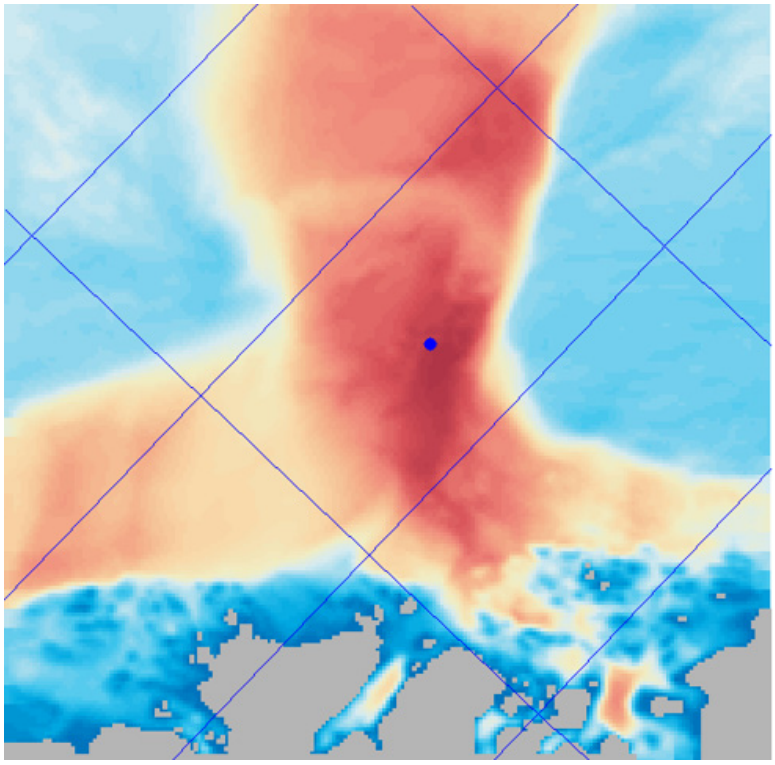
SINTEF A26953- Unrestricted

Report

Data assimilation with SINMOD

Author(s)
Morten Omholt Alver
Finn Are Michelsen



SINTEF Fisheries and Aquaculture
SINTEF Fisheries and Aquaculture
2015-05-20

SINTEF Fiskeri og havbruk AS
SINTEF Fisheries and Aquaculture

Address:
Postboks 4762 Sluppen
NO-7465 Trondheim
NORWAY

Telephone:+47 40005350
Telefax:+47 93270701

fish@sintef.no
www.sintef.no/fisk
Enterprise /VAT No:
NO 980 478 270 MVA

Report

Data assimilation with SINMOD

KEYWORDS:
Keywords

VERSION
1.0

DATE
2015-05-20

AUTHOR(S)
Morten Omholt Alver
Finn Are Michelsen

CLIENT(S)
Norges Forskningsråd

CLIENT'S REF.

PROJECT NO.
102004481

NUMBER OF PAGES/APPENDICES:
33 + Appendices

ABSTRACT

Data assimilation is the process of using measured data to correct the state of a mathematical model. This report documents the data assimilation process implemented in the ocean model system SINMOD, and the tests that have been run in the ELMO project. The report describes the choice of the EnOI assimilation method, its implementation in SINMOD, the input layer for providing measurement data to the model, and two test cases where the model has been run with assimilation of current data.

PREPARED BY
Morten Omholt Alver
SIGNATURE

CHECKED BY
Ingrid Helene Ellingsen
SIGNATURE

APPROVED BY
Gunvor Øie
SIGNATURE

REPORT NO.	ISBN	CLASSIFICATION	CLASSIFICATION THIS PAGE
SINTEF A26953	978-82-14-05881-9	Unrestricted	Unrestricted

Document history

VERSION	DATE	VERSION DESCRIPTION
Version No.	Date	"[Version description.Use TAB for new line]"

Table of contents

1 SINMOD model setup

In the ELMO project, sea water current predictions are obtained by running the coupled physical-biological ocean model SINMOD [*Slagstad and McClimans*, 2005]. The hydrodynamic model is a z-level model based on the primitive Navier-Stokes equations. The vertical layers are set up with different thicknesses, typically giving higher resolution in the upper parts of the water column. The horizontal resolution is uniform over the model area, and is a trade-off between computational requirements and the ability to resolve dynamics and geographical features. A nesting technique allows high resolution model setups to use boundary values produced by lower resolution setups, thereby absorbing large scale properties generated outside the local model domain.

For the tests in the ELMO project two different model configurations have been used. The first setup, *LoVe*, was used to extensively test and tune the data assimilation routine utilizing current measurements from the Lofoten and Vesterålen area in a hind cast setting. The second setup, *FroN*, was used for the full ELMO system test. Due to the lack of real current measurements in the system test, data assimilation was performed only using data from "virtual buoys" obtained from the Norkyst800 model. The choice of data assimilation algorithm, and the two model setups, are described in the following sections.

1.1 Data assimilation algorithm

The applications of supervising systems for ocean, near shore water and biomass state predictions are numerous. Examples include fish farming supervising, oil spill tracking and assisting dumping operations for masses in sea water. Ocean models are widely used in such applications for estimation and prediction of current, temperature, salinity etc. The reliability of these estimates relies on the inherent uncertainty in the model calculations due to unmodeled phenomena (or phenomena related to the physics are not modelled in sufficient detail), uncertainties in model parameters, driving forces, initial values and boundary conditions. Hence, model correction using measured data is necessary in order to achieve reliable results. This is known as state estimation or data assimilation (DA).

There are basically two classes of DA methods; (i) variational and (ii) sequential methods. As argued by [*Dee*, 2005], the distinction between the two classes are, however, somewhat artificial since all assimilation methods are fundamentally sequential in the sense that analyses are produced sequentially in time. Each analysis in the sequence is an approximate solution of a variational problem, in which observations are optimally combined with a model-generated state estimate. The variational methods back-calculate initial values, boundary values and driving forces such that assimilated variables, e.g. current, temperature and salinity, match measured data. Hence, systems based on these methods are known as inverse modelling systems. The calculations involve optimal search for initial conditions, boundary values and driving forces by minimizing a given cost function that measures the model to data misfit, see e.g. [*Erwig*, 2008] and [*Muccino et al.*, 2008]. One advantage with these methods is that the solutions are consistent with the balance equations in the ocean model. This is important since accurate initial conditions, boundary values and driving forces are crucial to the model solution partly due to the chaotic nature of the oceans [*Oey et al.*, 2005]. Further, systematic model and measurement errors are implicitly handled. A disadvantage is that they involve complex algorithms that result in a comprehensive implementation. The 4-dimensional variant, 4DVAR (i.e. time and space are free variables [*Rawlins et al.*, 2007]), is implemented in the widely applied Regional Ocean Modeling System (ROMS), see [*Moore et al.*, 2011].

In the ELMO project, focus is set on a simple sequential DA method. These types of methods are based on filter theory and give estimates of the system state sequentially forward in time by adding a statistically based correction term to the solution of the balance equations. In this way, the model estimate is a combination of a physically based and a statistically based solution. The advantage with these methods is

that the assimilation scheme and its implementation can be made simple. Further, they can be used when the errors are unknown, but unbiased. One disadvantage is that systematic errors (e.g. biases) have to be handled by expanding the correction scheme based on knowledge about the errors that is not always available, see e.g. [*Chepurin et al.*, 2005]. Another disadvantage is that the corrected estimate is not consistent with the conservation laws, which the ocean model is based on.

One popular class of methods is based on the Kalman Filter (KF) theory, originally proposed by Rudolf Kalman [*Kalman*, 1960]. For oceanographic models, the assimilation problem is complicated by high-dimensional and strongly nonlinear systems. This makes the standard Kalman Filter approach unpractical due to large computational costs. A practical approach for data assimilation of ocean models is the use of ensembles of data. One popular approach is the ensemble Kalman filter (EnKF) that was developed by [*Evensen*, 2003] who refers several publications that have proved the feasibility of ensemble based methods for real oceanographic problems. The EnKF is a Monte Carlo method where a number of model instances are run in parallel, and the variation between the instances of model state variables are represented by probability distributions. Other KF methods are the unscented KF that includes a minimal set of sample points (called sigma points) around the mean [*Julier and Uhlmann*, 1997] and Optimal Interpolation (KF with local analysis).

For real time data assimilation where the computational resources are limited, the ensemble optimal interpolation (EnOI) method, suggested by [*Evensen*, 2003], appears to be a practical and cost-effective approach ([*Oke et al.*, 2007]). This method uses a stationary ensemble of state estimates collected in advance, giving a constant Kalman filter gain. For small state ensembles, this approach typically does not span the full sub-space of the model. As a consequence, spurious correlations may occur in the model. In order to cope with this, a technique called localisation is applied. By localisation, covariances between distant elements of the model state vector are filtered. [*Oke et al.*, 2007] found that while EnOI is less optimal than the EnKF, for some cases the performance of EnOI may be comparable, or even superior, to that of the EnKF. EnOI for ocean models is earlier applied by [*Counillon and Bertino*, 2009] for assimilating altimetry data and by [*Xu et al.*, 2013] for assimilating current data, both from the Gulf of Mexico. Further, [*Fu et al.*, 2011] used EnOI to assimilate temperature and salinity profiles in the Baltic sea. Earlier results from the assimilation system presented in this article are published by [*Grythe et al.*, 2013].

The tests described in this report use an implementation of the EnOI method implemented for SINMOD. The following sections describe the relevant equations.

1.1.1 Basic filter equations

The minimum error-variance estimate of the analysed state, also known as a posteriori state estimate, is given by:

$$\boldsymbol{x^a = x^f + K\left(y - x^f\right)} \qquad (1)$$

where $\boldsymbol{x^f}$ is the forecast state estimate. The filter gain, also known as the Kalman gain matrix, weights the state covariances with the measurements noise, and is given by:

$$\boldsymbol{K = P^f H^T (H P^f H^T + R)^{-1}} \qquad (2)$$

P^f is the forecast-error covariance matrix, also known as background error covariance matrix and a priori covariance matrix. The measurement matrix H relates the measurements to the model states. It is essential that the measurements y are treated as random variables having a distribution with covariance matrix R. Further, the measurements are not restricted to Gaussian distributed noise.

K determines the relative weights of the forecast and the observations. The main challenge for a successful application of Eqs. (1) and (2) is the estimation of P^f.

1.1.2 The ensemble Kalman filter (EnKF)

The EnKF approximates P^f in Eq. (2) as

$$\boldsymbol{P^f = A' A'^T / (N-1)} \quad (3)$$

where the ensemble anomalies are given as:

$$\boldsymbol{A' = A^f - \overline{A^f}} \quad (4)$$

and

$$\boldsymbol{A^f = \left[x_1^f, \quad x_2^f, \ldots\ldots, x_N^f\right]} \quad (5)$$

A^f is the ensemble of N model forecasts provided from a Monte Carlo simulation, and $\overline{A^f}$ is the ensemble mean. Ideally, the model forecasts should be independent samples from the a priori distribution. However, they are not in general independent due to the model dynamics.

In periods of strong variability, where the model produces large errors, the assumed model forecast-error P^f is also large. This gives a rational for this approach [*Oke et al.*, 2007].

When the ensemble members in an EnKF are appropriately initialised and evolved in time, the EnKF arguably offers a nearly optimal method for assimilation. However, the EnKF is expensive, as a large N and calculation of a new filter at each assimilation time step are required.

1.1.3 Ensemble optimal interpolation (EnOI)

EnOI is an approximation to the EnKF where the ensemble A^f is stationary and multiplied by a constant α that scales the magnitude of the assumed forecast errors. As an example, an ensemble consisting of model states sampled over a long time period will have a variance which is too large to represent the actual error in the model forecast. In this case, α is used to reduce the variance to a realistic level. The state ensemble should be recorded from a long-term stationary period of time. Hence, this sub-optimal method doesn't require each ensemble member to be updated with time. This makes a stationary Kalman filter.

1.1.4 Localization

One important issue in assimilation is the range of significance of a measurement. Further, for many applications within ocean simulation, the computational resources limit the number of ensemble members to a modest size. Then, some correlations will often be spurious. This is because the ensemble does not span the same space as the true forecast errors. As a result, the a posteriori state estimates include large errors, and the assimilation of the measurements fails to give good match. Filtering can be applied to reduce these correlations. This is known as localization. By this, measurements will only affect states in the neighbourhood, i.e. located within a certain distance from the measurement location. Another result is that localisation increases the rank of the system [*Oke et al.*, 2007], which is computationally advantageous.

Localisation is implemented by the use of an element wise multiplication of the covariance matrix and a correlation function with local support [*Houtekamer and Mitchell*, 2001]. This can be any smooth function

that gives a weight depending on the horizontal distance to the measurement position. The following function $\beta \in (0,1]$ is used in this study:

$$\boldsymbol{\beta = 1, r < r_0} \quad (6)$$

$$\boldsymbol{\beta = e^{-(r-r_0)^2/w^2}, r \geq r_0} \quad (7)$$

where r_0 and w define the shape of the localization function.

3.5 Model correction

The integration of the ocean model is made by:

$$\boldsymbol{x^f_{k+1} = F(x^a_k)} \quad (8)$$

$$\boldsymbol{x^a_{k+1} = x^f_{k+1} + K\left(y_{k+1} - Hx^f_{k+1,}\right)} \quad (9)$$

where k is the discrete time index, F is the dynamic ocean model and H is the measurement matrix.

The ocean current is modelled in two horizontally perpendicular directions U and V. These consist of two parts; a barotropic part A and a baroclinic part B:

$U = U_A + U_B$, $V = V_A + V_B$

By this, the state vector is defined by:

$$x = [U_A, U_B, V_A, V_B] \quad (10)$$

where each current part covers all grid cells in the model setup.

The barotropic parts are defined by depth integration of U and V:

$$U_A = \frac{1}{h}\int_0^h U(z)dz, \quad V_A = \frac{1}{h}\int_0^h V(z)dz \quad (11)$$

Since these A parts of the model state represent much faster modi than the B parts and the rest of the ocean model, the time step for integration of the A parts is chosen much smaller than the time step for integration of the rest of the model. A factor of 24 is used in the results as presented here.

1.2 LoVe model setup

A model with 160 m horizontal resolution and a grid size of 200x200, hereafter called the LoVe model, is nested from a 800m model. The depth grids of these models are shown in Figure 1 and Figure 2. The 800m model is nested from at 4km model, which in turn is nested from a 20km model, which is the main SINMOD model for the Norwegian and Arctic seas.

The vertical layers are set up with different thickness, giving higher resolution in the upper and lower parts of the water column. Table 1 shows the grid sizes in various depth ranges.

Table 1: Grid sizes of vertical layers

Depth range from the ocean surface [m]	Grid size
1-25	3, 1, 1, 1, 1,1, 1, 1, 2, 3, 5, 5
25-150	5, 10, 10, 25, 25, 50
150-200	50
200-225	25
225-235	10
235-	5

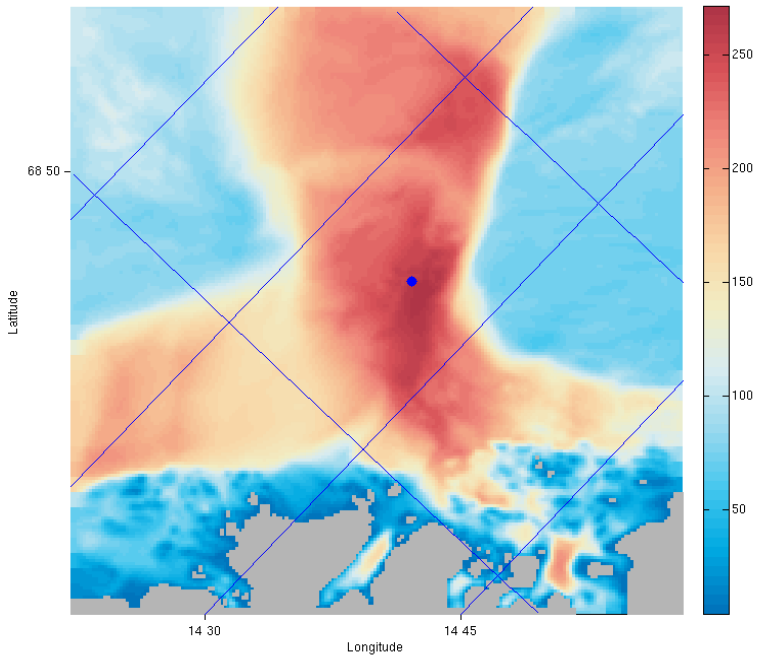


Figure 1: Depth grid of the 160 m LoVe model. The blue dot shows the measurement buoy position.

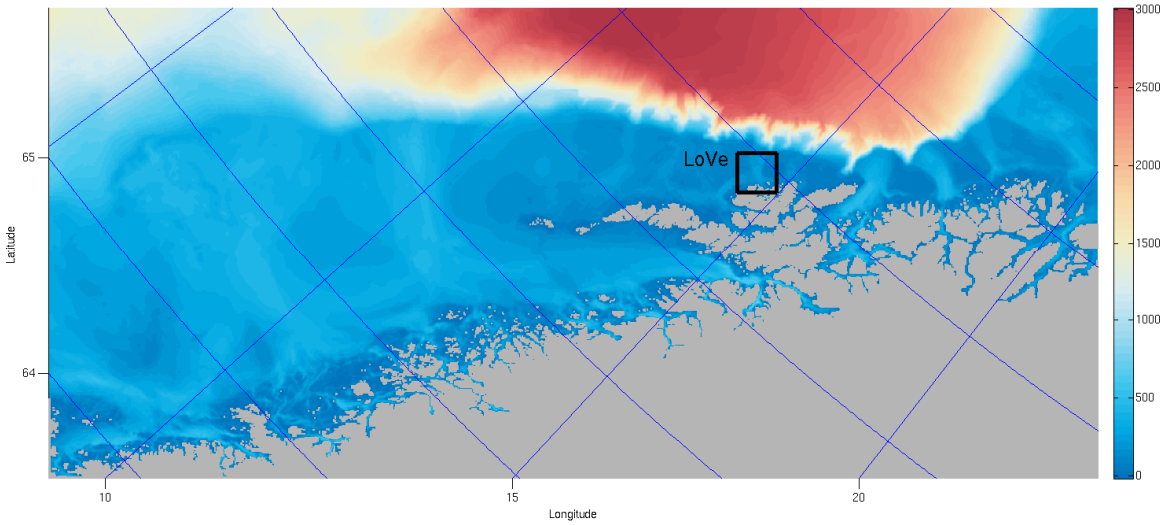


Figure 2: Depth grid of the 800 m model, covering the coastal areas of northern Nordland and south part of Troms. The black square shows the extent of the 160m model area for LoVe.

1.3 Frohavet Nord model setup

A model with 160 m horizontal resolution and a grid size of 510x430, hereafter called the FroN model, is nested from the Norkyst800 operational model system[1] developed and operated by the Institute of Marine Research and the Meteorological Institute. The depth grid of the FroN model is shown in Figure 3.

The vertical layers are set up with different thicknesses, giving higher resolution in the upper part of the water column. Table 2 shows the grid sizes in various depth ranges.

Table 2: Grid sizes of vertical layers

Depth range from the ocean surface [m]	Grid size
1-25	3, 1, 1, 1, 1, 1, 1, 1, 2, 3, 5, 5
25-200	5, 5, 5, 10, 25, 25, 50, 50
200-900	50
900-1000	100
1000-	500

[1] http://www.imr.no/temasider/modeller/kystmodellen/kystmodellen_norkyst800/en

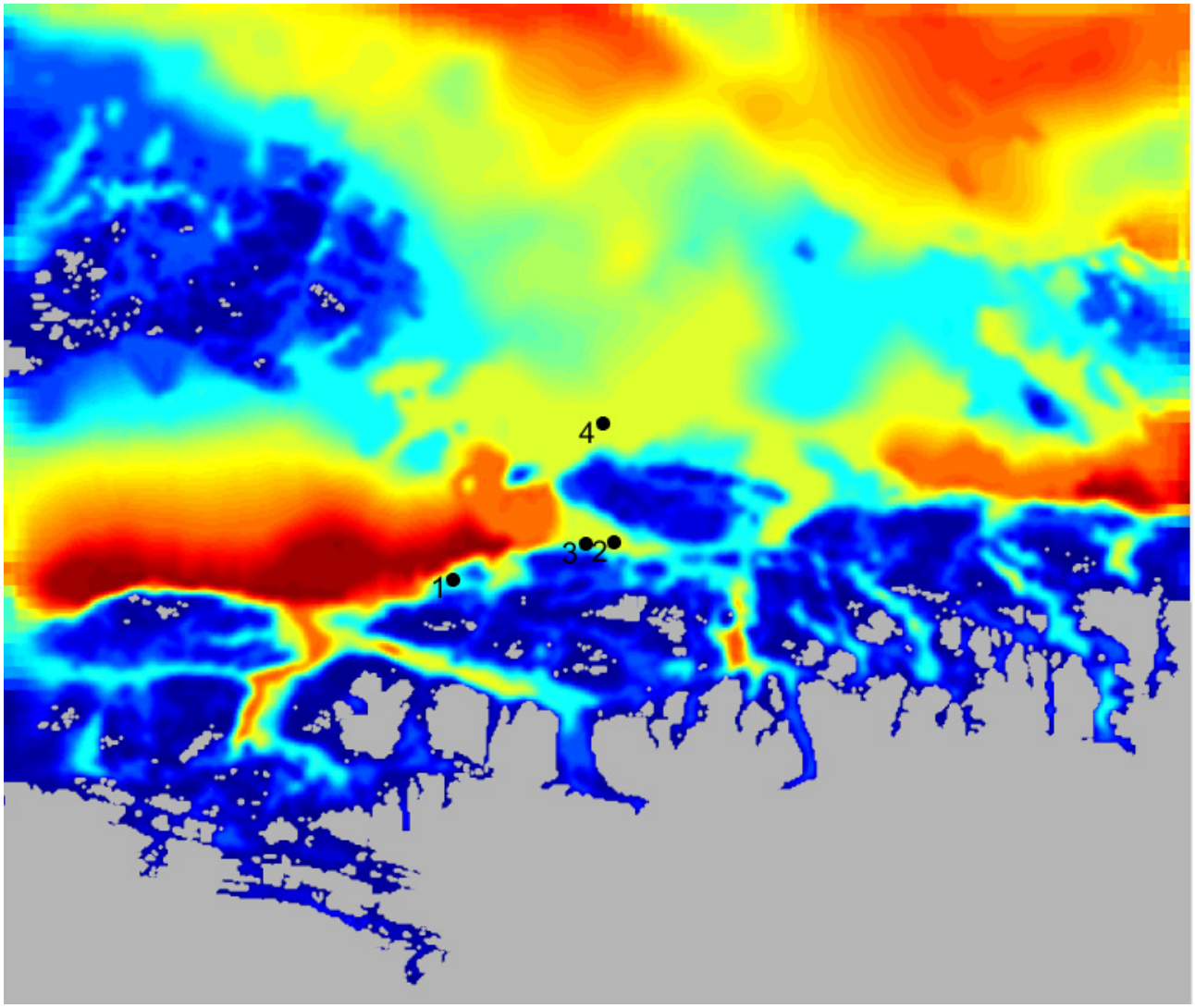


Figure 3: Depth grid of the 160 m Frohavet Nord model, covering parts of the coastal areas of Middle Norway. The numbered black circles show the locations of the four virtual measurement buoys.

Atmospheric input data are downloaded from the operational Hirlam12 system. Both the Norkyst800 and Hirlam12 data are distributed through Meteorological Institute's data server[2].

2 Measurement data

2.1 LoVe measurement data

The measurement buoy includes two ADCPs, one Continental and one Aquadop from Nortek, logging current at the same position, 68° 54.596' N, 14° 23.384' E, at about 260 m depth, as shown in Figure 1. The sensor depth is estimated from the mean pressure measurement. The Aquadop sensor has 2 m depth resolution and covers a sub range 220-255 m of the depth range for the Continental sensor, 115-255 m with 5 m resolution. Hence, only the Continental sensor is used for data assimilation.

[2] http://thredds.met.no

An example of the first lines of a data file is shown below:

```
11  01  2013  10  41  17      0    48   11.6 1485.0    2.6    2.6   -1.9 259.359    8.69
0      0 3   20
  1     0.229     0.036     0.028 125 126 126
  2     0.279     0.116    -0.009 107 109 108
  3     0.359     0.134     0.000  96  99  96
```

The first line is header, where the six first columns are the **date**, followed by **Error code**, **Status code**, **Battery voltage** (V), **Sound speed** (m/s), **Heading** (degrees), **Pitch** (degrees), **Roll** (degrees), **Pressure** (dbar)**, Temperature** (°C), **Analog input 1**, **Analog input 2**, **Number of beams**, **Number of cells**.

The next lines include: **Cell number**, **Velocity** (Beam1|X|East) (m/s), **Velocity** (Beam2|Y|North) (m/s), **Velocity** (Beam3|Z|Up) (m/s), **Amplitude** (Beam1) (counts), **Amplitude** (Beam2) (counts), **Amplitude** (Beam3) (counts)

.

Only those columns show with bold face text above are read from the data files.

The horizontal sensor accuracy is 2.2 cm/s independent of the depth, while the measurement signal noise which depends on the measured signal strength, is read from the amplitude/counts column. The lower limit for the Continental sensor is set to 55 counts and 25 counts for the Aquadop sensor. The average signal strength is used to flag measurement samples that are not trusted for data assimilation. Figure 4 shows an example of the strength profile along the depth in a case of good weather conditions. In this case, the signal strength increases towards the ocean surface since the signal reflection is high at the surface.

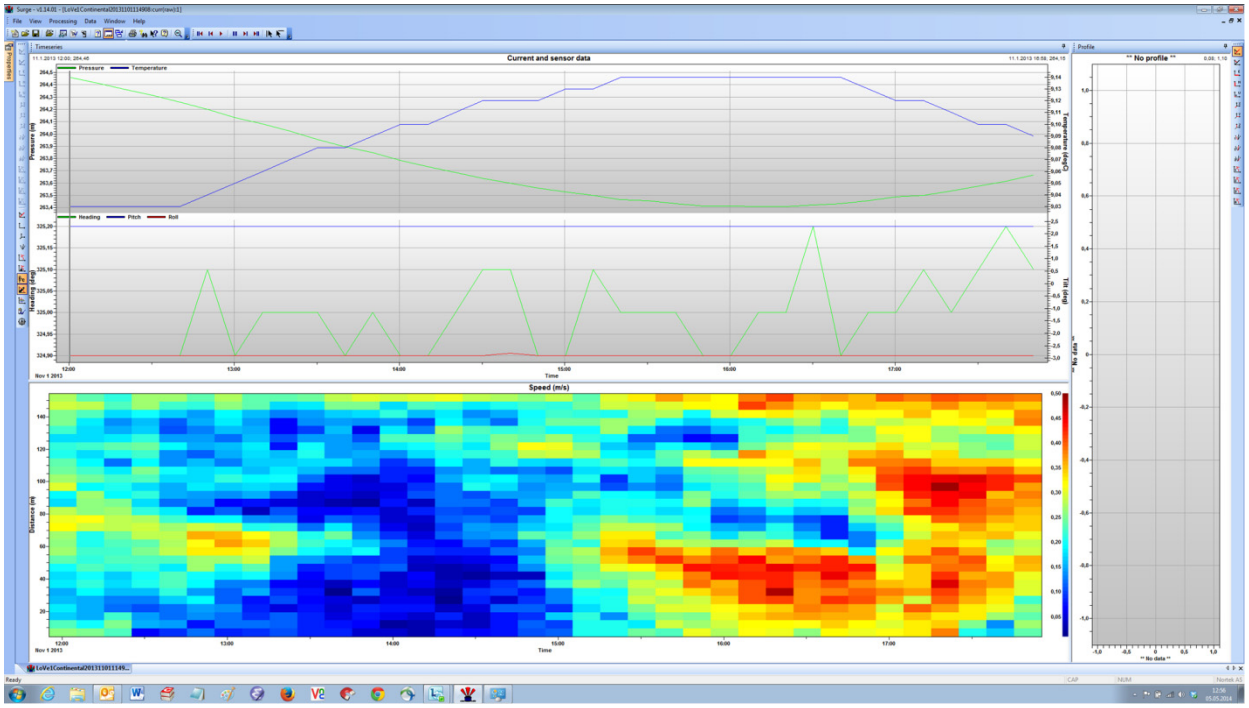


Figure 4: Signal plots from the Continental sensor

2.2 Frohavet Nord measurement data

No measurement data was available for the Frohavet Nord. As the primary aim of this test was of the interaction of the components of the ELMO system, the presence of measurement data was simulated by extracting hourly values for current speed and direction from the forecasts of the Norkyst800 model system. Since this is the same model providing the boundary values for the SINMOD setup, the simulated data take on reasonable values for assimilation into SINMOD.

Four virtual buoy locations were used:

64° 6.92' N 9° 49.6' E
64° 12.4' N 9° 56.5' E
64° 11.5' N 9° 45.0' E
64° 15.0' N 9° 48.1' E

Hourly values were extracted from the Norkyst800 data sets for these four positions and made available at the ELMO THREDDS server.

3 Data layer for input from measured data to SINMOD

Using measured data from a variety of sources as input to correct the current model involves a certain amount of preprocessing to get the data into the correct form, as well as precautions to avoid problems in case of bad or missing data. The preprocessing step needs to be tailored to any new data sources that are introduced.

3.1 Conversion of LoVe measurement data to NetCDF files

Some data files are incomplete; some files are missing, some are empty, and some have missing samples either at the beginning of the file, in the middle or at the end. All the cases are treated individually by the conversion algorithm. The following assumptions are made:

- Missing samples are replaced by the last valid sample if the number of missing samples is three or less. Otherwise, the current speeds u and v are set to 100 to flag uncertain samples that should not be assimilated.
- If the first file in the data set is missing, this is ignored.
- The first sample in the first data file is not missing.
- The data files normally have a fixed number of samples, but some may have more.
- The sampling time is 10 minutes.

The data is filtered with a moving average algorithm. A filter factor, f, defining the filter window is an input parameter. The filter window is given by the integer value of number of samples per file n/f.

Input parameters:

- Buoy position (latitude, longitude)
- Number of samples per file, n
- Filter factor, f
- Sampling time [min]
- Lower limit of signal strength [number of counts]
- Vector of distances from sensor head defining the resolution along the depth axis [m]

Stored variables:

- Buoy position (latitude, longitude)
- Sensor depth [m]
- Buoy station name
- Data responsible
- Sampling time [min]
- Number of depth layers
- Depth vector [m]
- Time axis, [days], ex.: 2000-03-01 15:45:17
- Current speed east [m/s]
- Current speed north [m/s]
- Resulting current speed [m/s]
- Resulting current direction [deg]
- Pressure at sensor head [bar]
- Temperature at sensor head [°C]

3.2 Input parameter files to SINMOD

The data layer between input data and SINMOD consists of two txt files including paths to input files, start and stop times for the simulation, and a number of parameters controlling the data assimilation process. Parameter names from each of the text files are listed below:

simSettings.txt:

savePath	Path to directory where output files are stored
atmoPath	Path to atmospheric data
boundsPath	Path to boundary value files
initFile	Path to file providing initial model values
sample_no	Sample number in initFile for initial values
startTime	Starting time for the simulation, format: 2013,10,04,0,0,0
endTime	End time for the simulation, format: 2013,11,01,0,0,0
dataAssim	Flag for activation of data assimilation (true or false)
saveBounds	Flag for saving of boundary values (true or false)

AssimSettings.txt:

assimUV	Flag for assimilation of current speed (true or false)
N_UV	Number of assimilated measurement locations
Assimileringsavbrudd	Flag for halting or stopping the data assimilation (true or false)
HardConstraints	Flag for putting hard constraints on the model corrections (true or false)
enoiAlpha_init	The alpha parameter in the EnOI assimilation algorithm
stdDevUV_init	Standard deviation of the measurement accuracy
enoiAlpha_fin	0.0
stdDevUV_fin	0.02
uvAssimInterval	Number of simulation samples between each assimilation time
updateAInterval	100
FirstEnsembleSample	The first sample of the ensemble set from the ensemble file
EnsembleSampleInterval	Number of samples between the samples chosen for the ensemble set
N_ensemble_1	Number of samples in the ensemble set
ensembleFile	Ensemble file including the path
uvDataFile1	Measurement data file 1, including the path
uvDataFile2	Measurement data file 2, including the path
uvDataFile3	Measurement data file 3, including the path
uvDataFile4	Measurement data file 4, including the path
UV_MEAS_FIRST	Sample number for the first sample that is read from the measurement file
Save_assimvar	Flag for storing assimilation variables, i.e. corrected model estimate (true or false)

At startup, SINMOD reads these parameter files. The *simSettings.txt* file is used to set up basic parameters of the model run, such as the initial and boundary values, start time and duration of simulation, and location of output files. If the *dataAssim* flag is set to *true*, the *AssimSettings.txt* file will be used to get the data assimilation settings and the location of measured data.

4 Test cases

4.1 LoVe test

In this section we present a test of the data assimilation process performed using the LoVe model setup, and data from the LoVe project for the period October 2.-November 10. 2013. This test was done in anticipation of a full ELMO system test with an additional oceanographic buoy in this area, which was eventually cancelled.

4.1.1 Measurements

Figure 5 and Figure 6 show current roses for the two sensors. At about 218 m depth, which is the depth farthest from the Aquadop sensor head, this sensor shows as expected higher current speeds than the Continental sensor. The Continental sensor shows current directed mainly in the northeast-southwest directions, while the Aquadop sensor mainly show currents in the east-northeast direction. Both the Continental and Aquadop sensors show that the current is directed mainly in the northeast-southwest direction at the bottom of the water column.

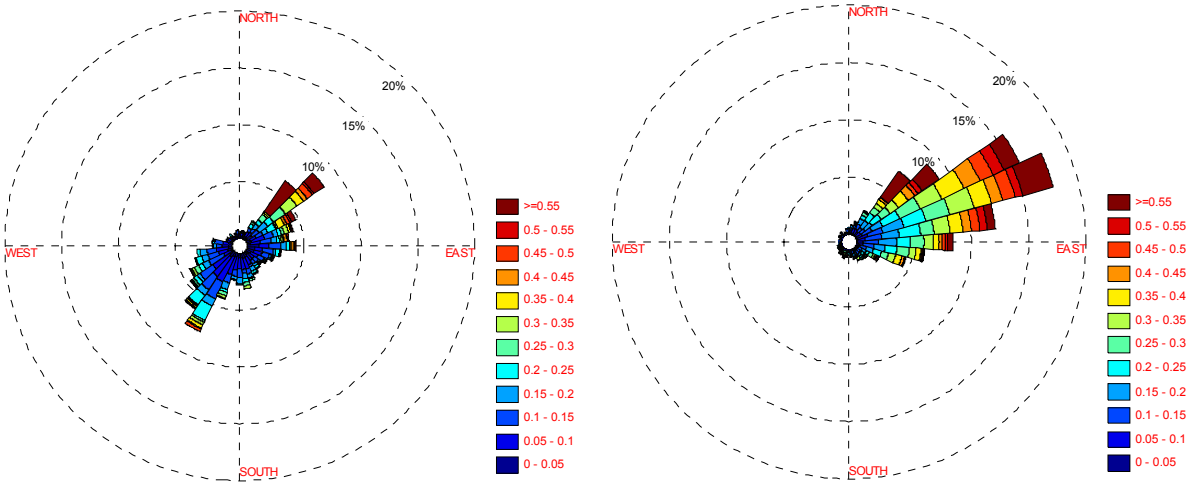


Figure 5: Current roses for Continental (left) and Aquadop sensors (right) at 218 m.

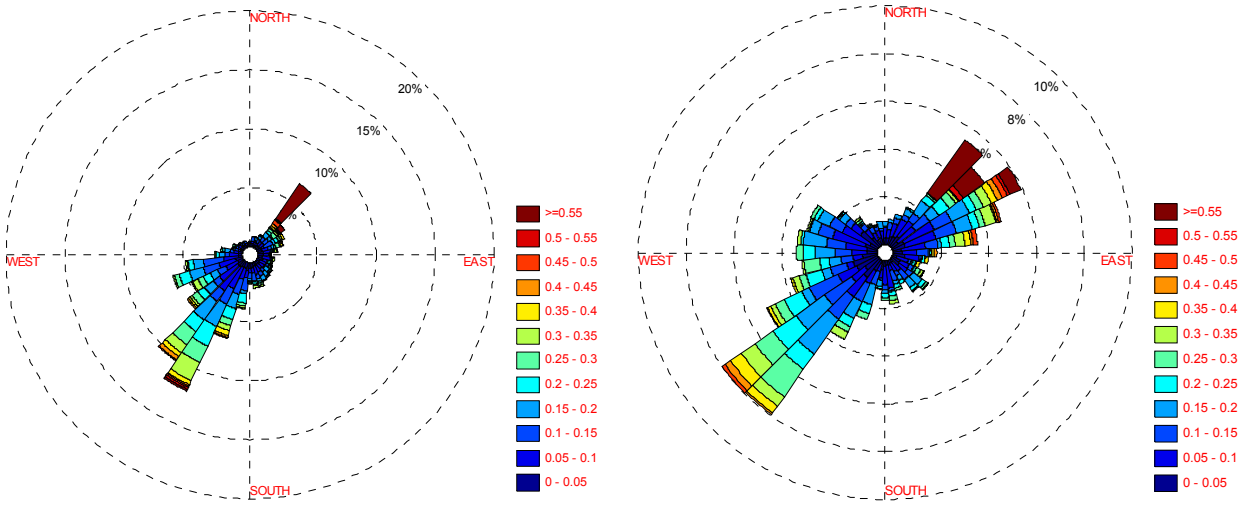


Figure 6: Current roses for Continental (left) and Aquadop sensors (right) at 252 m.

The directional distributions shown by the two sensors deviate significantly, in particular at 218 m. Both sensors are characterized by a lack of currents near the north direction, which is surprising considering the bathymetry of the area around the buoy and the typical pattern of currents in this area. The dominating current directions at 218 m are approximately at a normal angle to what we would expect from the bathymetry in the area. The local bottom topography may have a local influence on the current directions that is picked up by the sensors, but this also leads us to question the quality of the measured data.

4.1.2 Non-assimilated current estimates

Figure 7 and Figure 8 show current roses for estimated current from SINMOD without data assimilation. The modelled directional distributions seem reasonable considering the bathymetry of the area and the larger scale current systems in the area, although the distributions deviate substantially from the measured current directions, with a rotation of about 90 degrees.

Despite the large deviations in the current directions, the current speed is quite similar to the measurements in the middle of the water column as well as at the bottom (time series at 257.5 m shown in Figure 9). The figure also shows that the main dynamic components are the same in the model as in the measurements. A tidal component analysis shows that K1 and O1 are the main tidal components in this period.

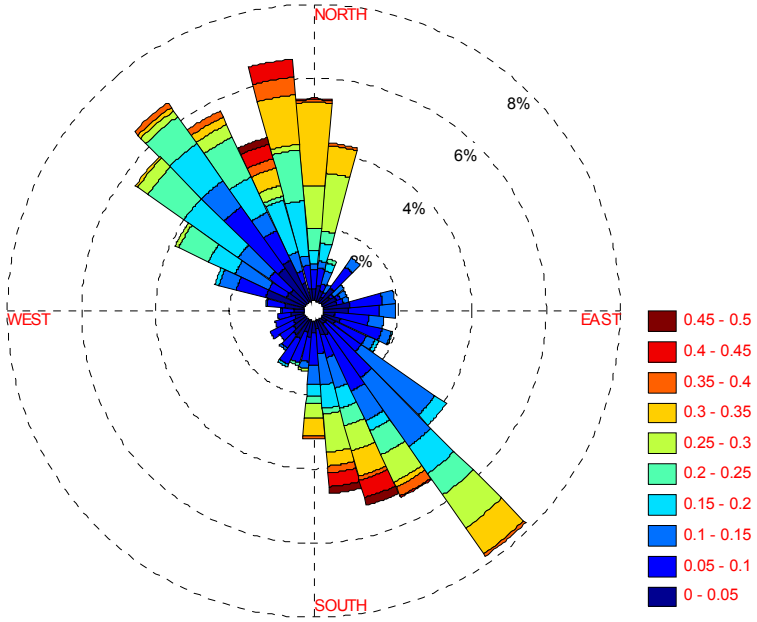


Figure 7: Current rose for non-assimilated current estimate at 212.5 m.

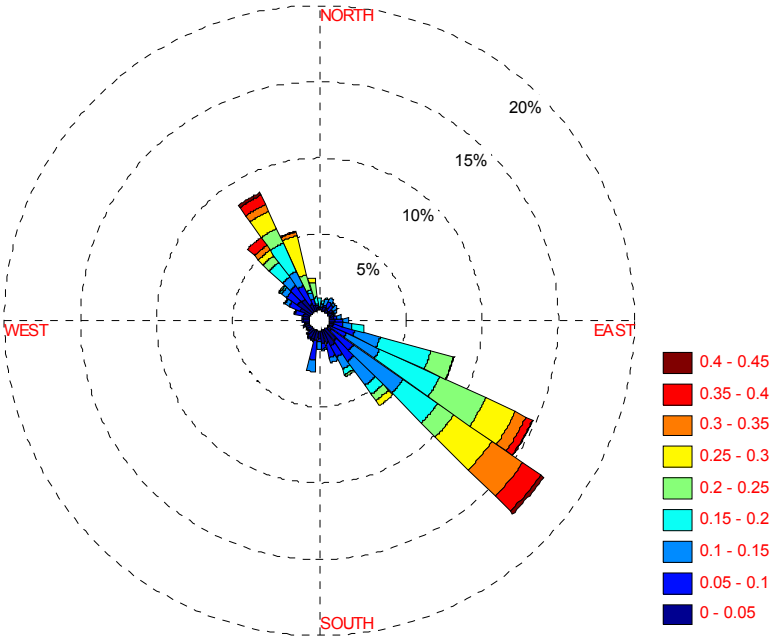


Figure 8: Current rose for non-assimilated current estimate at 252 m.

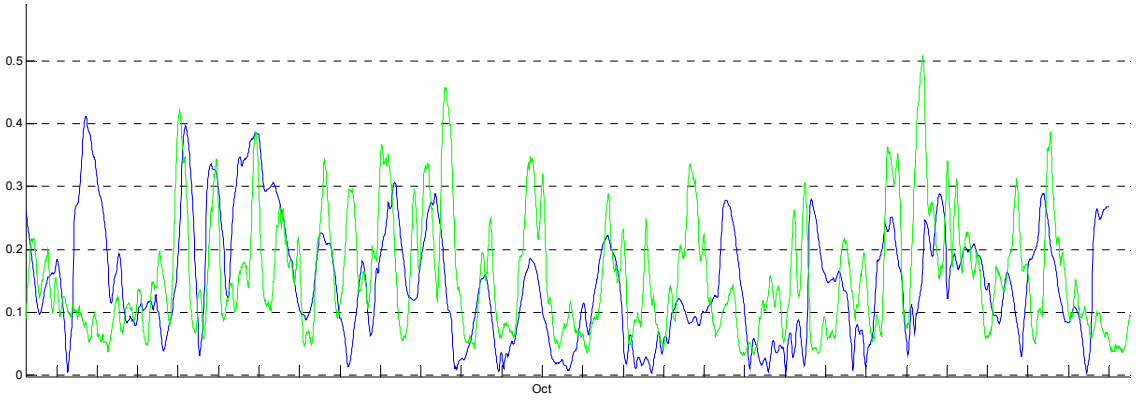


Figure 9: Non-assimilated modelled (blue) vs. measured (green) current speed [m/s] at 257.5 m. The marks on the x-axis denote days.

4.1.3 Assimilated current estimates

Only the Continental sensor measurements are used for data assimilation. At 212.5 m, Figure 10 shows that the model and the measurement currents are fairly well aligned, while at other depths, like 252 m in Figure 11, the model shows more spread in the current direction. This is partly due to the fact that the measurement buoy is located in a position with varying local eddies.

Despite the corrected current directions Figure 12 shows that the model corrections perturb the current speed quite strongly, giving frequently large spikes. These spikes can be reduced by choosing a smaller gain in the correction filter at the cost of a weaker correction of the current direction. The figures also show that more high frequency components are added to the current speed. The main tidal components are, however, retained.

The model corrections are typically largest close to the measurement location and gradually smaller farther away from this location. The current pattern after assimilation is quite complicated, as shown by the assimilated estimates in Figure 13. As noted earlier, the directional distribution of the measured currents are very different from what we would expect at this location, and this leads to the corrections significantly perturbing the model state. Without additional measurement points it is difficult to verify the modified current pattern.

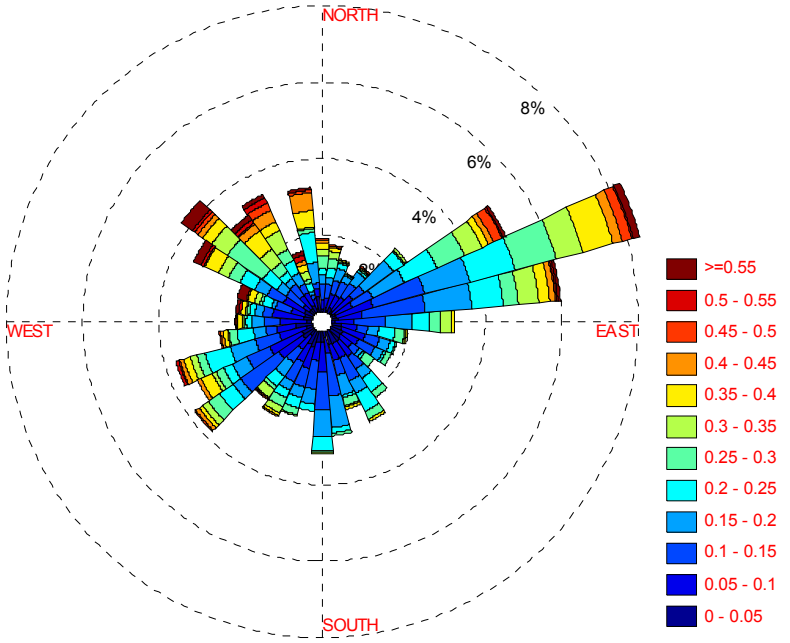


Figure 10: Current rose for assimilated current estimate at 212.5 m.

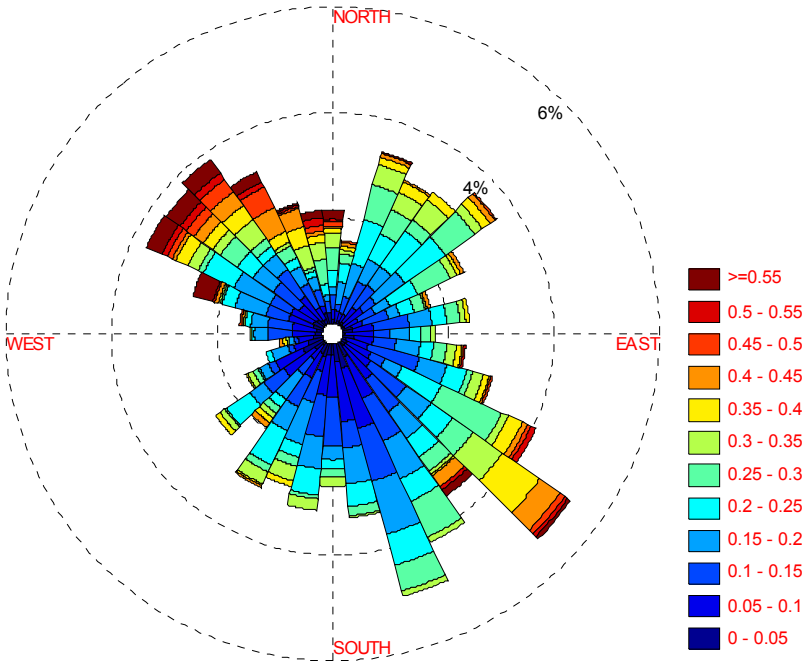


Figure 11: Current rose for assimilated current estimate at 252 m.

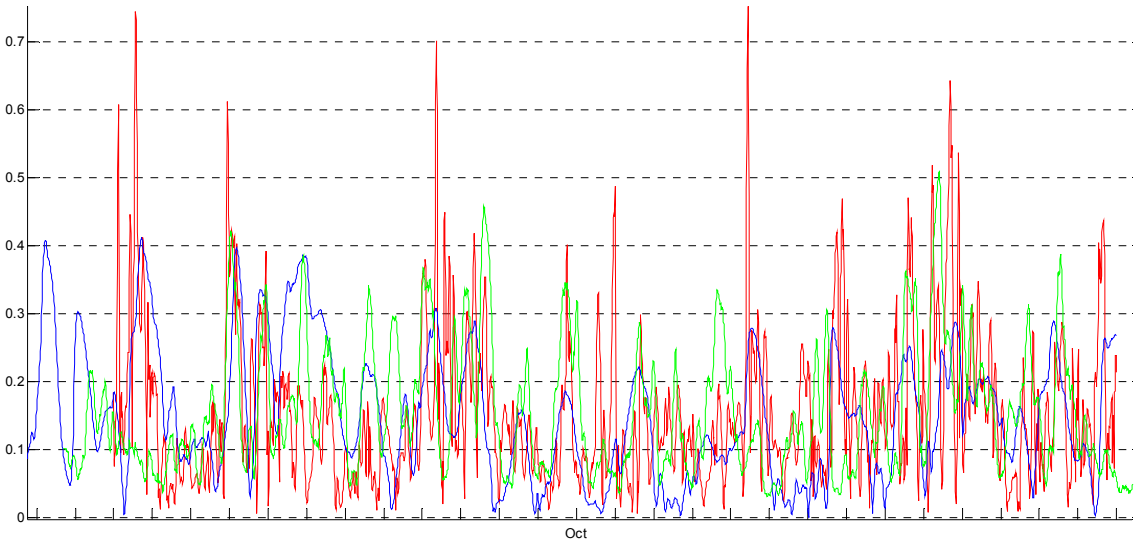


Figure 12: Non-assimilated (blue) vs. assimilated (red) vs. measured (green) current speed [m/s] at 257.5 m. The marks on the x axis denote days.

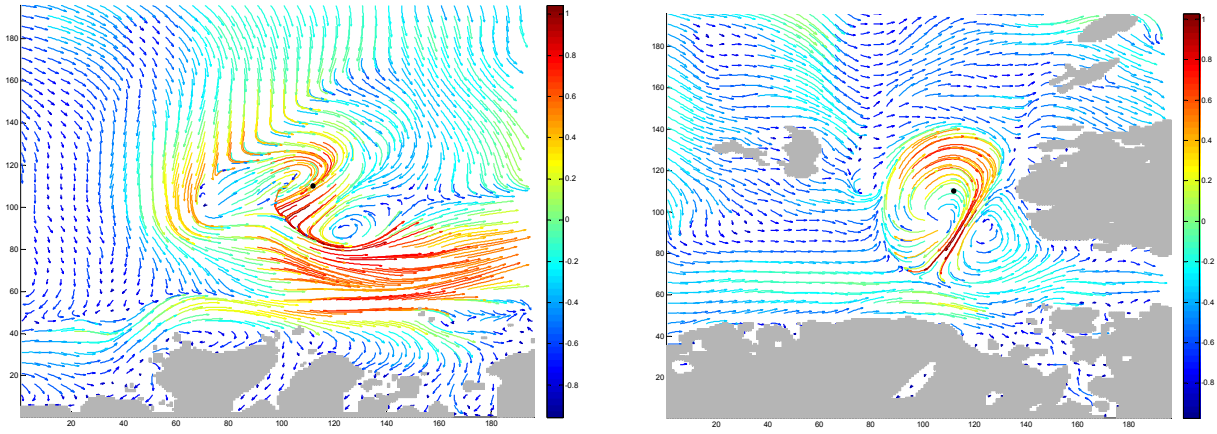


Figure 13: Examples of current patterns at the sea surface (left) and at 100 m (right) after assimilation.

4.1.4 Discussion

There are several possible explanations for the deviations seen in directional distribution between the SINMOD simulations and the measured data. The measurement location is in an area where the overall current direction changes as one moves northwards. Deviations in the location of the current sensor, and in the local bathymetry, may therefore give deviations in the dominating current directions. Comparisons between SINMOD model results and ADCP measurements from Statoil's LoVeCur project at a location 9 km north of the station discussed here, show good agreement both in current speeds and directional distribution.

The size of the ocean area that is modelled and the spatial resolution are important to the model results and the effects of data assimilation. When the ocean area is too small it appears to be hard to tune the filter parameters to give current estimates close to the measurements. This is reasonable since smaller models

generally include less information about the statics and dynamics of the ocean due to the sampling of the boundaries. In this case the model is largely driven by the boundary values. It appears that the selected model domain should be enlarged to give better match to the measurements.

Comparing model estimates with measurements at the measurement location confirms that the model is assimilated only at this location. A more interesting result is to evaluate how well the model is assimilated at other locations in the ocean. By using more measurements, a better evaluation can be made by comparing the model estimate at measurement locations that are not included in the assimilation scheme.

4.2 Frohavet Nord test

The main objective of this test was to test all the parts of the ELMO system together. From the viewpoint of SINMOD, this includes the following tasks:

1. Run the ocean model operationally and make output values available to the other ELMO components
2. Acquire current measurements from the ELMO server, and assimilate this into model runs.

In the absence of measurement data, this test does not allow the quality of the assimilated output to be assessed. As the effects of the data assimilation function were investigated for the LoVe case, a full tuning and assessment of the data assimilation algorithm has not been done for the FroN case.

4.2.1 Operational model system

The operational model controller runs regularly at 30 minute intervals, performing the following operations:

1. At the earliest opportunity, download boundary values from Norkyst800, and atmospheric input from Hirlam12, and convert these to SINMOD format.
2. If initial value data from previous day's simulation is missing, download and convert initial value field from Norkyst800.
3. When boundary and initial value data is available, start 48 hour model run with data assimilation disabled.
4. Whenever the non-assimilating model run has been started, and no model run with assimilation is currently running, check if updated current data is available. If so, start a new 48 hour run assimilating all currently available data.
5. Whenever a model run is finished, add a symbolic link to make the model output available on the THREDDS server.

4.2.2 Comparison between SINMOD and Norkyst800

Comparing the simulated currents from SINMOD with Norkyst800 has value primarily for validation of the operational setup. SINMOD runs using boundary values fron Norkyst800, and the models are therefore expected to show overall agreement. The SINMOD model runs at a higher resolution, with higher resolution bathymetry that may also be based on different input data, and since there are differences between the two models we expect to see differences in the output as well. As there are no measurements available, we have no basis for assessing which of the models produce the most correct output. Figure 14 shows a snapshot of surface currents from Norkyst800 and SINMOD compared.

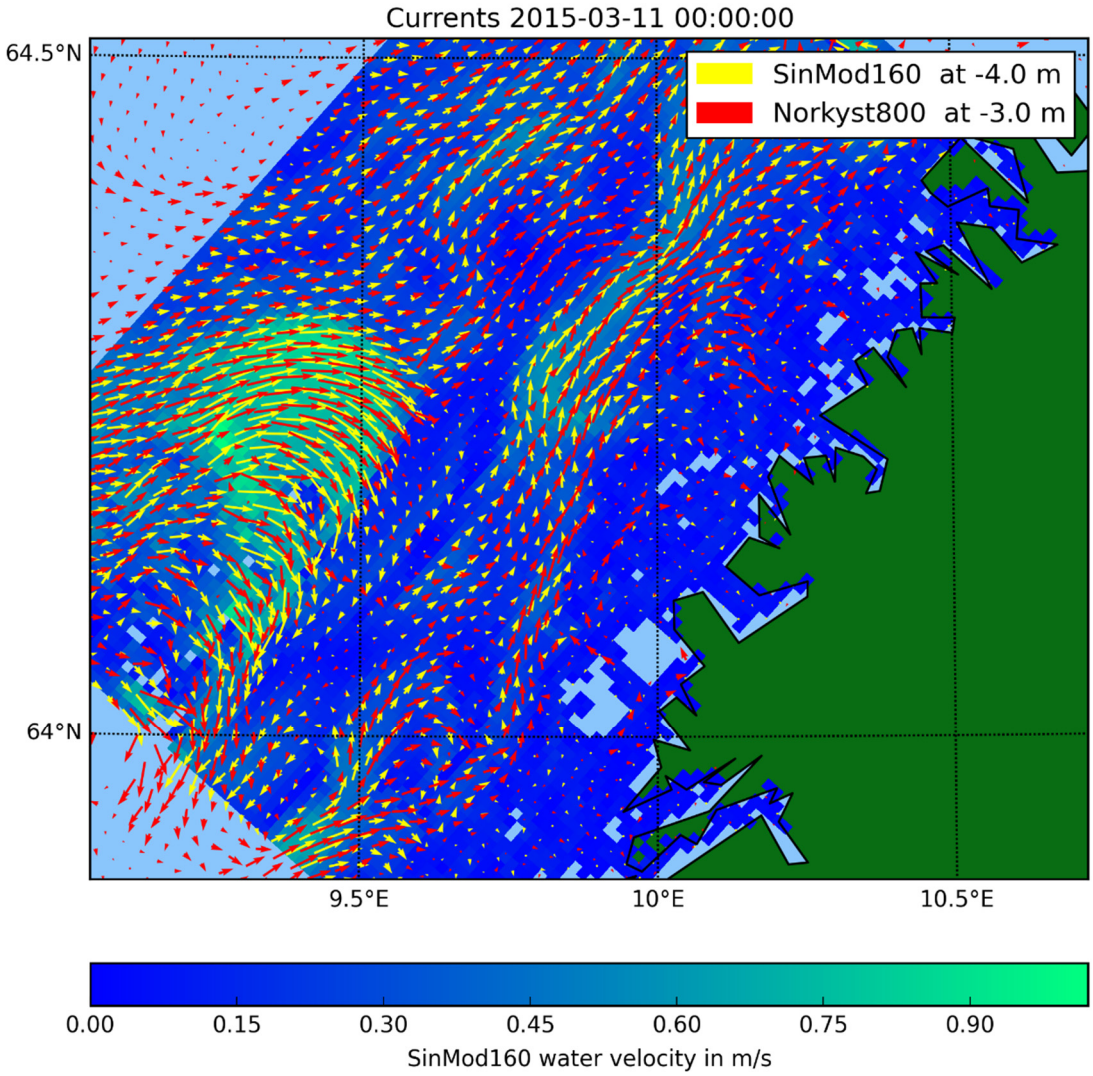


Figure 14: Snapshot comparing near surface currents betwen SINMOD (yellow arrows) and Norkyst800 (red arrows). Note that SINMOD operates at 5 times higher resolution, meaning the SINMOD data are subsampled 5 times more than the Norkyst800 data. Figure by Jan Prosi.

We have extracted time series for the four virtual measurement stations for a short period from 5.-10. April 2015 from the models. The time series for a selection of depths, as well as current roses for 100 m depth, are shown in Figure 15 to Figure 18. Current roses comparing speeds and directional distribution at 100 m depth at the four stations are shown in Figure 19 to Figure 22.

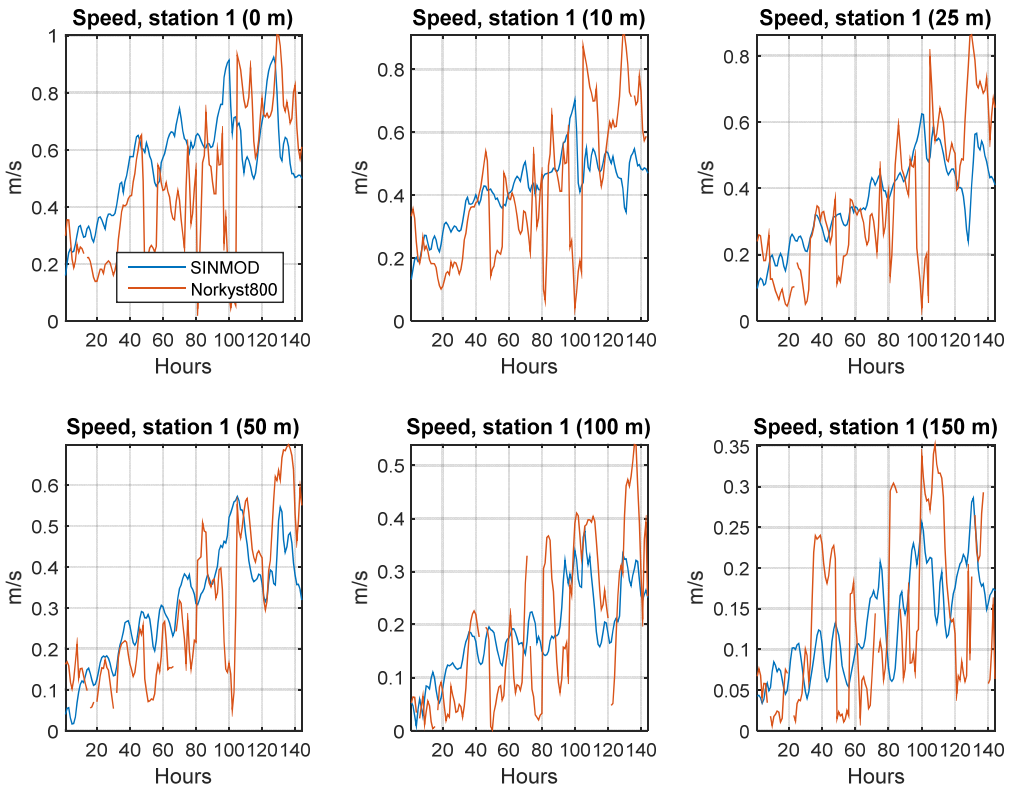


Figure 15: Comparison of current speed in the period from 5th to 11th April 2015 at station 1 at a selection of depths.

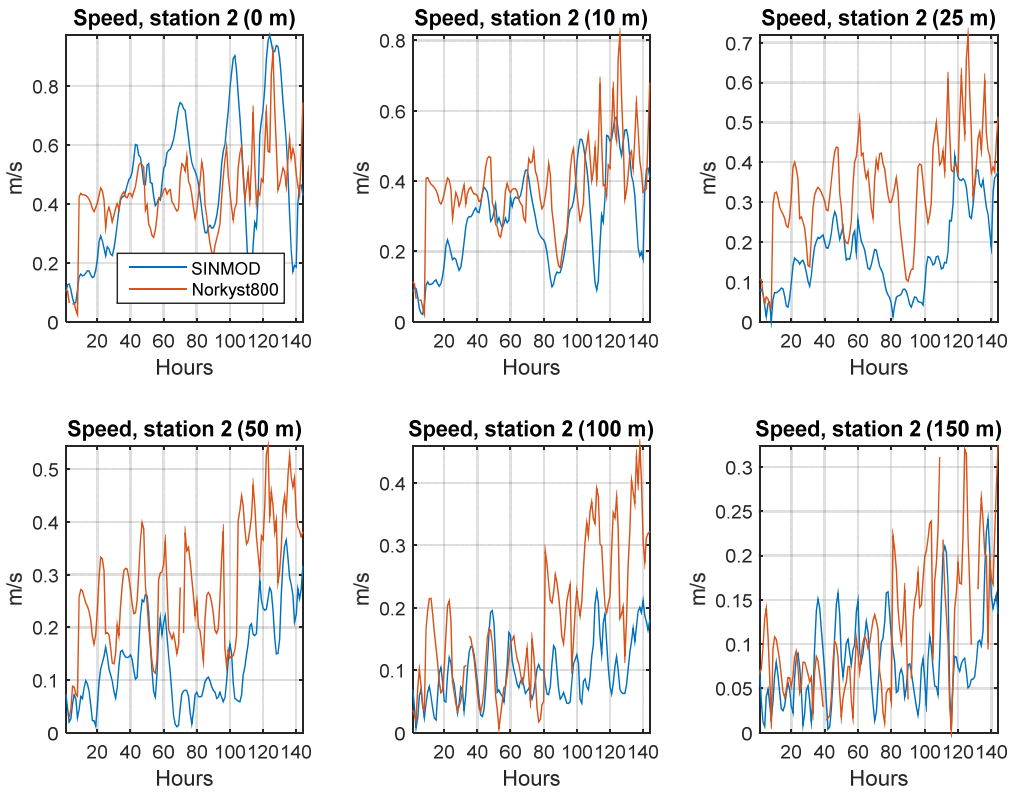


Figure 16: Comparison of current speed in the period from 5th to 11th April 2015 at station 2 at a selection of depths.

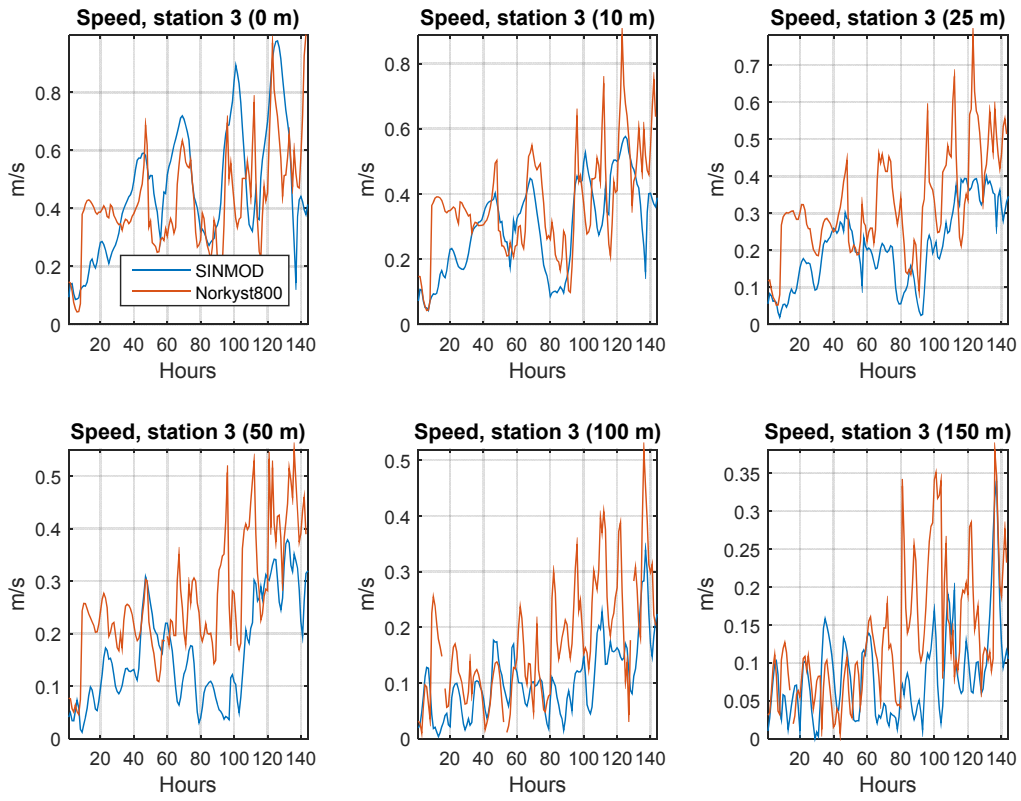


Figure 17: Comparison of current speed in the period from 5th to 11th April 2015 at station 3 at a selection of depths.

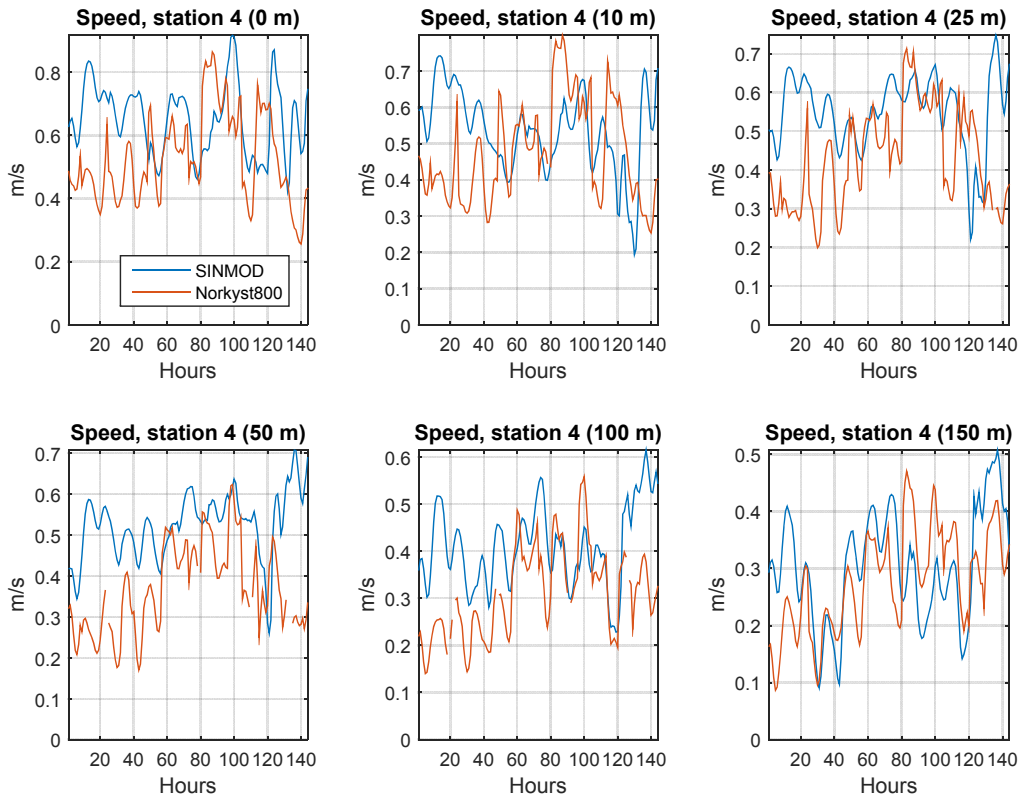


Figure 18: Comparison of current speed in the period from 5th to 11th April 2015 at station 4 at a selection of depths.

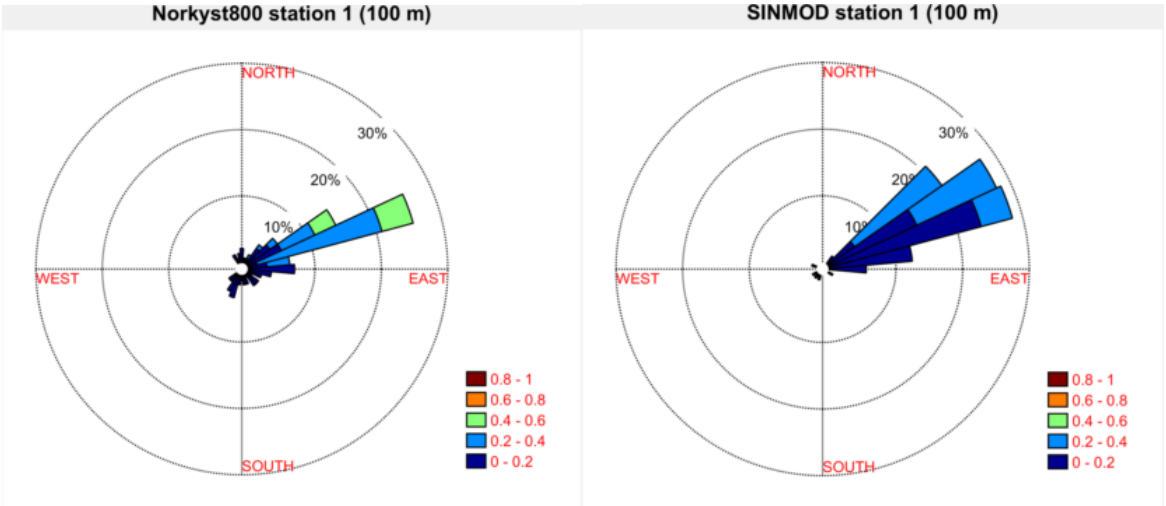


Figure 19: Comparison of speed and directions between Norkyst800 (left) and SINMOD (right) in the period from 5th to 11th April 2015 at Station 1.

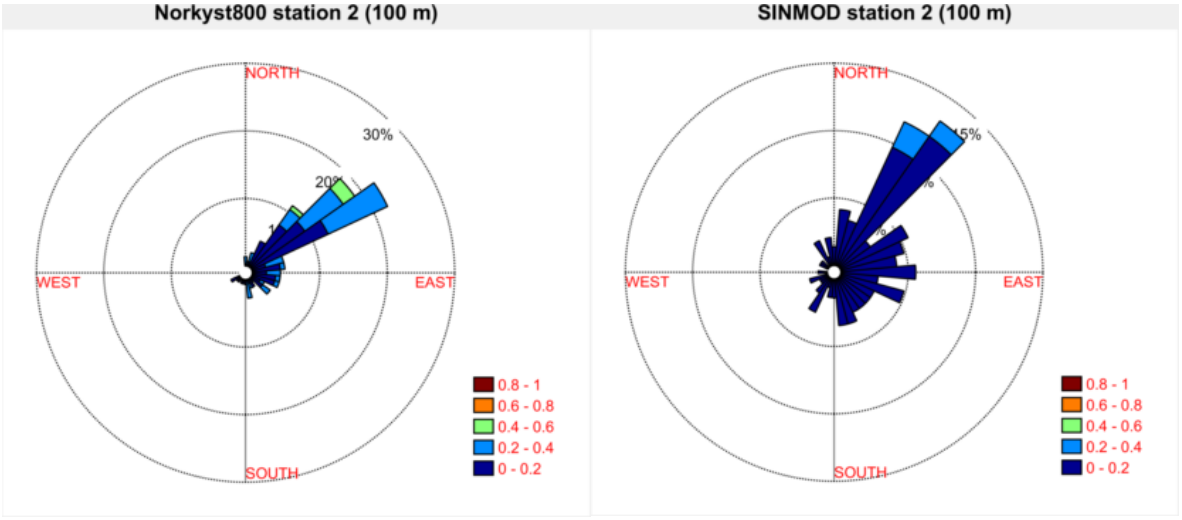


Figure 20: Comparison of speed and directions between Norkyst800 (left) and SINMOD (right) in the period from 5th to 11th April 2015 at Station 2.

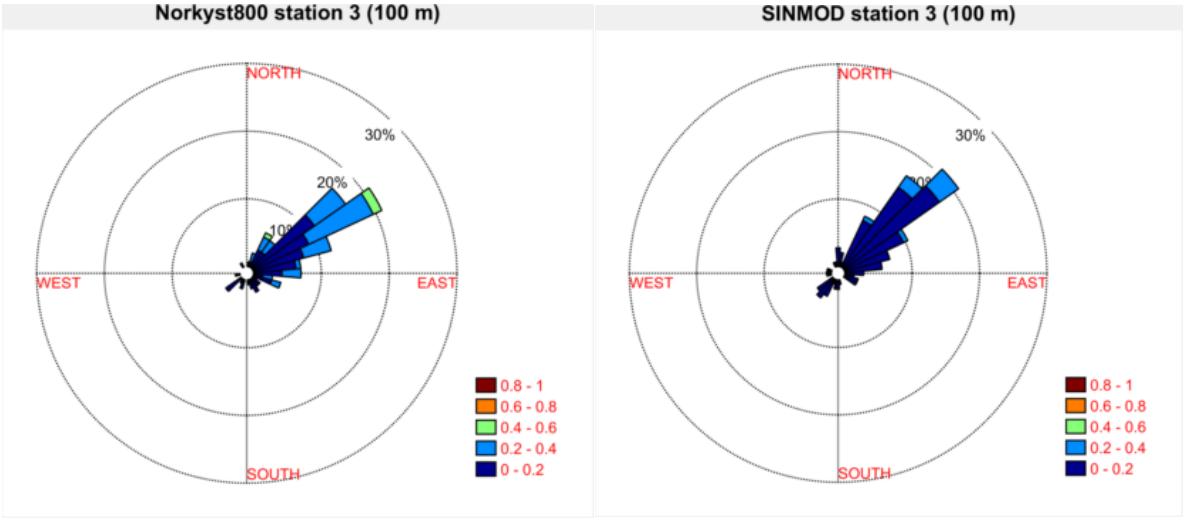


Figure 21: Comparison of speed and directions between Norkyst800 (left) and SINMOD (right) in the period from 5th to 11th April 2015 at Station 3.

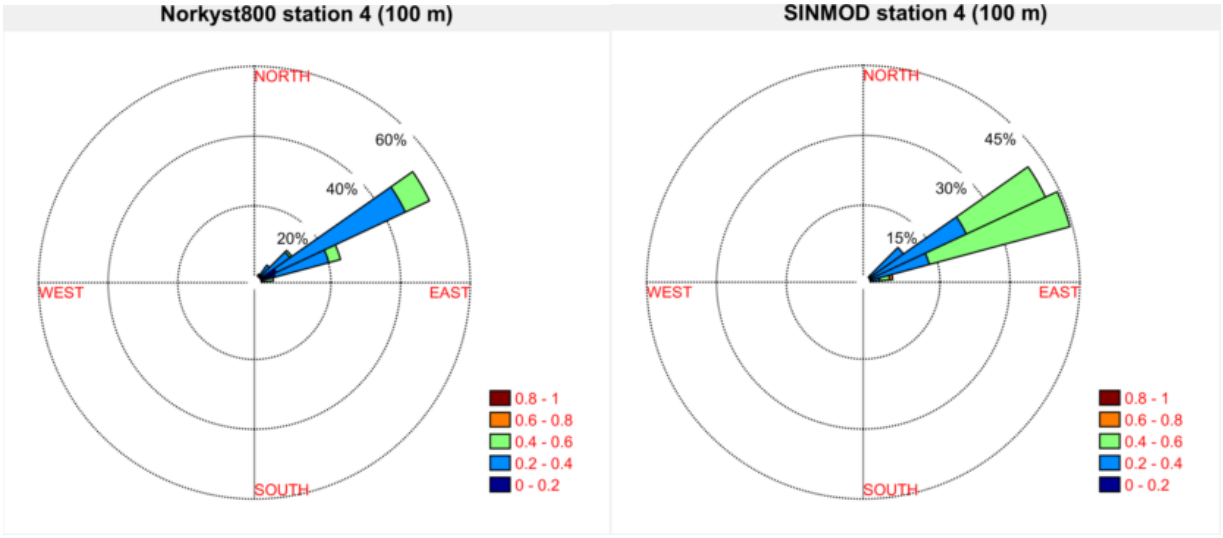


Figure 22: Comparison of speed and directions between Norkyst800 (left) and SINMOD (right) in the period from 5th to 11th April 2015 at Station 4.

4.2.3 Data assimilation

Figure 23 and Figure 24 shows an example of a current field produced by SINMOD directly, compared to a current field after assimilating data at Station 1.

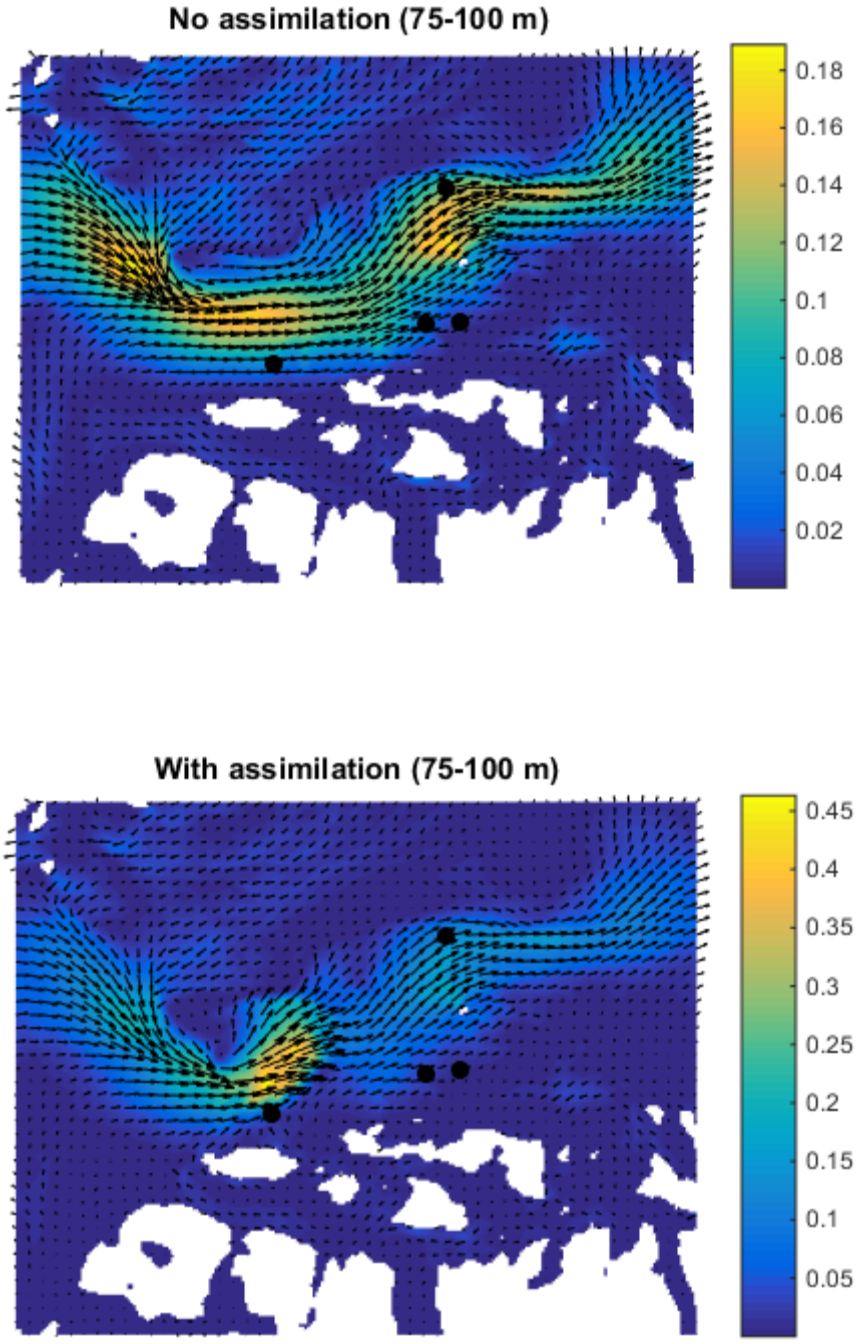


Figure 23: Example of current field (extract from the FroN model area) without assimilation (upper panel) and with assimilation at the lower left station.

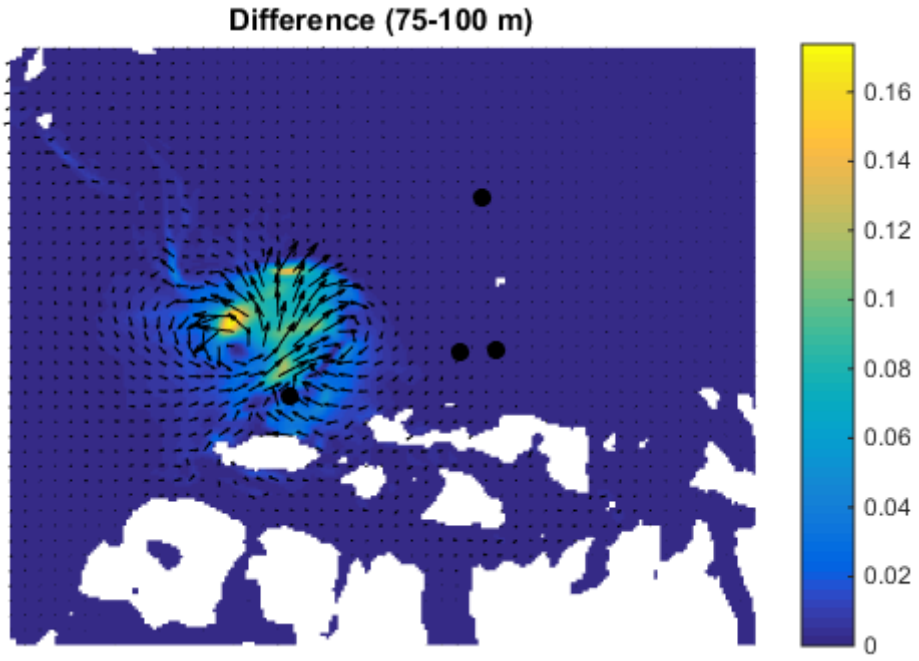


Figure 24: Difference between non-assimilated and assimilated current field.

5 Conclusion

This report has documented the implementation of data assimilation in SINMOD in terms of the choice and implementation of assimilation algorithm, the preprocessing of measured data, and the specification of assimilation settings. The EnIO assimilation algorithm is a light weight method suited to applications such as ocean modelling, where it often makes sense to sacrifice the advantages of a dynamic ensemble in favor of reserving the computational power for running the model at a sufficiently high speed and resolution.

Two test cases have been presented – the LoVe case where the assimilation algorithm was tuned and tested using ADCP data from Statoil's LoVe ocean laboratory, and the FroN case where SINMOD was set up to run based on boundary conditions from the Norkyst800 operational ocean modelling system. Both cases were originally expected to be enhanced by the deployment of an oceanographic measurement buoy, but the deployment was cancelled. Therefore, none of the test cases provided a fully satisfactory test setup.

In the LoVe case, strong deviations in measured vs. modelled current directions were found, with no clear explanation, while the current speeds were similar in the model and measurements. After application of the data assimilation algorithm, the model was forced into producing current directional distributions significantly more similar to the measurements, in particular in the middle part of the water column.

In the FroN case, data assimilation was applied using virtual buoys providing data extracted from the Norkyst800 model. As this was the same model used to drive the boundaries of the SINMOD setup, this could only be a test of the data assimilation algorithm itself, without any validation against real conditions.

6 References

Chepurin, G. A., J. A. Carton, and D. Dee (2005), Forecast Model Bias Correction in Ocean Data Assimilation, *Monthly Weather Review*, *133*(5), 1328-1342.

Counillon, F., and L. Bertino (2009), Ensemble Optimal Interpolation: multivariate properties in the Gulf of Mexico, *Tellus*, *61A*, 296–308.

Dee, D. P. (2005), Bias and data assimilation, *Q. J. R. Meteorol. Soc.*, *131*, 3323–3343.

Erwig, M. (2008), The Inverse Ocean Modeling System. Part I: Implementation, *Journal of Atmospheric and Oceanic Technology*, *September*.

Evensen, G. (2003), The Ensemble Kalman Filter: theoretical formulation and practical implementation, *Ocean Dynamics*, *53*, 343–367.

Fu, W., J. She, and S. XZhuang (2011), Application of an Ensemble Optimal Interpolation in a North/Baltic Sea model: Assimilating temperature and salinity profiles, *Ocean Modelling*, *40*, 227–245.

Grythe, K., I. Jensen, A. Lie, T. A. Reinen, M. Omholt Alver, G. Eidnes, F. A. Michelsen, M. Reed, and D. Slagstad (2013), Ocean space surveillance system - OSS, *Oceans'13, MTS/IEEE Bergen, June 10-13*.

Hillier, F. S., and G. J. Lieberman (2005), Introduction to operations research, *McGraw-Hill*.

Houtekamer, P. L., and H. L. Mitchell (2001), A Sequential Ensemble Kalman Filter for Atmospheric Data Assimilation, *Monthley wether review, American Meteorological Society 129*.

Julier, S. J., and J. K. Uhlmann (1997), A new extension of the Kalman filter to nonlinear systems, *Int. Symp. Aerospace/Defense Sensing, Simulation and Controls*, *3*.

Kalman, R. E. (1960), A new approach to linear filtering and prediction problems, *Transactions of the ASME - Journal of Basic Engineering*, *82*(1), 35–45.

Moore, A. M., H. G. Arango, G. Broquet, B. S. Powell, A. T. Weaver, and J. Zavala-Garay (2011), The Regional Ocean Modeling System (ROMS) 4-dimensional variational data assimilation systems Part I – System overview and formulation, *Progress in Oceanography*, *91*, 34–49.

Muccino, J. C., et al. (2008), The Inverse Ocean Modeling System. Part II: Applications, *JOURNAL OF ATMOSPHERIC AND OCEANIC TECHNOLOGY*, *25*(9), 1623-1637.

Oey, L.-Y., T. Ezer, G. Forristall, C. Cooper, S. DiMarco, and S. Fan (2005), An exercise in forecasting Loop Current and eddy frontal positions in the Gulf of Mexico, *Geophys. Res. Lett.* , *32*(L12611, 2005GL023253).

Oke, P. R., P. Sakov, and S. P. Corney (2007), Impacts of localisation in the EnKF and EnOI: experiments with a small model, *Ocean Dynamics*, *57*, 32–45.

Rawlins, F., S. P. Ballard, K. J. Bovis, A. M. Clayton, D. L. Inverarity, G. W. , A. C. Lorenc, and T. J. Payne (2007), The Met Office global four-dimensional variational data assimilation scheme, *Quart. J. Roy. Meteor. Soc.*, *133*, 347-362.

Slagstad, D., and T. A. McClimans (2005), Modeling the ecosystem dynamics of the Barents Sea including the marginal ice zone: I. Physical and chemical oceanography, *Journal of Marine Systems*, *58*.

Xu, F.-H., L.-Y. Oey, Y. Miyazawa, and P. Hamilton (2013), Hindcasts and forecasts of Loop Current and eddies in the Gulf of Mexico using local ensemble transform Kalman filter and optimum-interpolation assimilation schemes, *Ocean Modelling*, *69*, 22–38.

Technology for a better society

www.sintef.no