model (M) and for all forecasting horizons h. Then we will determine the average horizon RMSFE by

$$\overline{RMSFE}^{(M)} = \frac{1}{L}\sum_{t=T}^{T+L} RMSFE_t^{(M)}. \tag{32}$$

At the end we will have 3 average forecast errors for each of the 3 different physical model data (A,B,C) used in each model. We will compare the forecasting performances of the models to that of each respective physical model forecasts.

5.2 Implementation Suitability

In the sections up until here we have focused on modeling techniques to improve the forecasting capabilities of the software model. Now we define which characteristics of the models are important for the next step of the process which is implementing the model into the software. First of all, the model must run on a on-board computer together with the current software. This brings along the complexity of the model: it should not take so much computing power that it drastically slows down the forecasting procedure. Another important feature is the robustness of the method. As the data encountered can change quite strongly due to the operation area or seasonal differences, the model should not be very dependent on exact or defined settings. Next to that, it should also not need too much length of data to evaluate the parameters, since forecasts are needed as soon as the vessel reaches the operation zone. This means that after acquiring the forecasting results we will evaluate the models on the following three points:

- Computing power and time
- Robustness to data and settings
- Data length needed

We will measure the computing time of each model on the same data sample to assess the first point. The robustness can be tested by using different data and by looking critically the settings used in each method. The third is measurable, as some methods need certain amounts of data to estimate parameters, and might fail to live up to expectation on smaller data sets.

6 Results

The forecasting results of each model will be compared to the physical model forecast performances which, as explained in section 2, do not vary for different forecasting horizons and have the following RMSFE performance: 0.053, 0.057 and 0.048 for data supplier A, B and C respectively. At the end of this section we will use different subsets of the data to analyze whether the RMSFE performances hold.

6.1 Results of the Simple Component Model

Resembling the situation in practice where one could estimate the coefficients of this model at the beginning of the operation by using the first observations that come in, we use the first 10 observations to estimate the simple component model. The coefficients c_1 and c_2 are estimated through linear regression. We find $c_1 = 0.418$ and $c_2 = 0.084$. The respective forecast performances can be found in Table 6. We can see that in these forecasting horizons the simple component model does not improve the physical model forecasts in any case. This is regarded to be due to this model not using the physical model forecasts to its advantage. As we have seen in section 2, it is very difficult to find a trend in the observations themselves but by using the physical properties of the weather models, a lot can be accomplished. The 8 hour ahead ($h = 32$) forecast generated at observation 160 with this model, using metocean data supplier A, is plotted in Figure 8. The grey part is the forecast area. We can use this model to estimate the time series, which gives the result in Figure 9.

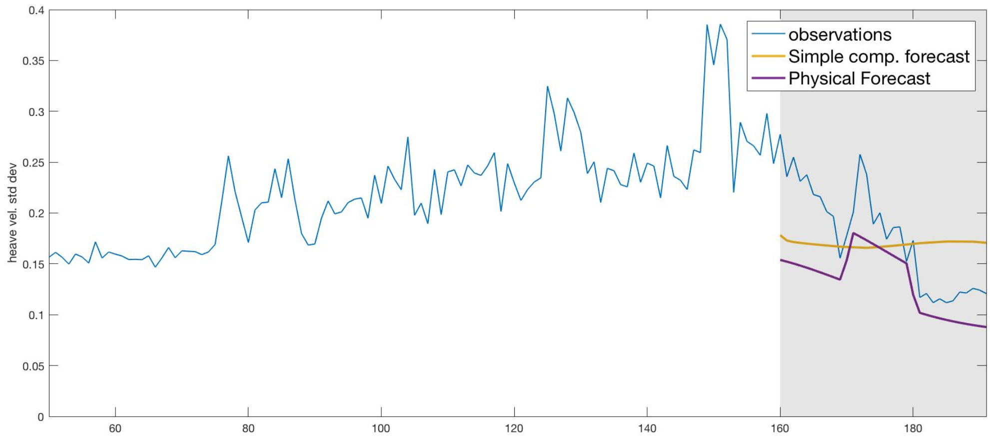

Figure 8: Simple component and physical model forecast for $h = 32$, using supplier A

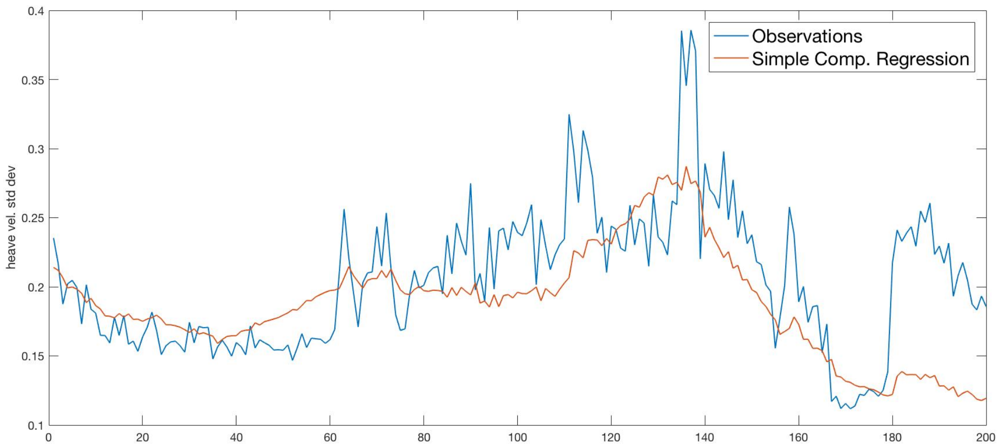

Figure 9: Simple component regression model result

Table 6: RMSFE Performance for all models

Model	Data	12h	8h	4h
	A	0.071 (-33%)	0.067 (-69%)	0.063 (-18%)
Simple Component	B	0.064 (-12%)	0.060 (-6%)	0.058 (-2%)
	C	0.055 (-14%)	0.054 (-11%)	0.051 (-6%)
	A	0.057 (-6%)	0.051 (5%)	0.046 (14%)
Local Level model (a)	B	0.058 (-2%)	0.049 (14%)	0.045 (21%)
	C	0.047 (3%)	0.042 (12%)	0.039 (18%)
	A	0.046 (14%)	0.042 (21%)	0.036 (33%)
Local Level model (b)	B	0.042 (27%)	0.035 (38%)	0.032 (44%)
	C	0.035 (27%)	0.032 (33%)	0.030 (38%)
	A	0.047 (11%)	0.045 (16%)	0.042 (22%)
Dynamic Mixture (a)	B	0.044 (23%)	0.039 (32%)	0.037 (36%)
	C	0.041 (11%)	0.034 (30%)	0.030 (32%)
	A	0.048 (11%)	0.043 (20%)	0.037 (30%)
Dynamic Mixture (b)	B	0.038 (33%)	0.039 (32%)	0.033 (41%)
	C	0.034 (30%)	0.038 (22%)	0.032 (34%)

Average RMSFE and % Improvement compared to the physical model data source (A,B,C) in each respective model for forecasting horizons 12, 8 and 4 hours.

6.2 Results of the Local Level Model, Case (a)

The local level model (a) is the one where during the forecasting procedure the weighting coefficients are re-estimated at each time step, when a new observation becomes available. A burn in sample of 100 and a desired sample size of 1000 ($B = 100$, $B + G = 1000$) is used in the sampling scheme discussed in section 4.1 and on the data of operating window 2. This results in the parameter estimates in Table 7, which are slightly different for each physical forecast input. Details and histograms of the posterior distributions from which they are sampled are given in appendix A.12.

Table 7: Parameter estimates of local level model (a)

Model	σ_ε	σ_η	ϕ_1	ϕ_2
A	0.050	0.002	0.842	0.139
B	0.050	0.002	0.861	0.139
C	0.050	0.002	0.845	0.146

Model parameters using each physical model data source (A,B,C).

This gives the state estimate plotted together with the data in Figure 10. We see here that the state follows general patterns in the data relatively easily, but does not have a strong local variation.

If we consider physical forecast data of supplier A and generate the local level model forecast at one point in time for a horizon of 8 hours, we get the result in Figure 11. The gray area is the forecasting horizon. In this example we can clearly see that there is an advantage of starting the forecast at the estimated level. The physical forecast was on a different level than the actual measurements, and starting the forecast from the estimated state level gives an improved forecast. The forecast is then drawn towards the physical forecast. In this case it leads to a correct estimate of the trend, but it could also result in less accurate estimates. Examples of this will be given in the next section. The weighting coefficients are re-estimated before each forecasting exercise, leading to widely varying coefficients. In some forecasts the coefficient $\hat{\phi}_1$ is estimated as low as 0.64 and $\hat{\phi}_2$ as high as 0.29. Because we forecast multiple periods ahead, this means that the influence of the time series disappears after a few forecasting steps and very strong influence of the physical forecast model is found.

The forecasting performance of local level model (a) expressed by the Average RMSFE can be found in Table 6 for the case of each data supplier. When forecasting 4 hours ahead we find forecasting improvements of 14%, 21%, and 18% considering metocean data supplier A, B and C respectively. For the 8 hour horizon we found forecasting improvements of 5%, 14% and 12% when using metocean data supplier A, B and C. For the 12 hours horizon, 2 out of 3 forecasts under perform on average compared to the physical forecasts (using metocean supplier A and B) and we find a marginal improvement of 3% when using supplier C.

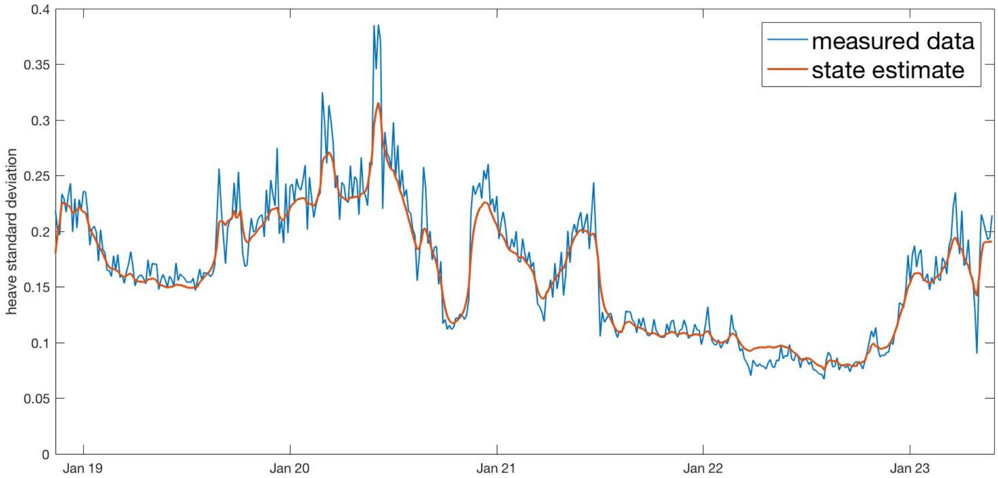

Figure 10: State estimates using the local level model (a)

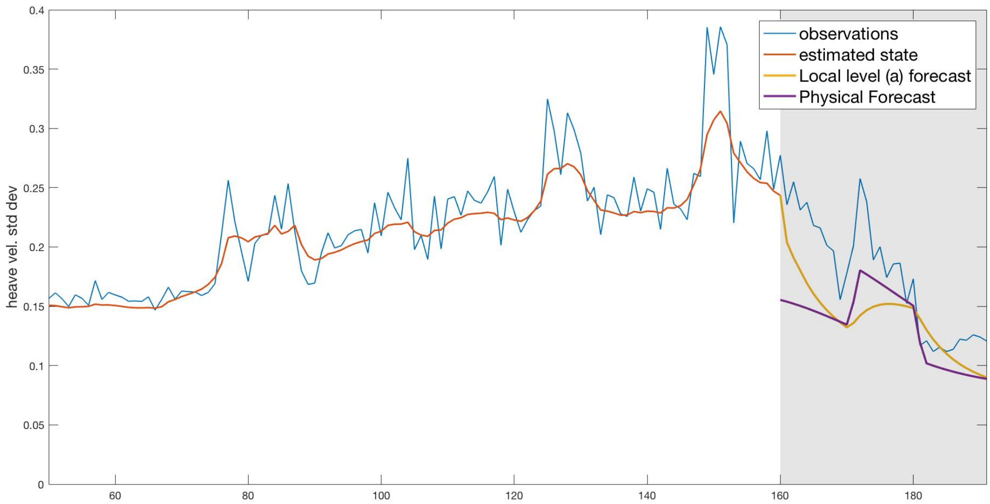

Figure 11: Local level model (a) and physical model forecast for $h = 32$, at observation 160

6.3 Results of the Local Level Model, Case (b)

Case (b) of the local level model is the one where we use the same weighting coefficients throughout the whole forecasting procedure. The parameter estimates are similar to the estimates in model (a) when using the full data set of operating window 2, which is $\sigma_\varepsilon = 0.050$ and $\sigma_\eta = 0.016$ for all sub models A, B and C. This results in similar state estimates, but slightly different forecasting characteristics, as can be seen in Figure 12. Again, this figure is made by using physical forecast data A. Compared to Figure 11 we see that it in this forecasting snapshot the weight is less towards the physical model forecast and more to the observation data, as the forecast reacts less strong to the low values of the physical model forecasts visible in the gray area. This is due to the fact that model (a) has the freedom to calculate the ϕ's before each forecasting exercise based on the data that is available up until then. The figure shows a positive scenario, where the local level model forecast is drawn towards the physical forecast and that it results in correct forecasts. This may not always be the case. In Figure 13 we see an example forecast where the physical forecast draws the model forecast in the wrong direction. This case still leads to an improvement compared to the physical forecast as it at least starts at the correct level. Figure 14 shows a case where the physical forecast is more accurate because it accounts for a large shift in the level. Here the local level model reacts to slowly.

Fixing the coefficients in $\boldsymbol{F}$ as $\phi_1 = 0.9$ and $\phi_2 = 0.1$, as done in this model, leads to improved forecasting results overall. The result can be seen in Table 6. In the 4 hour ahead forecasts we find improvements between 33% and 44% on the average RMSFE, in the 8 hour ahead forecasts we find improvements between 21% and 38% and in the 12 hour ahead forecasts we find improvements between 14% and 27%, all depending on which metocean data supplier was used. This means this model leads to stronger improvement in all cases compared to local level model case (a) where the coefficients are re-estimated before each forecasting procedure.

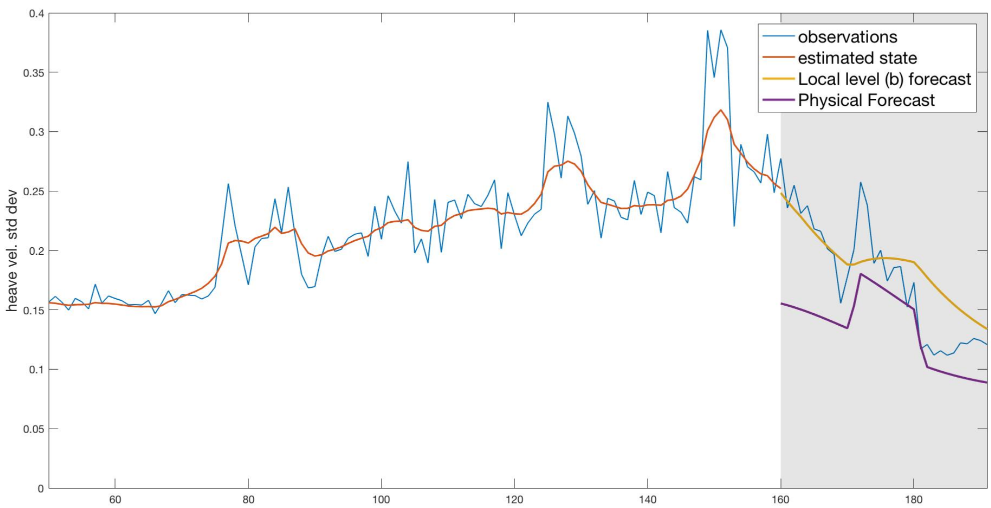

Figure 12: Local level model (b) and physical model forecast for $h = 32$, at observation 160

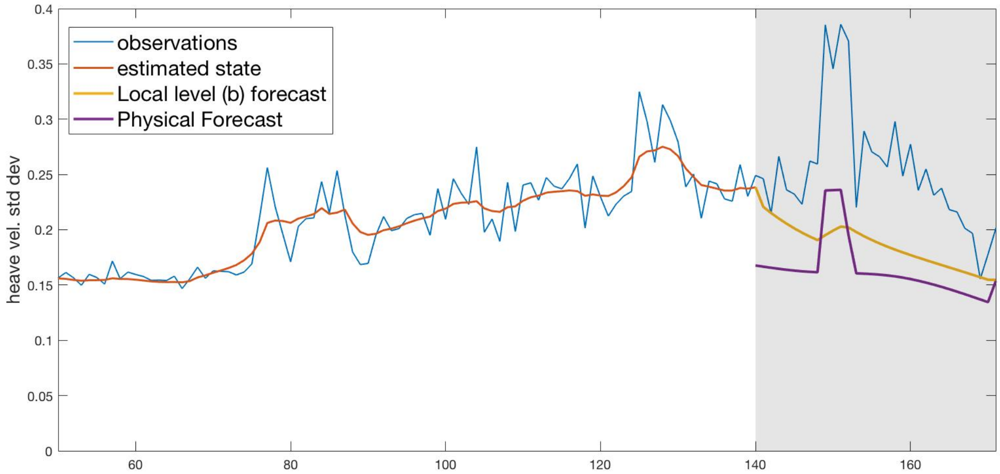

Figure 13: Local level model (b) and physical model forecast for $h = 32$, at observation 140.

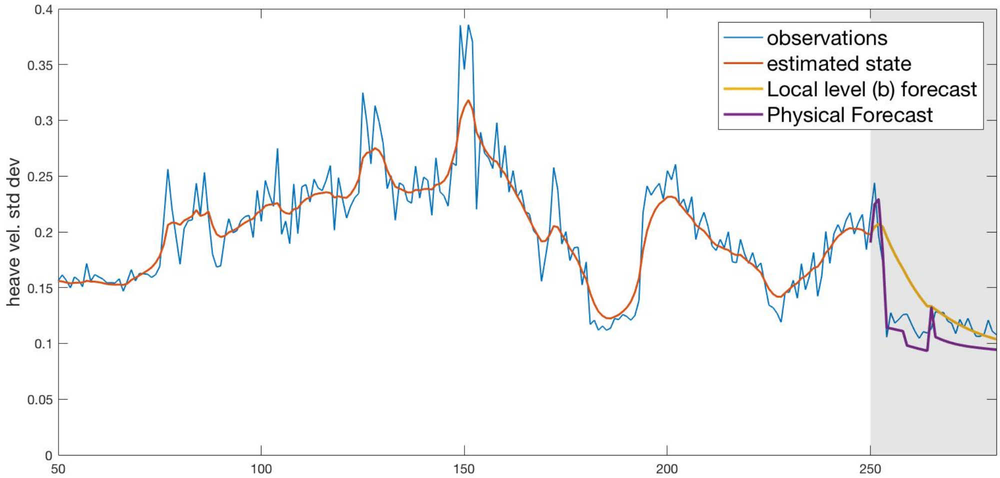

Figure 14: Local level model (b) and physical model forecast for $h = 32$, at observation 250

6.4 Results of the Dynamic Mixture Model, Case (a)

In the dynamic mixture models we estimate a state space model which has breaks in the level of the state, and in case (a) we re-estimate the model, and especially the weighting coefficients in the state equation, in the forecasting procedure each time a new observation becomes available. The dynamic mixture model applied to the full data of operating window 2, using the physical forecast A, we find the parameters in Table 8. The outcome is in line with prior expectation, with a slightly higher signal to noise ratio ($q \approx \frac{1}{2}$). It turns out that the outcome is not very sensitive to the starting values, which can be expected from the amount of observations.

Table 8: Parameter estimates for the Dynamic Mixture model

Model	σ_u	σ_v	ϕ_1	ϕ_2
A	0.019	0.008	0.856	0.142
B	0.019	0.008	0.873	0.127
C	0.019	0.008	0.859	0.141

Model parameters using each physical model data source (A,B,C)

The result of the estimated state with breaks is given in Figure 15. For this purpose, as discussed before, we find it desirable that the state breaks relatively quickly, and in this case each break is justified by the requirement of staying on the level for more than 3 observations, as this means staying there for an hour. That would be long enough to account for the change in expected motions during an offshore operation.

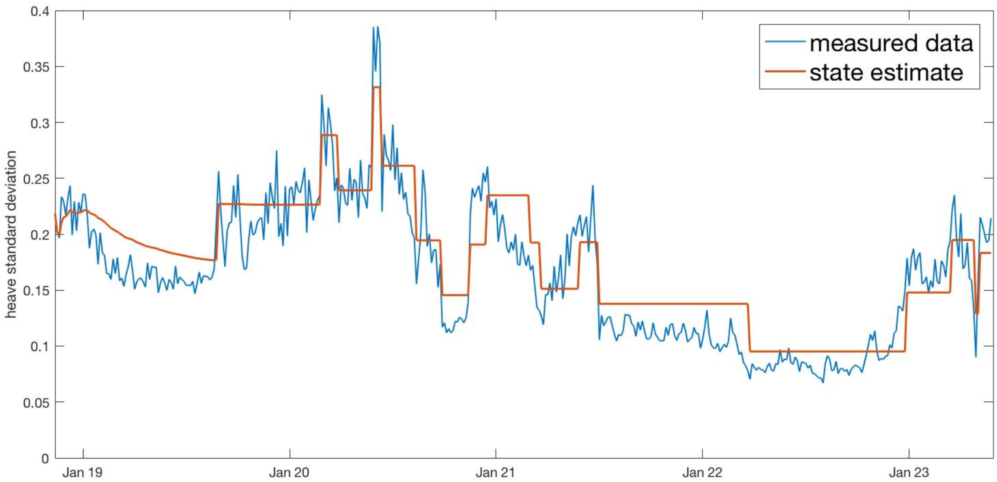

Figure 15: State estimate of a Dynamic Mixture on operating window 2

Generating a forecast with the dynamic mixture model (a), at observation number 160 for this data set, for a forecasting window of 32 periods (8 hours) gives the result in Figure 16. The gray area is the forecasting horizon. Here we have also plotted the physical forecast related to this model (supplier A). Again we see the same advantage of using the combined data as we did in the local level model. The forecast starts at the estimated level of the state, instead of at a seemingly very incorrect place, and then it is drawn towards the physical model forecast as it runs through the multiple period ahead forecast. In this case it gives a desirable effect, but of course this may also give an incorrect and undesirable effect in case it strays away from the true values due to poor quality of the physical model forecast, which was shown in section 6.3. We also see that the forecast will not react as strong as the original physical model forecast to inputs such as heading change. In practice this will matter less, as the forecast will be generated for a given heading as the captain or operator will insert it as input.

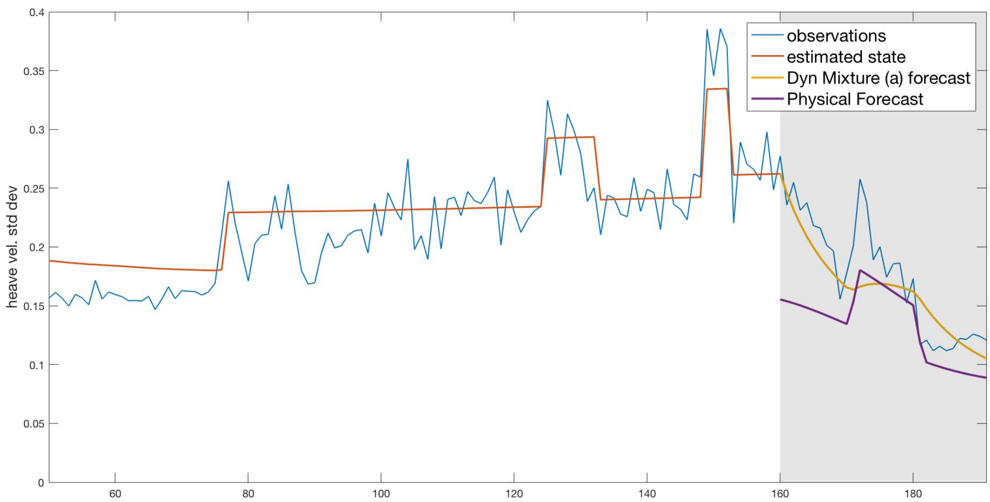

Figure 16: Dynamic mixture and physical model forecast for $h = 32$, at observation 160

The dynamic mixture model is subject to the quality of the physical model that is used and this will also influence the forecasting performance of the model. For each respective model we have calculated the average root mean square error when the model is used on the data of operating period 2, for forecasting procedures with 4, 8 and 12 hours ahead forecasts. The results for can be found in Table 6. The biggest improvements of 22%, 36% and 32% are found on the shortest horizon of 4 hour ahead forecasts, which is as expected because there the time series contribution was expected to be the most. The 8 hour ahead forecasts with this model give improvements of 16%, 32% and 30% when using metocean supplier A, B and C respectively. The results for the 12 hour horizon are also positive and are 11%, 23% and 11% for the respective data suppliers, and can certainly be seen as an improvement. Whether 11% improvement is enough reason to apply this complex model to the software is something that should be analyzed in follow up case studies and discussions with engineers.

6.5 Results of the Dynamic Mixture model, Case (b)

In case (b) of the dynamic mixture model the weighting coefficients in the state equation are not re-estimated during the forecasting exercise. This leads to the results in Table 6. In the 4 hour ahead forecasts improvements are found of 30%, 41% and 32% for metocean data supplier A, B and C respectively. In the 8 hour ahead forecasts improvements of 20%, 32% and 22% are found depending on the data supplier. Finally in the 12 hour ahead forecasts, improvements of 11%, 33% and 30% are found when metocean data supplier A, B or C is used respectively. The adjustment compared to case (a) gives slight improvements, in most cases a few percent. Again the same forecasting moment as in case (a) can be seen in Figure 17, at observation number 160. Again this leads to a positive outcome for forecasting due to starting at the estimated state. On the contrary, Figure 18 shows that when the state estimate is not correct, that is, the level did not shift back, we see that the forecast performance is poor and the physical forecast would have given better results for that moment. The advantage of including the level shift can be seen in Figure 19, as here the state shifts dramatically at observation 255 leading forecasts on the proper level. Clearly, when the physical forecast strays, the model will follow. This inevitably leads to poor performance. An example can be seen in Figure 20, when the forecast is made at observation 305.

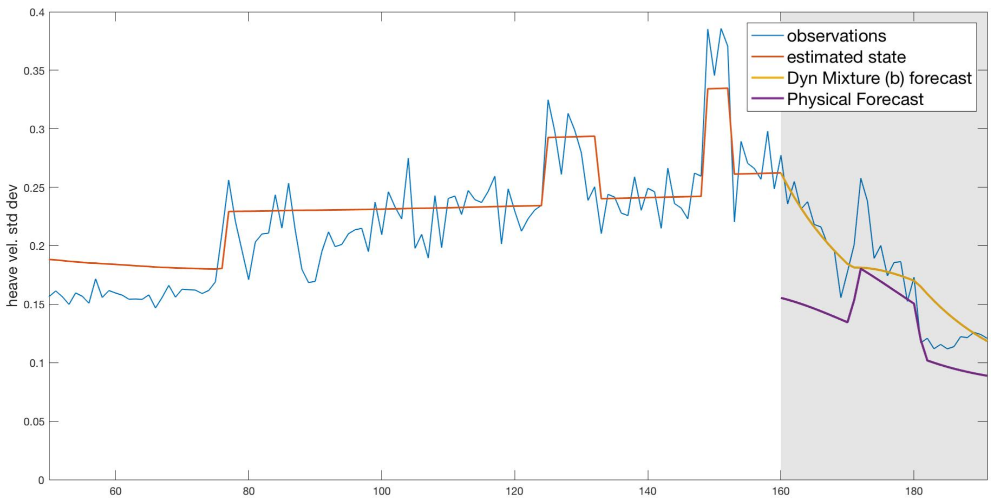

Figure 17: Dynamic mixture and physical model forecast for $h = 32$, at observation 160

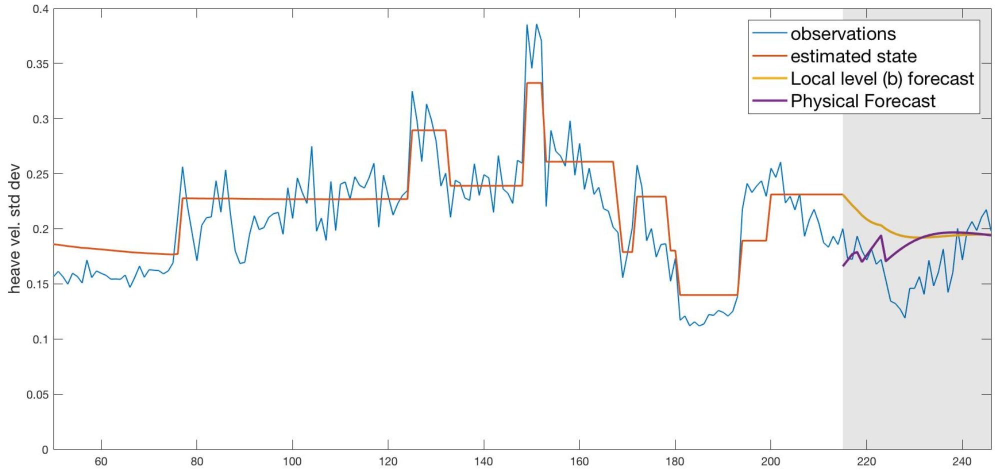

Figure 18: Dynamic mixture and physical model forecast for $h = 32$, at observation 215

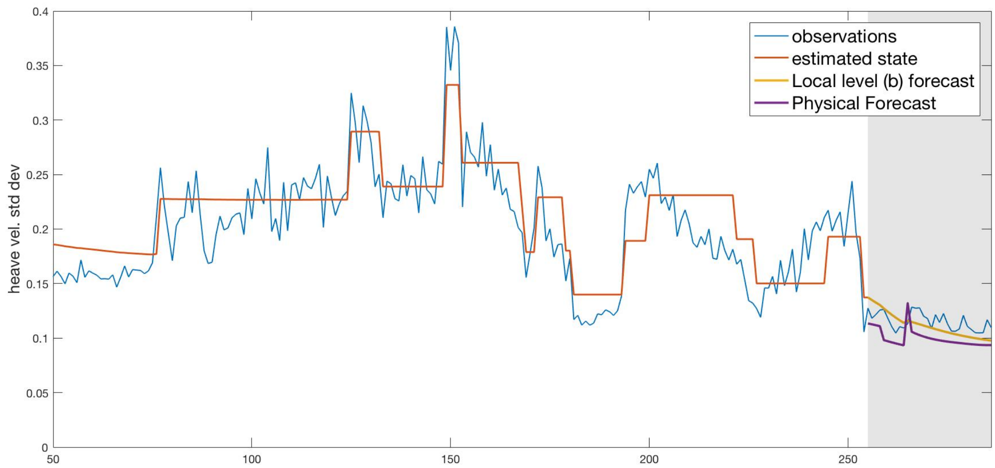

Figure 19: Dynamic mixture and physical model forecast for $h = 32$, at observation 255

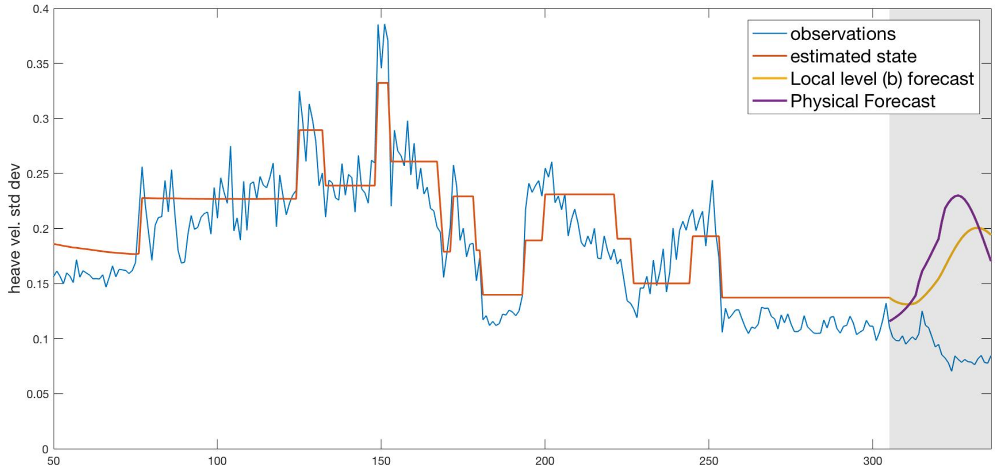

Figure 20: Dynamic mixture and physical model forecast for $h = 32$, at observation 305

6.6 Performance Review

The results presented in the previous section give the average root mean square errors of the forecasting procedures. The figures that come out of that calculation are already not a single test outcome, but as explained in section 5 they are the averages of a large amount of multiple period ahead forecasts. In this section we evaluate whether the results hold when tested on other subsets of the data. This is done by repetitively moving the start of the forecasting window 10 observations into the future. In each new subset the model is recalculated and the forecasting performance is determined, leading to 18 test cases for each metocean data supplier and each model. This allows us to investigate whether the performances found before hold or whether it was found by coincidence. We compare the forecasting performance of each of the models when using the metocean data suppliers A, B and C. It is also compared to the forecasting performance of the physical model when using the same metocean data. We do this only for the 8 hours ahead forecasts.

Figures 21, 22 and 23 show the result of this benchmarking procedure, if the line of the model is below the blue line of the physical forecast, an improvement is found. From these figures we make up that the increase in forecasting performance stays valid for most of the investigated cases, as the RMSFE is lower than that of the physical model forecast. Also, local level model (b) where the weighting coefficients are not re-estimated, seems to outperform in general, which is especially clear when looking at the cases where data of supplier B and C are used. When using supplier A a different characteristic is visible.

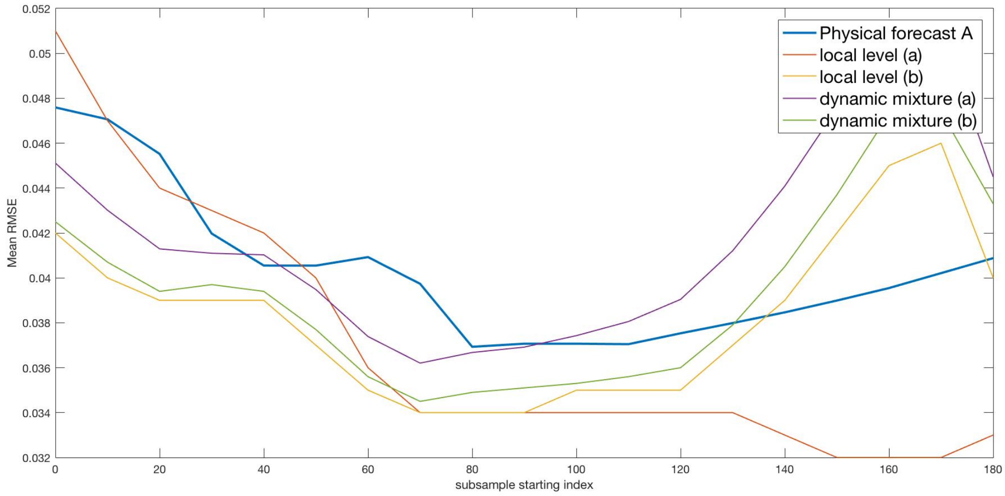

Figure 21: Forecast performance over multiple subsamples for each model compared to the physical model forecast, using metocean data supplier A

When using metocean data supplier A, a sharp decrease can be seen in the performance of local level model (b) and both the dynamic mixture models, when the subsample reaches the starting index of 120. This is likely due to poor quality of data supplier A, and this effects the performance of the models as they tend to forecast towards this data, which is visible in Figure 26 in appendix A.2. The reason that local level model (a) outperforms in this case is likely due to the flexibility in the coefficients, as it may decrease the weight of the physical model forecast strongly in this part. In the dynamic mixture this freedom is far less because of the strong prior put on the sum of $\phi_1 + \phi_2$ to lie between 0.999 and 1.001.

When using metocean data supplier B and C, more consistent results are found. The improved performances compared to the physical model found in the previous sections hold in general, as well as the order of performance between the respective models. The downward slope in Figure 22 seems to be due to the performance of the physical forecast itself. This is less apparent when using metocean data supplier C. Here the change in accuracy of the physical forecast is not so strong but still the performance of the model forecasts changes strongly with the various subsets.

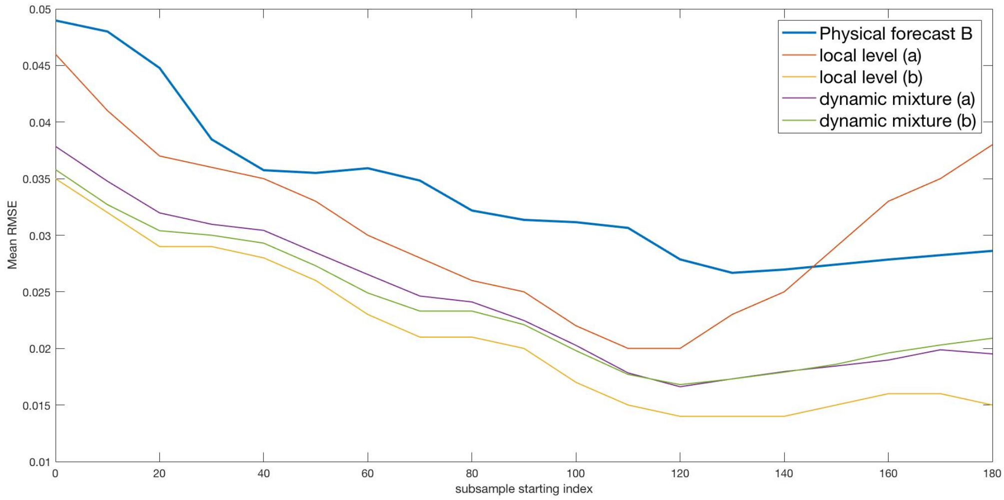

Figure 22: Forecast performance over multiple subsamples for each model compared to the physical model forecast, using metocean data supplier B

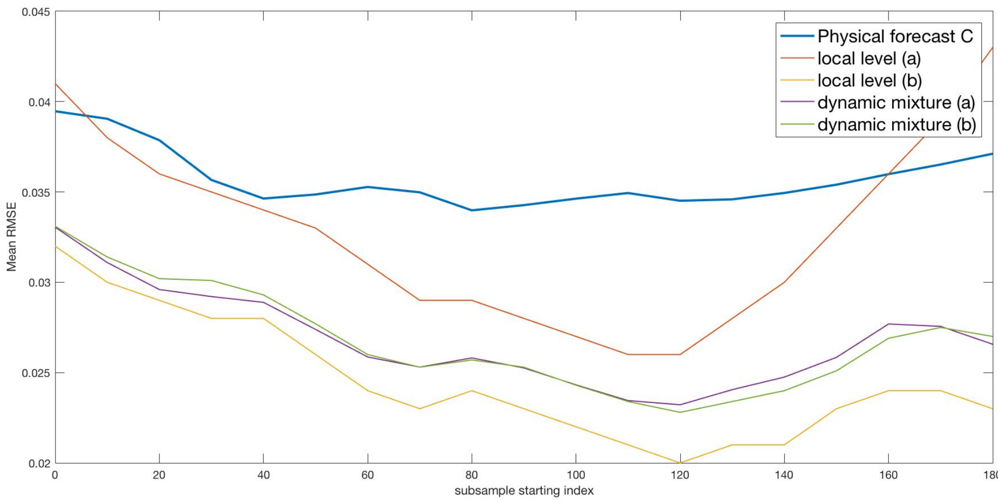

Figure 23: Forecast performance over multiple subsamples for each model compared to the physical model forecast, using metocean data supplier C

7 Discussion

Considering the forecasting performance we have found that the local level model with fixed coefficients leads to the strongest improvement compared to only using the physical model. Next to the forecasting performance we have defined other characteristics as well to determine the suitability of the models.

First of all, the simple component model led to a basic benchmark for the other more complex models. The shortcoming of this model is that it follows the observational data points strongly and does not use a stable level to generate the forecast from. Also, it may only work well when using this type of observations, as the linear coefficient we determined might not be applicable on other measurements such as roll and pitch. On the other hand, its computing time is very low, it is robust to different data and specific settings. Next to that we do not need much data before applying the method.

The local level model with re-estimated coefficients led to values for ϕ_1 and ϕ_2 close to the prior expected values when the sample data was fairly large. When this model was tested on smaller sets of observations (less than 50) the method fails to properly estimate $\boldsymbol{F}$, where sometimes the sum of ϕ_1 and ϕ_2 does not come close to 1 (sum as low as 0.3 is found). During the forecasting procedure, and especially in practice, the model should be able to cope with these sample sizes and therefore the method is not optimal, as it would not work in practice when an operation has just started. In the forecasting procedure where the smallest sample size was 135 observations, these problems where mostly avoided and combinations such as $\hat{\phi}_1 = 0.64$ and $\hat{\phi}_2 = 0.29$ can be found. Apparently this flexibility does not lead to a positive effect on the forecasting performance.

The problems encountered above are mostly solved by fixing ϕ_1 and ϕ_2, leading to the fixed coefficient model. First of all, when using the fixed coefficients we see that the minimum data requirement drops completely. When using suitable starting values for σ_ε and σ_η we can actually start forecasting with the model at the beginning of the data set. Further, the forecasting performance is improved strongly. Most likely because now there are no inconsistent combinations of the coefficients and the relative high weight of ϕ_1 seems to be beneficial.

The dynamic mixture model proved to be a complex model to estimate properly. First of all the prior specification strongly influences the way the level shifts are evaluated: adjusting them slightly can lead to finding either a lot of shifts or to finding almost none. This means the method is not robust to various data and needs exact settings. It means that the method might fail when moving to a new region or when a different season is entered, where the sea state characteristics change and the priors will have to be redetermined. Considering the sample size we find that relevant results are still found when using more than 80 observations. This is a full day of data, which is not optimal as mentioned above for practical reasons.

The computing time to estimate the dynamic mixture model is approximately 20 times higher than for the local level model with re-estimated coefficients and almost 40 times higher than for the fixed coefficient model (4.77 seconds versus 0.23 and 0.13 seconds on the same data set, with an 2.2 GHz Intel Core i7). This has implications for the implementation into the software. This research focuses on a simplified situation as we are looking at one degree of freedom, one motion measurement and known heading. In practice we will require the software to generate forecasts for multiple motions by giving input for the heading, resulting in an interactive type of forecasting. If the captain has to wait too long each time a forecast is generated this is a major drawback of the model.

Fixing the coefficients to the pre-determined value leads to improved performance of the dynamic mixture model. This removes the rejection and resampling scheme and improves its speed. Now the computing time drops to 3.45 seconds measured on the same data set as the other models, which is still 15 to 25 times slower than the local level models. The sensitivity to prior parameters remains and the amount of data to find relevant level shifts does not improve either.

8 Conclusion

Improving the operating efficiency of vessels working on offshore projects can turn out to be worth a lot of money. An element of improving the operation is correct planning and estimation of workability, as in this case time is really money. For this reason we investigated whether it is possible to improve the forecasting accuracy of a vessel operating at sea by combining time series measurements with physical model forecasts in a model. This could lead to an improved planning of the operation and subsequently to cost reductions.

To achieve the improvement, we use and compare the forecasting capabilities of multiple models to find the most suitable and best performing way to combine the time series measurements with the physical model forecasts. The first one is a simple component model, which makes use of a linear combination of the physical forecast and one autoregressive parameter where the coefficients are estimated through least squares regression. Based on the average root mean square errors it does not perform better than the physical model forecasts. It shows that a more involved model is needed to find improvements in forecasting.

The second one is based on a local level model where the physical model forecast is added through the state equation. A weight is given to the previous state estimate and the physical model. We use two variants of this model during the forecasting exercise, one where the weighting coefficients are re-estimated in each time step and one where they are estimated up front and then kept fixed throughout forecasting. The first model works relatively well when enough data is available but with less data the freedom in the coefficients lead to inconsistent results and subsequently to poor performance during the forecasting exercise. When using the fixed coefficient variant this is solved and the forecasting performance increases in all sub-models and forecasting horizons compared to the first model, with improvements in average root mean square error of 14-44% compared to the physical model forecasts.

The third model is a dynamic mixture model which includes the physical model forecast data in the same way as in the local level models through the state equation. This model includes the possibility for level shifts which allows the model to react to sharp increases or decreases of the estimated state. With the dynamic mixture model with varying coefficients we find improvements of the forecasts compared to the current situation of 11-36%. This model is sensitive to prior parameter settings and if it is not correctly specified the model fails to find relevant level shifts. Next to that, it needs at least a day of measurement data to function properly and it is 20 and 40 times slower in computation than the local level models with re-estimated and fixed coefficients respectively. To ensure proper state estimation a rejection scheme for the coefficients was necessary, which leads to 50 times the amount of sampling. Fixing the parameters solves this problem and also increases the forecasting performance. It does not solve the other disadvantages of the model regarding the complexity of estimating the level shifts.

The models are run on multiple data samples to investigate whether the positive results where found by coincidence or due to a beneficial sample data. This leads to the conclusion that when using metocean data supplier B and C the results are found to be very consistent and the local level model with fixed coefficients leads to the best forecasting improvement in almost all tested cases. When using the data of supplier A this is less evident, and it appears that due to accuracy issues of this data it sets off the performance of the models which do not allow for different weighting coefficients.

Considering the results above we recommend to use the local level model with fixed coefficients as a base to implement an improvement in the on-board forecasting system of MO4. This model is more suitable for this purpose due to its simplicity and independence of specific data and it can start to run shortly after starting the operation.

Future research can focus on attaining dynamic coefficients for the fixed coefficient model without the drawbacks of the varying coefficient model. It would be beneficial if the weighting coefficients depend on the recent performance of the specific physical model forecast, which was not accomplished by re-estimating the coefficients in the models. On the other hand, the level shifting behavior of the dynamic mixture model ensures that this model can react to peak responses adequately, where the local level model would not estimate the level to go up so high and therefore it may miss peaks. In offshore operation, this would be a desirable feature. Future research could investigate how to setup such a type of model with less of the computational drawbacks and find robust specifications such that the model can run on smaller and more diverse data sets. Further, the study has limited itself to the investigation of one variable of vessel motion. In a future study it would be interesting to see whether the method is useful when extended to different measurements, or whether combining multiple variables from the measurements can improve the performance.

The aim of the study has been reached as we have found a model that combines time series measurement data with physical model forecasts and is able to give superior forecasts compared to the current situation of using only the physical forecasts, and is considered to be a suitable basis for implementation in the current software. This finding is a relevant contribution as it results in better forecasting performance and combining these data sources has not been done before.

References

Carlin, B. P., Polson, N. G., & Stoffer, D. S. (1992). A monte carlo approach to nonnormal and nonlinear state-space modeling. *Journal of the American Statistical Association*, *87*(418), 493–500.

Carter, C. K., & Kohn, R. (1994). On gibbs sampling for state space models. *Biometrika*, *81*(3), 541–553.

Durbin, J., & Koopman, S. J. (2012). *Time series analysis by state space methods* (Vol. 38). OUP Oxford.

Gerlach, R., Carter, C., & Kohn, R. (2000). Efficient bayesian inference for dynamic mixture models. *Journal of the American Statistical Association*, *95*(451), 819–828.

Giordani, P., & Kohn, R. (2008). Efficient bayesian inference for multiple change-point and mixture innovation models. *Journal of Business & Economic Statistics*, *26*(1), 66–77.

Greenberg, E. (2012). *Introduction to bayesian econometrics.* Cambridge University Press.

Hamilton, J. D. (1994). *Time series analysis* (Vol. 2). Princeton university press Princeton.

Jazwinski, A. (1970). *Stochastic processes and filtering theory, 1970.* Academic Press.

Massie, W., & Journée, J. (2001). Offshore hydromechanics. *Delft University of Technology: Delft, The Netherlands*.

McCulloch, R. E., & Tsay, R. S. (1993). Bayesian inference and prediction for mean and variance shifts in autoregressive time series. *Journal of the American Statistical Association*, *88*(423), 968–978.

Petris, G., Petrone, S., & Campagnoli, P. (2009). Dynamic linear models. In *Dynamic linear models with r* (pp. 31–84). Springer.

Shephard, N. (1994). Partial non-gaussian state space. *Biometrika*, *81*(1), 115–131.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *the Annals of Statistics*, 1701–1728.

Varela, J. M., & Soares, C. G. (2011). Interactive simulation of ship motions in random seas based on real wave spectra. In *Grapp* (pp. 235–244).

A Appendix

A.1 MO4 Vessel Motion Forecasting

- Text and figures adopted from the MO4 Theoretical Manual -

MO4 is a vessel motion forecast tool that is installed on the ship's bridge. The system provides a motion forecast that is relevant for the operations in the coming hours or days. It is both a forecasting as well as a monitoring system due to its integrated sensors. The interface has been designed to be simple and intuitive. The presented data allows the captain and OIM to fully optimise their short term planning. MO4 increases the workability and safety of all weather critical operations.

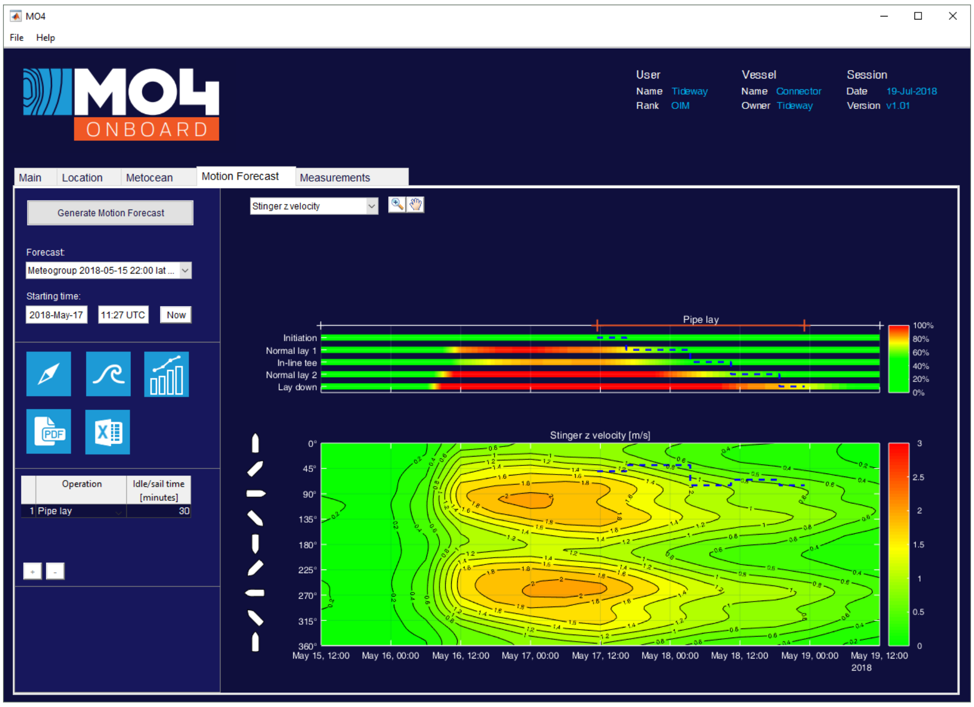

Figure 24: MO4 interface showing results of a multiphase operation

The figure below gives a high level interpretation of how the MO4 system is placed within the offshore operation and vessel.

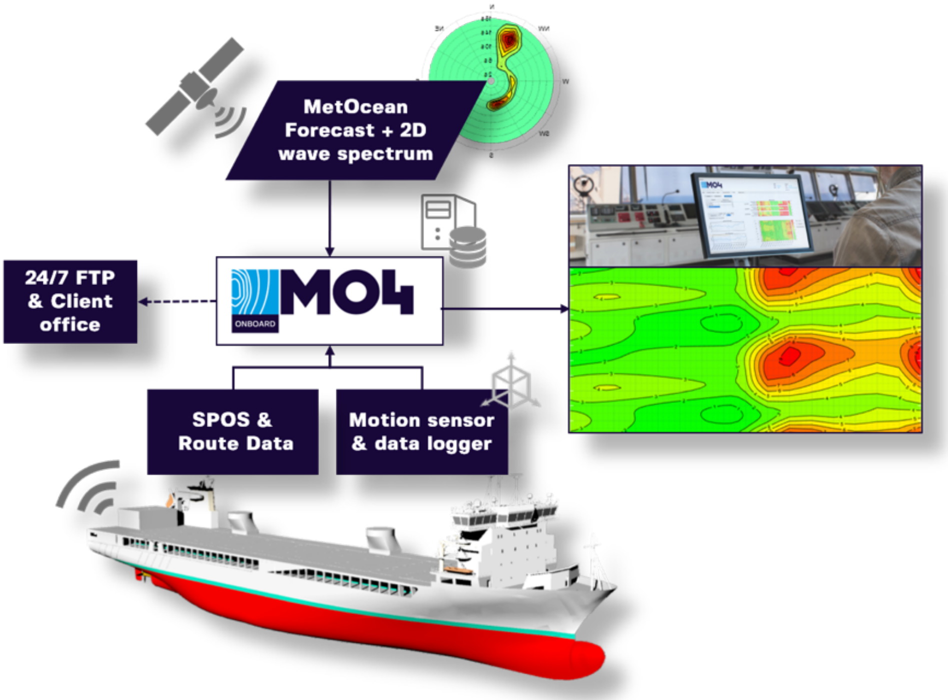

Figure 25: MO4 system

1. The MO4 computer contains a database that includes all the data relevant for the motion calculation except for the metocean forecast data.

2. Metocean forecasts are downloaded either via FTP-connection of directly through an API-connection with the metocean data provider.

3. Based on the user input on vessel operational data, criteria and vessel draft the MO4 algorithm can calculate the motions based on the forecasted wave data. This is what is referred to as forecasted motions.

4. Using the user specified or pre-loaded operational criteria the MO4 tool can define for each operation the total utilisation over time for any given vessel heading.

5. Using the input from the motion sensor the user can assess the accuracy of the previous forecast.

A.2 Forecast data of suppliers A, B and C

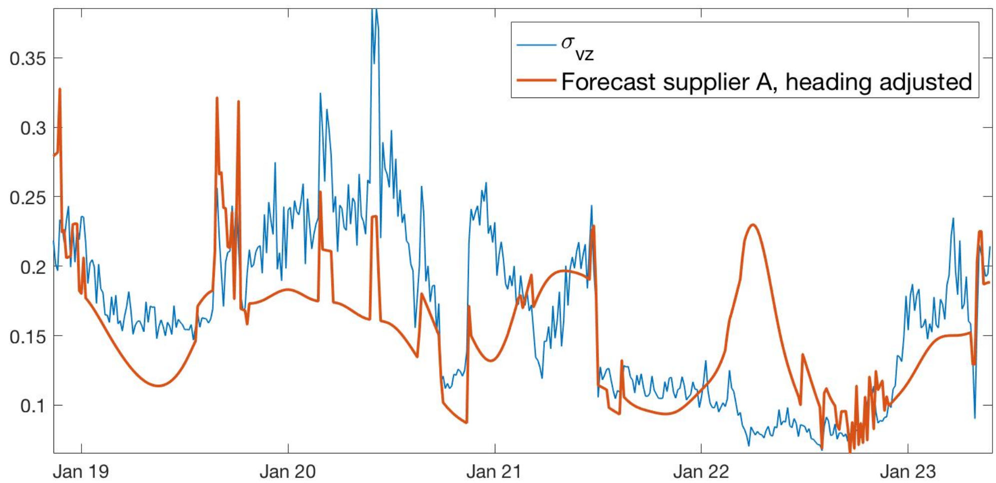

Figure 26: Forecast of supplier A adjusted for the measured headings of operating window 2

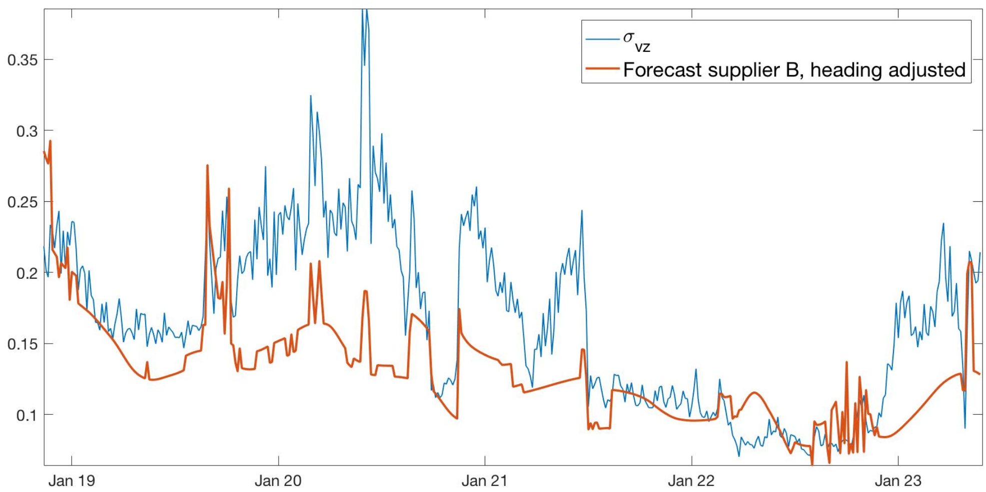

Figure 27: Forecast of supplier B adjusted for the measured headings of operating window 2

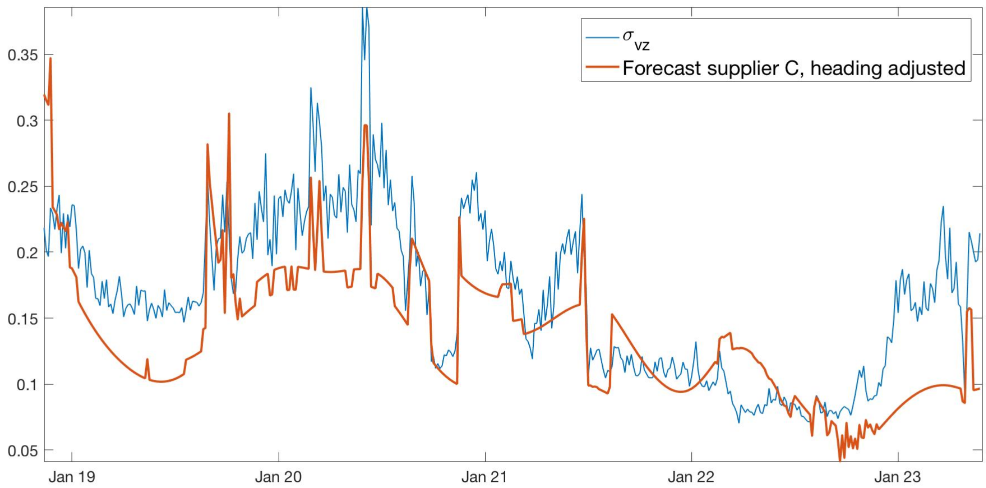

Figure 28: Forecast of supplier C adjusted for the measured headings of operating window 2

A.3 Data with Heading

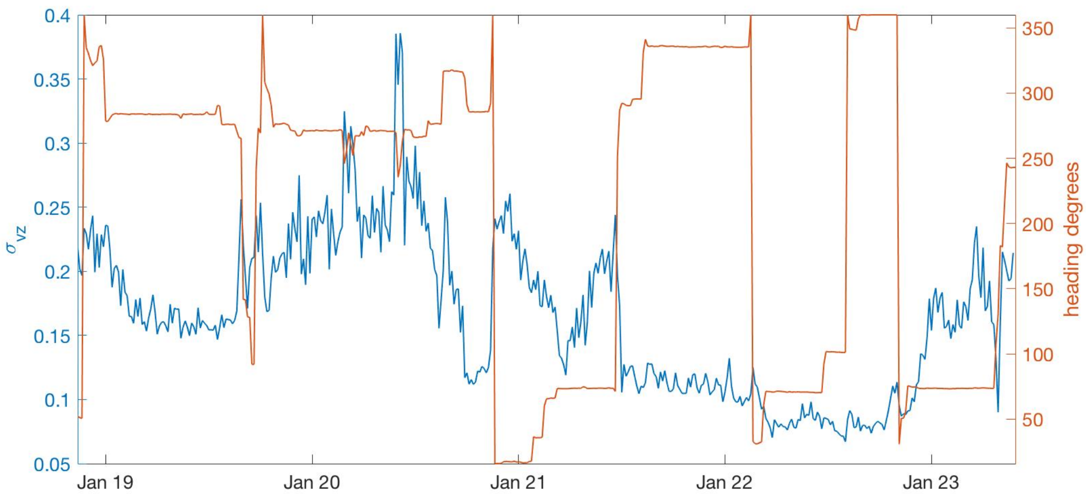

Figure 29: Heave velocity standard deviation including measured heading

A.4 Data for Prior specification

We base our prior specification on investigation on the data of one operating window earlier. This period lies about 1 week from the next.

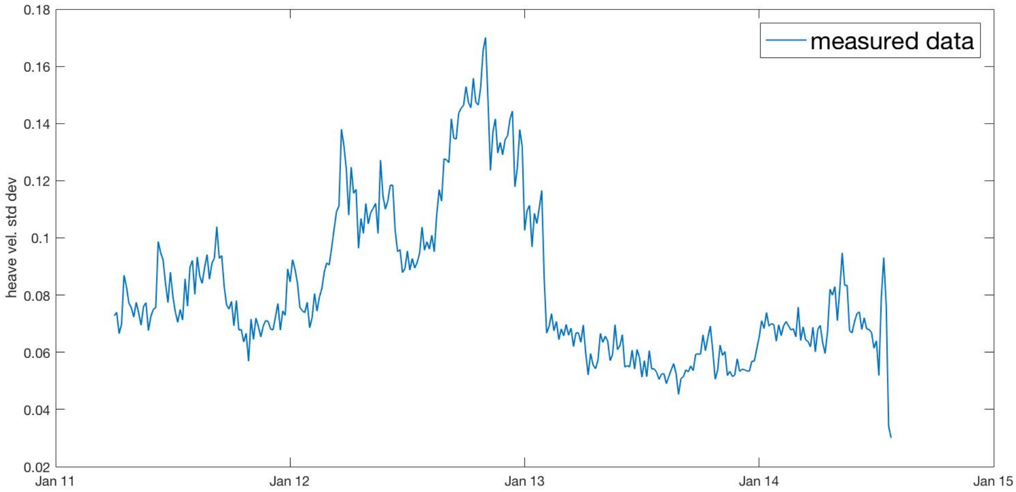

Figure 30: Data of the previous operating window, or operating window 1

A.5 Bayesian versus Frequentist view on probability

The difference in characteristics between a Bayesian and a Frequentist view on probability are explained in the following example adjusted from Greenberg (2012).

Example: From a frequentist point of view, probability theory could tell you something about the distribution of the data for a given parameter (θ). If we take the example of a coin toss, we can view the data (y) as an outcome of a large number of repetitions of tossing the coin n times for some θ. The parameter itself is not given a probability distribution. From the the subjective probability view, we regard θ as an unknown quantity. As we have uncertainty about its value we regard it as a random variable. We might have a belief about its distribution before seeing the data, which we assign as its *prior* distribution $\pi(\theta)$. Conditioning on having observed the data y, the distribution of the random variable θ is constructed as $\pi(\theta|y)$; the posterior distribution. Bayesian inference centers on this posterior distribution. The review of literature that will follow next makes use of this point of view in statistics.

A.6 Linear Relation Wave Height & Vessel Motion

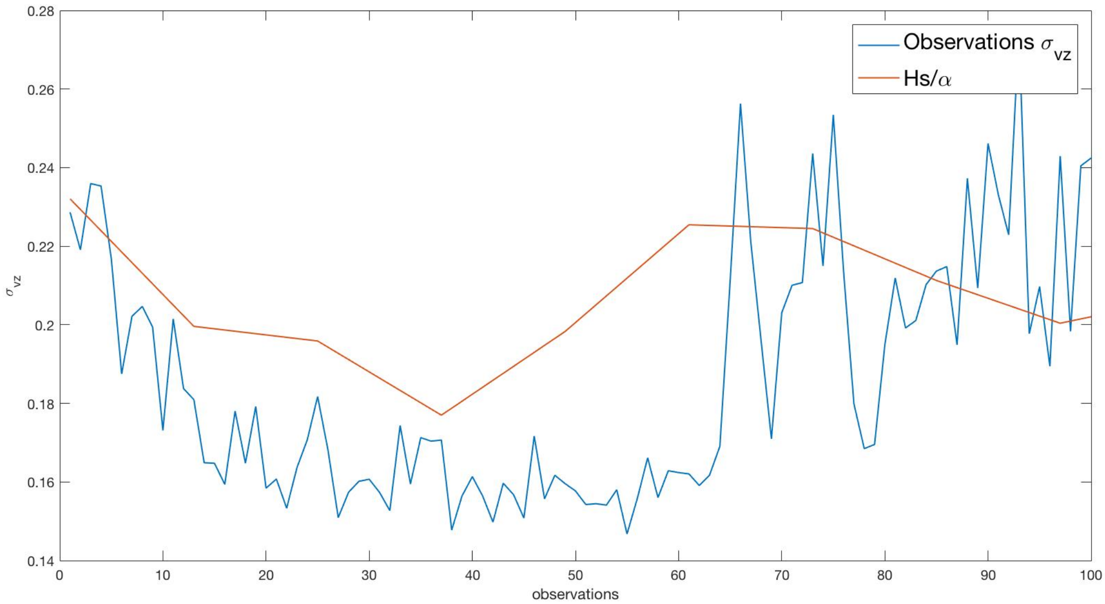

Figure 31: Linear relation between significant wave height H_s and σ_{v_z} with $\alpha = 6.5$

A.7 General Gibbs Sampling Algorithm

In this method, different sets of variables are called 'blocks', which can be vectors of variables, or the parameters to estimate. First of all we have the nonstandard joint distribution we are interested in represented by $f(x_1, ..., x_d)$ for d blocks. Then we have the conditional distributions $f(x_i|x_{-i})$ for which necessarily simulation algorithms should be known, and where x_{-i} are all the variables in the joint distribution other than x_i. Then the following algorithm given by Greenberg (2012) can be used:

Gibbs Algorithm with d blocks

1. Choose a starting values $x_2^{(0)}, ..., x_d^{(0)}$.

2. At the first iteration, draw

$$\begin{aligned}
&x_1^{(1)} \text{ from } f(x_1|x_2^{(0)}, ..., x_d^{(0)}) \\
&x_2^{(1)} \text{ from } f(x_2|x_1^{(1)}, x_3^{(0)}, ..., x_d^{(0)}) \\
&\quad\vdots \\
&x_d^{(1)} \text{ from } f(x_d|x_1^{(1)}, ..., x_{d-1}^{(1)}).
\end{aligned}$$

3. At the gth iteration, draw

$$
\begin{aligned}
&x_1^{(g)} \text{ from } f(x_1|x_2^{(g-1)}, ..., x_d^{(g-1)}) \\
&x_2^{(g)} \text{ from } f(x_2|x_1^{(g)}, x_3^{(g-1)}, ..., x_d^{(g-1)}) \\
&\quad\vdots \\
&x_d^{(g)} \text{ from } f(x_d|x_1^{(g)}, ..., x_{d-1}^{(g)})
\end{aligned}
$$

until the desired number of iterations is obtained.

After the Markov chain has converged (at iteration $g = g^*$), the simulated values $\{x_i^{(g)}, g \leq g^*\}$ can be used as a sample from the joint posterior distribution $p(x_1, ..., x_d|y)$. Next to that the parameter estimates can be calculated by (33), after defining a burn in sample of B and a desired sample size B+G.

$$\hat{x}_d = \frac{1}{G-B} \sum_{g=B+1}^{B+G} x_d^{(g)} \tag{33}$$

This base algorithm will be used and expanded in the next section to estimate our model. Important elements in this are the state estimates, the prior definitions, the likelihood functions and the posterior distributions.

A.8 Normal Lemma

Here we give the normal lemma for variables z_1 and z_2 that have a joint normal distributions:

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \sim N\left[\begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}\right]. \tag{34}$$

Then the distribution of z_2 conditional on z_1 (state conditional on data) is $N(\tilde{\mu}_2, \Sigma_2)$ where

$$
\begin{aligned}
&\tilde{\mu}_2 = \mu_2 + \Omega_{21}\Omega_{11}^{-1}(z_1 - \mu_1) \\
&\Sigma_2 = \Omega_{22} - \Omega_{21}\Omega_{11}^{-1}\Omega_{21}.
\end{aligned} \tag{35}
$$

A.9 Maximum Likelihood Parameter Estimation

To estimate the parameters ($\boldsymbol{\theta}$), we can maximize the likelihood of the parameters of the model in combination with the data. However, the observations y_t and y_{t+1} are not independent because of the dynamic specification of the state equation for μ_{t+1}, which normally simplifies this procedure as then the joint pdf can be written as a product of the individual probability density functions. Using the law of conditional probability, the joint pdf $p(y|\boldsymbol{\theta})$ can be rewritten as the product of conditional pdf's. Then, as loglikelihood is generally used in practice, this gives the loglikelihood as defined in chapter 7 of Durbin and Koopman (2012) as

$$logL(y) = \sum_{t=1}^{T} log\ p(y_t|y_{1,t-1}) \tag{36}$$

The conditional distribution of y_t given the data and the parameters is again normal, which gives the prediction error decomposition:

$$logL(y) = -\frac{T}{2}log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}(log|V_t| + u_t' V_t^{-1} u_t) \tag{37}$$

Where we substituted $N(E(y_t|y_{1,t-1}), Var(y_t|y_{1,t-1})) = N(\mu_t, V_t)$ for $p(y_t|y_{1,t-1})$ in the previous loglikelihood function. Further, $u_t = y_t - \mu_t$ and T is the amount of observations in the sample set. The expectation and the variance can be found by using the prediction step of the Kalman Filter ($y_{T+1}|y_{1,T}$), and then we can calculate the likelihood from the Kalman filter outputs. For this one runs the filter at some initial values ($\boldsymbol{\theta}^{(0)}$), which gives an estimate of $\hat{\mu}_{t|t-1}(\boldsymbol{\theta}^{(0)})$. Next, this estimate can be used to find a new estimate $\boldsymbol{\theta}^{(1)}$ which has a bigger value for the predictive log-likelihood in (37). These steps can be repeated to maximize the log-likelihood function.

A.10 Calculation of Full Conditional Distributions for State Space Variances

Using the convenient inverse-gamma family of priors, Petris et al. (2009) identify the full conditional distributions of the variances. First they let $\psi_1 = V^{-1}$ and $\psi_2 = W^{-1}$ (V is the variance of the observation equation and W of the state equation), and assume that ψ_1 and ψ_2 are a priori independent, with

$$\psi_i \sim G(a_i, b_i), \qquad i = 1, 2. \tag{38}$$

Then based on the fact that the full conditional distribution is proportional to the joint distribution of all the random variables considered, the standard approach they use for ψ_1 is

$$\begin{aligned}
\pi(\psi_1|\psi_2, \theta_{0:T}, y_{1:T}) &\propto \pi(\psi_1, \psi_2, \theta_{0:T}, y_{1:T}) \\
&\propto \pi(y_{1:T}|\theta_{0:T}, \psi_1, \psi_2) \propto \pi(\theta_{0:T}|\psi_1, \psi_2)\pi(\psi_1, \psi_2) \\
&\propto \prod_{t=1}^{T} \pi(y_t|\theta_t, \psi_1)\pi(\theta_t|\theta_{t-1}, \psi_2)\pi(\psi_1)\pi(\psi_2) \\
&\propto \pi(\psi_1)\psi_1^{T/2}\exp\left(-\frac{\psi_1}{2}\sum_{t=1}^{T}(y_t - \theta_t)^2\right) \\
&\propto \psi_1^{a_1+T/2-1}\exp\left(-\psi_1\left[b_1 + \frac{1}{2}\sum_{t=1}^{T}(y_t - \theta_t)^2\right]\right).
\end{aligned} \tag{39}$$

Where from the previous equations it can be deduced that ψ_1 and ψ_2 are conditionally independent, given $\theta_{0:T}$ and $y_{1:T}$, and

$$\psi_1|\theta_{0:T}, y_{1:T} \sim G\left(a_1 + \frac{T}{2}, b_1 + \frac{1}{2}\sum_{t=1}^{T}(y_t - \theta_t)^2\right). \tag{40}$$

For ψ_2 the same argument shows that we then have for the state equation

$$\psi_2|\theta_{0:T}, y_{1:T} \sim G\left(a_2 + \frac{T}{2}, b_2 + \frac{1}{2}\sum_{t=1}^{T}(\theta_t - \theta_{t-1})^2\right). \tag{41}$$

A.11 Recursion & Lemmas from Gerlach et al. (2000)

The Recursion Lemmas

Lemma 1. Let $r_{t+1} = \mathrm{var}(y_{t+1}|\boldsymbol{x}_t, \boldsymbol{K}^{1,t+1})$. Then the following holds:

a $E(y_{t+1}|\boldsymbol{x}_t, \boldsymbol{K}^{1,t+1})$

$$= g_{t+1} + \boldsymbol{h}'_{t+1}(\boldsymbol{f}_{t+1} + \boldsymbol{F}_{t+1}\boldsymbol{x}_t) \tag{42}$$

$$r_{t+1} = (\boldsymbol{h}'_{t+1}\boldsymbol{\Gamma}_{t+1} + \boldsymbol{G}_{t+1})(\boldsymbol{h}'_{t+1}\boldsymbol{\Gamma}_{t+1} + \boldsymbol{G}_{t+1})', \tag{43}$$

$$E(x_{t+1}|\boldsymbol{x}_t, y_{t+1}, \boldsymbol{K}) = \boldsymbol{a}_{t+1} + \boldsymbol{A}_{t+1}\boldsymbol{x}_t + \boldsymbol{B}_{t+1}y_{t+1},$$

and

$$\mathrm{var}(x_{t+1}|\boldsymbol{x}_t, y_{t+1}, \boldsymbol{K}) = \boldsymbol{C}_{t+1}\boldsymbol{C}'_{t+1},$$

where

$$\begin{aligned}
\boldsymbol{a}_{t+1} &= (\boldsymbol{I} - \boldsymbol{B}_{t+1}\boldsymbol{h}'_{t+1})\boldsymbol{f}_{t+1} - \boldsymbol{B}_{t+1}g_{t+1}, \\
\boldsymbol{A}_{t+1} &= (\boldsymbol{I} - \boldsymbol{B}_{t+1}\boldsymbol{h}'_{t+1})\boldsymbol{F}_{t+1}, \\
\boldsymbol{B}_{t+1} &= \boldsymbol{\Gamma}_{t+1}(\boldsymbol{h}'_{t+1}\boldsymbol{\Gamma}_{t+1} + \boldsymbol{G}'_{t+1})/r_{t+1}
\end{aligned}$$

and

$$\begin{aligned}
&\boldsymbol{C}_{t+1}\boldsymbol{C}'_{t+1} \\
&\quad = \boldsymbol{\Gamma}_{t+1}\left\{\boldsymbol{I} - \frac{1}{r_{t+1}}(\boldsymbol{\Gamma}'_{t+1}\boldsymbol{h}_{t+1} + \boldsymbol{G}'_{t+1})(\boldsymbol{\Gamma}'_{t+1}\boldsymbol{h}_{t+1} + \boldsymbol{G}'_{t+1})'\right\}\boldsymbol{\Gamma}'_{t+1}. \qquad (44)
\end{aligned}$$

The right side of (44) can be factored to get a matrix of $\boldsymbol{C}_{t+1}$ that is either null or has full rank.

b One can write

$$\boldsymbol{x}_{t+1} = \boldsymbol{a}_{t+1} + \boldsymbol{A}_{t+1}\boldsymbol{x}_t + \boldsymbol{B}_{t+1}y_{t+1} + \boldsymbol{C}_{t+1}\boldsymbol{z}_{t+1}$$

where $z_{t+1} \sim N(0, I)$ and is independent of $\boldsymbol{x}_t$ and y_{t+1} (conditional on $\boldsymbol{K}$).

Lemma 2. For $t = 1, ..., n-1$ the conditional density $p(\boldsymbol{y}^{t+1,n}|\boldsymbol{x}_t, \boldsymbol{K})$ is independent of $\boldsymbol{K}^{1,t}$ and can be expressed as

$$p(\boldsymbol{y}^{t+1,n}|\boldsymbol{x}_t, \boldsymbol{K}) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{x}'_t\boldsymbol{\Omega}_t\boldsymbol{x}_t - 2\boldsymbol{\mu}'_t\boldsymbol{x}_t)\right\} \qquad (45)$$

The terms $\boldsymbol{\Omega}_t$ and $\boldsymbol{\mu}'_t$ are computed recursively as follows:

- $\boldsymbol{\Omega}_n = 0$ and $\boldsymbol{\mu}'_n = 0$.
- For $t = n-1, ..., 1$,

$$
\begin{aligned}
\boldsymbol{\Omega}_t =& \boldsymbol{A}'_{t+1}(\boldsymbol{\Omega}_{t+1} - \boldsymbol{\Omega}_{t+1}\boldsymbol{C}_{t+1}\boldsymbol{D}_{t+1}^{-1}\boldsymbol{C}'_{t+1}\boldsymbol{\Omega}_{t+1})\boldsymbol{A}_{t+1} \\
&+ \boldsymbol{F}'_{t+1}\boldsymbol{h}_{t+1}\boldsymbol{h}'_{t+1}\boldsymbol{F}_{t+1}/r_{t+1},
\end{aligned} \tag{46}
$$

$$
\begin{aligned}
\boldsymbol{\mu}_t =& \boldsymbol{A}'_{t+1}(\boldsymbol{I} - \boldsymbol{\Omega}_{t+1}\boldsymbol{C}_{t+1}\boldsymbol{D}_{t+1}^{-1}\boldsymbol{C}'_{t+1}) \\
&\times (\boldsymbol{\mu}_{t+1} - \boldsymbol{\Omega}_{t+1}(\boldsymbol{a}_{t+1} + \boldsymbol{B}_{t+1}y_{t+1})) \\
&+ \boldsymbol{F}'_{t+1}\boldsymbol{h}_{t+1}(y_{t+1} - g_{t+1} - \boldsymbol{h}'_{t+1}\boldsymbol{f}_{t+1})/r_{t+1},
\end{aligned} \tag{47}
$$

and

$$
\boldsymbol{D}_{t+1} = \boldsymbol{C}'_{t+1}\boldsymbol{\Omega}_{t+1}\boldsymbol{C}_{t+1} + \boldsymbol{I} \tag{48}
$$

Matrix $\boldsymbol{D}_{t+1}$ may be low dimensional because $\boldsymbol{C}_{t+1}$ has low rank. If $\boldsymbol{C}_{t+1}$ is null, $\boldsymbol{D}_{t+1} = \boldsymbol{I}$, and no matrix inversions are necessary in 47. Which may happen if for given $\boldsymbol{K}$, $\boldsymbol{x}_{t+1}$ is a linear combination of $\boldsymbol{D}_t$ and y_{t+1}.

Let $\boldsymbol{m}_t = E(\boldsymbol{x}_t|\boldsymbol{y}^{1,t}, \boldsymbol{K})$, $\boldsymbol{V}_t = \mathrm{var}(\boldsymbol{x}_t|\boldsymbol{y}^{1,t}, \boldsymbol{K})$, and $R_t = \mathrm{var}(\boldsymbol{y}_t|\boldsymbol{y}^{1,t-1}, \boldsymbol{K})$. The next lemma is the Kalman filter for the state space model 2.

Lemma3.

a

$$
R_t = \boldsymbol{h}'_t\boldsymbol{F}_t\boldsymbol{V}_{t-1}\boldsymbol{h}'_t\boldsymbol{F}_t + (\boldsymbol{h}'_t\boldsymbol{\Gamma}_t + \boldsymbol{G}_t)(\boldsymbol{\Gamma}'_t\boldsymbol{h}_t + \boldsymbol{G}'_t),
$$

$$
\boldsymbol{m}_t = (\boldsymbol{I} - \boldsymbol{J}_t\boldsymbol{h}'_t)(\boldsymbol{f}_t + \boldsymbol{F}_t\boldsymbol{m}_{t-1}) + \boldsymbol{J}_t(y_t - g_t),
$$

and

$$
\boldsymbol{V}_t = \boldsymbol{F}_t\boldsymbol{V}_{t-1}\boldsymbol{F}'_t + \boldsymbol{\Gamma}_t\boldsymbol{\Gamma}'_t - \boldsymbol{J}_t\boldsymbol{J}'_tR_t, \tag{49}
$$

where

$$
\boldsymbol{J}_t = [\boldsymbol{F}_t\boldsymbol{V}_{t-1}\boldsymbol{F}'_t + \boldsymbol{\Gamma}_t(\boldsymbol{\Gamma}'_t\boldsymbol{h}_t + \boldsymbol{G}'_t)]/R_t.
$$

b The conditional density

$$p(y_t|\boldsymbol{y}^{1,t-1}, \boldsymbol{K}^{1,t}) \propto R_t^{-1/2}\exp\left\{-\frac{1}{2R_t}(y_t - g_t - \boldsymbol{h}_t'(\boldsymbol{f}_t + \boldsymbol{F}_t\boldsymbol{m}_{t-1}))^2\right\}.$$

c We can write $\boldsymbol{V}_t = \boldsymbol{T}_t\boldsymbol{T}_t'$, where the matrix $\boldsymbol{T}_t$ either has full column rank if $\boldsymbol{V}_t \neq 0$ or is null if $\boldsymbol{V}_t = 0$. Conditional on $\boldsymbol{K}$ we express $\boldsymbol{x}_t$ as

$$\boldsymbol{x}_t = \boldsymbol{m}_t + \boldsymbol{T}_t\boldsymbol{\xi}_t,$$

where $\boldsymbol{\xi}_t$ is $N((0), \boldsymbol{I})$ and independent of $\boldsymbol{y}^{1,t}$.

The next lemma uses Lemma 3 to efficiently evaluate the factor $p(\boldsymbol{y}^{t+1,n}|\boldsymbol{y}^{1,t}, \boldsymbol{K})$ in (22).

Lemma 4.

$$\begin{aligned} p(\boldsymbol{y}^{t+1,n}|\boldsymbol{y}^{1,t}, \boldsymbol{K}) &= \int p(\boldsymbol{y}^{t+1,n}|\boldsymbol{x}_t, \boldsymbol{K}^{t+1,n})p(\boldsymbol{\xi}_t|\boldsymbol{K}^{1,t})d\boldsymbol{\xi}_t \\ &\propto |\boldsymbol{T}_t'\boldsymbol{\Omega}_t\boldsymbol{T}_t + \boldsymbol{I}|^{-1/2}\exp\left\{-\frac{1}{2}[\boldsymbol{m}_t'\boldsymbol{\Omega}_t\boldsymbol{m}_t - 2\boldsymbol{\mu}_t'\boldsymbol{m}_t\right. \\ &\left. -(\boldsymbol{\mu}_t - \boldsymbol{\Omega}_t\boldsymbol{m}_t)'\boldsymbol{T}_t(\boldsymbol{T}_t\boldsymbol{\Omega}_t\boldsymbol{T}_t + \boldsymbol{I})^{-1}\boldsymbol{T}_t'(\boldsymbol{\mu}_t - \boldsymbol{\Omega}_t\boldsymbol{m}_t)]\right\} \end{aligned} \tag{50}$$

The recursion for generating $\boldsymbol{K}$ in $O(n)$ operations is given next.

Recursion 1: Sampling $\boldsymbol{K}$ *From Reduced Conditionals*

Step 1. Given the current value for $\boldsymbol{K}$, calculate $\boldsymbol{\Omega}_t$ and $\boldsymbol{\mu}_t$ for $t = n-1, ..., 1$, using the recursions in Lemma 2.

Step 2. Given $E(\boldsymbol{x}_0)$, $\text{var}(\boldsymbol{x}_0)$, perform the following for $t = 1, ..., n$:

a For each value of $\boldsymbol{K}_t$:

- Obtain $\boldsymbol{R}_t$, $\boldsymbol{m}_t$ and $\boldsymbol{V}_{t-1}$ as in Lemma 3.
- Obtain $p(y_t|\boldsymbol{y}^{1,t-1}, \boldsymbol{K}^{1,t})$ as in Lemma 3, part b and obtain $p(\boldsymbol{y}^{t+1,n}|\boldsymbol{y}^{1,t}, \boldsymbol{K})$ as in Lemma 4.

b Obtain $p(\boldsymbol{K}_t|\boldsymbol{y}, \boldsymbol{K}_{s\neq t})$ for all values of $\boldsymbol{K}_t$ by normalization. Then generate $\boldsymbol{K}_t$.

c Update $\boldsymbol{m}_t$ and $\boldsymbol{V}_t$ as in Lemma 3, using generated values of $\boldsymbol{K}_t$.

A.12 Posterior Distribution Histograms

The conditional distributions found after convergence of the parameters are used to sample the parameters from, based on the full data of operating window 2. The sampled values can be used to identify the distributions. Impressions of these distributions are given here in form of their histograms.

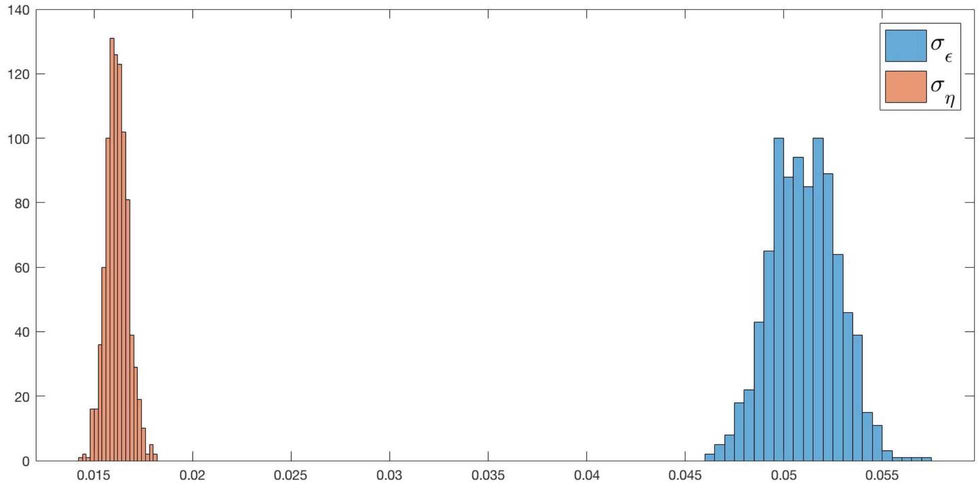

Figure 32: Posterior density histograms of σ_ε and σ_η in the local level model

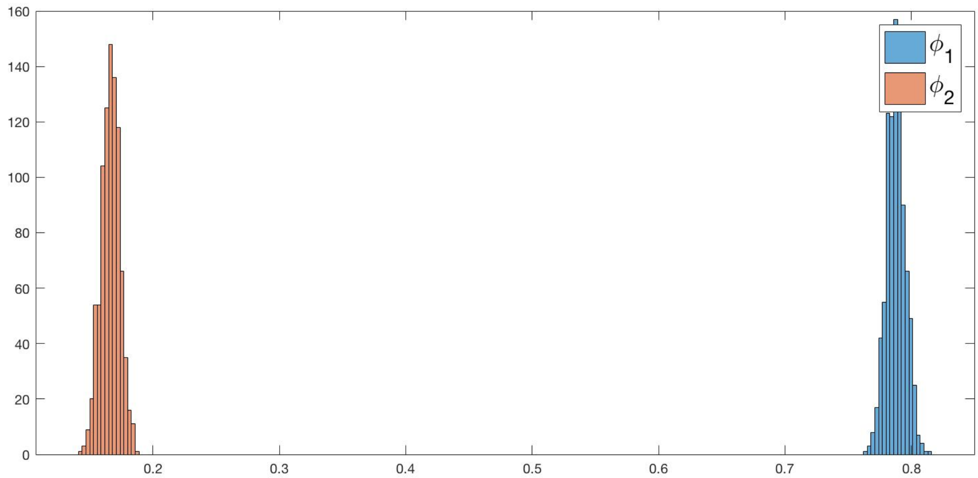

Figure 33: Posterior density histograms of ϕ_1 and ϕ_2 in the local level model

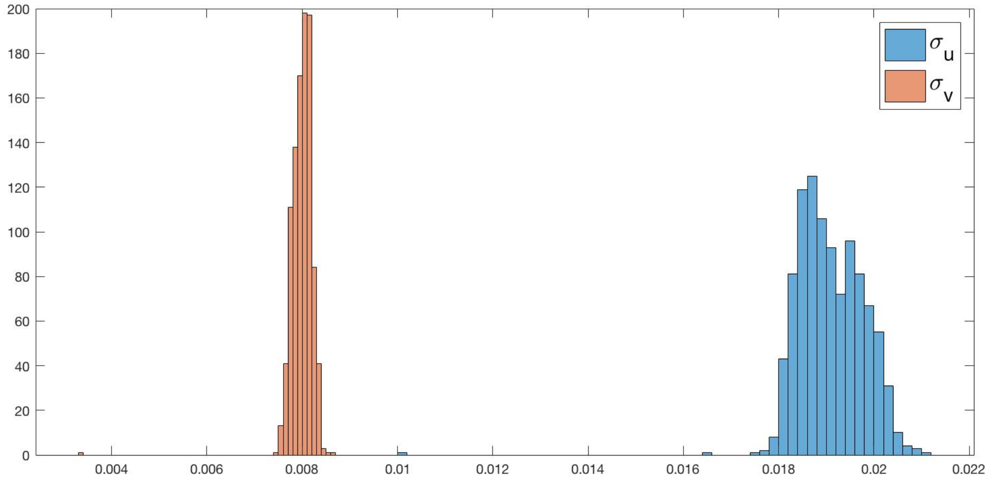

Figure 34: Posterior density histograms of σ_u and σ_v in the dynamic mixture model

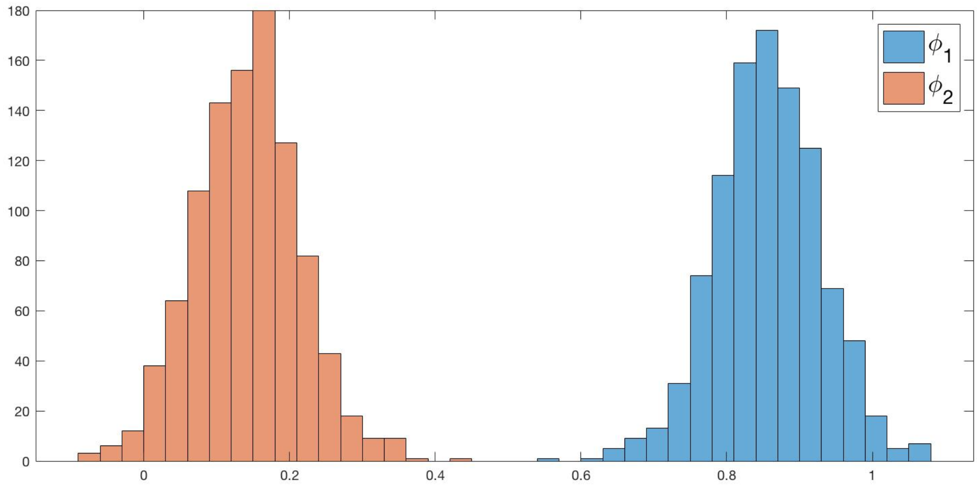

Figure 35: Posterior density histograms of ϕ_1 and ϕ_2 in the dynamic mixture model