

Data fusion and data assimilation of ice thickness observations using a regularisation framework

By NAZANIN ASADI*, K. ANDREA SCOTT, and DAVID A. CLAUSI

Department of Systems Design Engineering, University of Waterloo, Waterloo, Canada

(Manuscript received 5 January 2018; in final form 21 December 2018)

ABSTRACT

Accurate estimates of sharp features in the sea ice cover, such as leads and ridges, are critical for shipping activities, ice operations and weather forecasting. These sharp features can be difficult to preserve in data fusion and data assimilation due to the spatial correlations in the background error covariance matrices. In this article, a set of data fusion and data assimilation experiments are carried out comparing two objective functions, one with a conventional l2-norm and one that imposes an additional l1-norm on the derivative of the ice thickness state estimate. The latter is motivated by analysis of high resolution ice thickness observations derived from an airborne electromagnetic sensor demonstrating the sparsity of the ice thickness in the derivative domain. Data fusion and data assimilation experiments (using a 1D toy sea-ice model) are carried out over a wide range of background and observation error correlation length scales. Results show the superiority of using an l1–l2 regularisation framework. For the data fusion experiments it was found when both background and observation error correlation length scales are zero, the ice thickness root mean squared error for the l1–l2 method was 0.16 m as compared to 0.20 m for the l2 method. The differences between the methods were greater when the background error correlation length scale was relatively short (approximately five times the analysis grid spacing), and were not significant for larger background error correlation length scales (e.g. 10 times the analysis grid spacing). For data assimilation experiments it was found that openings in the ice cover were captured better with the l1–l2 regularisation, with reduced errors in ice thickness, concentration and velocity. In addition, the ice thickness derivatives in the analyses were found to be more sparse when the l1–l2 method was used and are closer to the those from the true model run.

Keywords: sea ice, data assimilation, ice thickness, regularisation, 3D-Var, sparsity

1. Introduction

According to recent records of sea ice concentration and thickness, there has been a significant decline in Arctic ice extent, and thinning of the ice cover, over the past two decades (Meier et al., 2014; Renner et al., 2014). It also has been recently shown that there is a significant correlation between reductions in ice concentration in the seasonal ice zone and increases in shipping activity (Pizzolato et al., 2016). However, the safety of navigation in ice covered regions depends in part on the availability of accurate estimates of the small scale details of ice concentration, ice thickness and ice pressure. Ships will tend to navigate through narrow openings in the ice cover, and if there are none present they will attempt to find a route where the ice appears to be thin and/or undeformed (Loiptien and Axell,

2014). Sudden changes in ice thickness (e.g. deformation ridges) and ice pressure play a significant role in vessel besetting events (Kubat et al., 2013).

Ideally, sea ice forecasts would be made in a similar manner to weather forecasts, using data assimilation (Carrieres et al., 2017). Data assimilation is a method for combining available observations with a background state from a numerical forecast model, to find the best estimate of the current state of the system, which is called the analysis (Lahoz et al., 2010). The analysis is then used to initialise the forecast model to run a short-term forecast, the output of which is used as the background state for the next data assimilation cycle. Using a data assimilation system would allow physically-based forecasts to be made of features in the ice cover that cannot be directly observed, such as regions of high ice pressure.

Data assimilation schemes involve an interpolation of observational information to a model grid. To carry out this interpolation certain criteria should generally be

*Corresponding author. e-mail: n2asadi@uwaterloo.ca; nazy.asady@gmail.com

satisfied, such as smoothness of the underlying fields (Wahba and Wendelberger, 1980; Lorenc, 1986). This poses a specific problem for sea ice because even though at large scales sea ice seems to be a nonrigid continuum, at small scales it includes important sharp features or discontinuities (Thomas et al., 2008).

One way to move towards a better representation of these sharp features in the state estimate is to revisit the definition of the data assimilation problem. Data assimilation can be considered an inverse problem where the objective is to find the best state, $\mathbf{x} \in \mathbb{R}^n$, that satisfies the relationship $\mathbf{y} = H(\mathbf{x}) + \varepsilon$, where $\mathbf{y} \in \mathbb{R}^m$ represents the available observations, $H: \mathbb{R}^n \rightarrow \mathbb{R}^m$ is the observation operator, which maps the state vector to the observation space and can be a physical or mathematical model, and $\varepsilon \in \mathbb{R}^m$ is the observation error. The standard objective function in an inverse problem is defined as the l_2 -norm of the misfit between the observations and model state as $\|\mathbf{y} - H(\mathbf{x})\|_2^2$. Since, usually there are not enough observations available (i.e. the dimension of $\mathbf{y}$ is less than $\mathbf{x}$) and/or the observations have errors, the data assimilation problem may be very ill-posed. This problem can be solved by adding regularisation terms to the standard objective function. For example, in variational data assimilation (VDA) an l_2 norm is added to the objective function that is the difference between the optimal state and a background state, where the latter is typically from a forecast model.

The relationship between VDA and Tikhonov regularisation was discussed by Johnson by reformulating the objective function of VDA as a regularisation problem (Johnson et al., 2005). In later studies, where sharp features in the analysis were desired, the l_2 -norm term in the objective function was replaced with the l_1 -norm (Freitag et al., 2010; Budd et al., 2011). This follows similar practice in image deblurring problems, where an original image is estimated from a given blurred image. The l_1 -norm regularisation yields better performance than l_2 norm regularisation when sharp edges need to be recovered (Hansen et al., 2006). The reason for the edge preserving property of the l_1 -norm is that edges in images lead to outliers in the regularisation term, which are penalised less in an l_1 -norm than an l_2 -norm. A similar l_1 regularisation was taken in Foufoula-Georgiou et al. (2014) and Ebtehaj et al. (2015), where it was demonstrated that precipitation states and global fields of temperature, humidity and geopotential height admit a nearly sparse representation in a specific domain (derivative or wavelet domain). In the interest of retaining this sparsity in a data assimilation framework, the VDA problem was solved by applying l_1 -norm regularisation on the analysis state derivative. In a similar spirit, Freitag et al. (2013) added a term to the standard l_2 objective function

constraining the l_1 -norm of the gradient of the analysis vector to follow a Laplacian density for experiments involving sharp fronts. Ebtehaj et al. (2014) extended this study by proposing a generalised regularisation framework and applying this framework to an advection-diffusion equation enforcing sparsity in wavelet and spectral domains. They found this l_1 - l_2 -norm regularisation framework performed better than the standard l_2 -norm regularisation method for their problem in idealised 4-D-Var experiments.

In this study, evidence of sparsity in the ice thickness derivative domain is demonstrated using both airborne survey data, and submarine sonar data. The impact of using a mixed variational regularisation framework, in which the l_2 term is augmented with an l_1 term in the derivative domain, to retain this sparsity in sea ice thickness analyses, is demonstrated using both data fusion and data assimilation experiments. For the data assimilation experiments a 1-D sea ice model is used (Stonebridge et al., 2018), whereas the data fusion experiments do not use a prognostic model. Note similar studies were carried out in Ebtehaj and Foufoula-Georgiou (2013) and Freitag et al. (2013) where the data assimilation was carried out using a 1-D heat equation, and 1-D advection equation, respectively. In addition to using a more complicated physical model, our experiments examine not only the impact of background error correlations on the analyses, but also the impact of observation error correlations. The latter is motivated by both past (Daley, 1993) and recent (Rainwater et al., 2015) studies demonstrating the role of observation error correlations in retaining small scale information in the analyses.

The article is constructed as follows: In Section 2 the data is described and presence of sparsity in the derivative field of data is statistically illustrated. In Section 3 the assimilation framework and the formulation of the l_1 -norm regularisation problem is presented. Following that, in Section 4 the experimental set-up is described, while in Section 5 the results of the experiments are shown and discussed. Finally in Section 6 concluding remarks are given.

2. Sparsity of ice thickness observations

At the small scales, similar to observations of precipitation and other geophysical variables (Foufoula-Georgiou et al., 2014; Ebtehaj et al., 2015), sea ice no longer behaves as a continuum (Flato, 1998), and exhibits sharp features or discontinuities. To visualise sharp features in the sea ice state, ice thickness measurements obtained using an airborne electromagnetic (AEM) sensor during an aircraft survey over the Beaufort Sea are examined. The AEM sensor measures the ice + snow thickness, which we will refer to as the ice thickness (Haas and

Howell, 2015), with a spatial distance of 4.5m–6m (ground distance) between measurement points. The data used here were acquired on 20 April 2015 in the Beaufort Sea. The measurement sample points indicating the flight path are shown in Fig. 1a. For more information about AEM ice thickness measurements see (Haas et al., 2009).

In the Beaufort Sea, the ice cover is highly heterogeneous, and consists of a mixture of thin and thick ice, first-year and multi-year ice, with many ridges. The ridges are spatially localised regions with higher ice thickness than the surrounding ice. While the AEM measurement is known to underestimate the maximum thickness of pressure ridges, it has high accuracy over level ice (± 0.1 m) (Haas et al., 2009, 2010) and it can represent the variability in ice thickness at small spatial scales, as can be seen in Fig. 1b.

To demonstrate the sparsity in the derivative field of AEM ice thickness, various methods are used. In a sparse distribution, the distribution is more peaked at zero, indicating many points with value zero, with fatter tails, as compared to a Gaussian distribution. The histogram of the first order spatial derivative of the AEM ice thickness measurements (taking into account all data points in Fig. 1b) is shown in Fig. 2a. Overlaid on this distribution are fitted Gaussian, Laplacian and generalised Gaussian distributions. The generalised Gaussian distribution is defined as

$$P_X(x) = \frac{p}{2\sigma\Gamma(1/p)} \exp\left(-\left|\frac{x}{\sigma}\right|^p\right), \quad (1)$$

where $\Gamma(z)$ is the Gamma function and p and σ are non-negative parameters describing the shape and width of the density, respectively. For special cases of $p=1$ and $p=2$, this probability distribution corresponds to Laplacian and Gaussian distribution, respectively. The number of bins in the histograms of Fig. 2 was set to the maximum of the Sturges and Freedman Diaconis estimators (Sturges, 1926; Freedman and Diaconis, 1981) and the AEM measurement points were resampled to have an equal spacing of 7 m as this simplifies application of the methods used.

It can be seen in Fig. 2a that the data exhibits a symmetric representation in the derivative space with a large mass around zero. The large number of zero coefficients in the derivative domain is caused by the uniformity of ice thickness over a large area, whereas the large tails are associated with the ridges and/or openings in the ice cover. Note the density of the histogram values around zero is greater than the fitted Gaussian distribution.

For comparison, histograms of the first order spatial derivatives of sea ice draft from submarine upward looking sonar (National Snow and Ice Data Center (NSIDC), 1998) and ice thickness retrieved from Cryosat data¹ (Laxon et al., 2013) are also shown in Fig. 2. Submarine

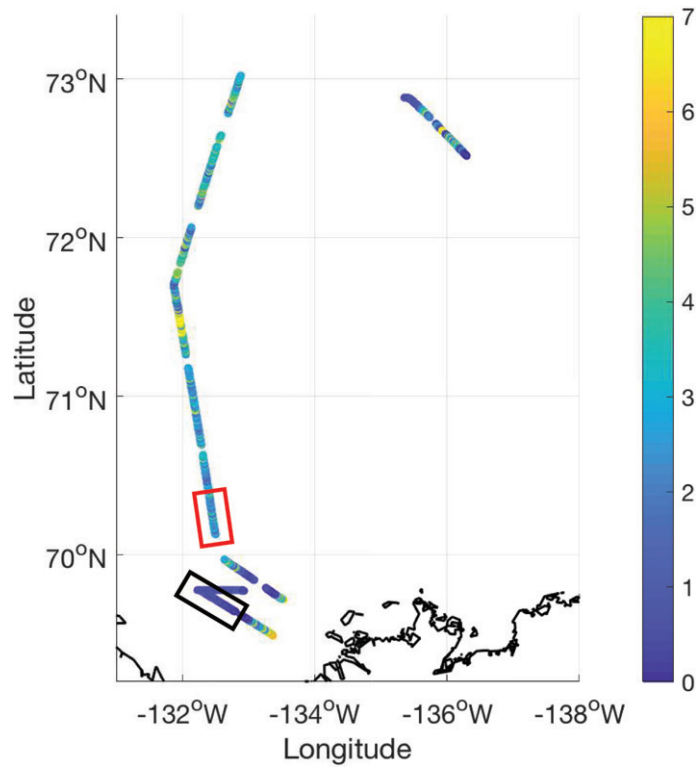
data in Fig. 2b is from an archive of data acquired in Arctic ocean in July 2005 and is resampled at 10m. It was expected that these data would be less sparse than the AEM data, since the underside of the ice is known to be irregular, containing fewer regions than the surface that can be considered level (Rothrock and Thorndike, 1980). This is supported by the histogram in Fig. 2b, where it can be seen that while the fit is closest to the Laplacian distribution, derivative coefficients have lower concentration around zero than that of the AEM data showing that AEM data peak is sharper. As another example of ice thickness sparsity, Fig. 2c shows the histogram of the derivative field of the Cryosat sea ice thickness measurements at 5 km grid from March and April of 2015. The distribution of ice thickness derivatives is highly peaked around zero, showing a better fit to Laplacian and Generalised Gaussian distributions rather than a Gaussian distribution.

To quantitatively describe the goodness of fit of the represented distributions in Fig. 2 to the ice thickness derivative field distribution, Kullback-Leibler divergence (D_{KL}) (Kullback and Leibler, 1951) is used. D_{KL} is a measure of the distance between two distributions. For the discrete probability distributions P and Q it is defined as

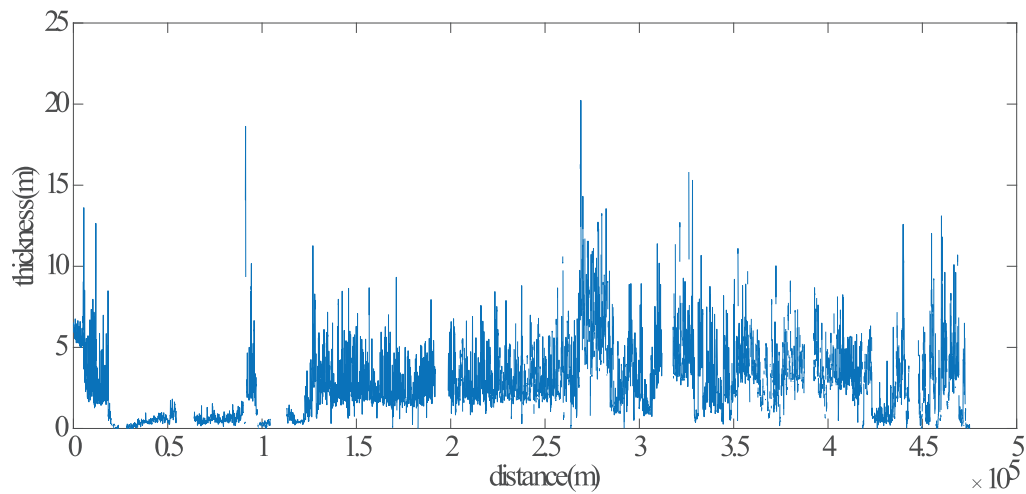
$$D_{KL}(P||Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)} \quad (2)$$

where i is the size of the discrete probability distributions defined by the number of bins in the histogram. Table 1 represents the D_{KL} values for the three datasets in Fig. 2. For AEM data, the D_{KL} values of two subsets of data, selected as smooth and deformed first year ice (FYI) corresponding to the black and red boxes in Fig. 1, respectively, are also included. For each distribution, D_{KL} was calculated by substituting the histogram bin values for P and the fitted distribution values for each bin by Q into (2). The D_{KL} values of the first column (AEM_all) reflect what can be seen in the histogram, which is that the distribution of the AEM ice thickness derivatives is statistically closer to a Laplacian or generalised Gaussian distribution (with $p=0.8$ in (1)) than a Gaussian distribution. While the D_{KL} values for the Cryosat data are relatively large, the Laplacian distribution is a better fit in terms of D_{KL} values to the AEM data and other datasets compared to Gaussian distribution.

To further evaluate the fit of the proposed distributions to the AEM data, a Q-Q plot (Chambers, 1983) is used. In a Q-Q plot the sample data points are plotted against the quantiles of an assumed theoretical distribution in such a manner that the points should form approximately a straight line. Departure from the straight line indicates departure from the specified distribution. Fig. 3 shows that Laplacian and generalised Gaussian



(a) Measurement path of the AEM ice thickness survey
on April 20, 2015 in the Beaufort Sea.



(b) The sequence of the AEM thickness data for the measurement path shown in panel (a)

Fig. 1. Sea ice thickness measurements from airborne electromagnetic (AEM) sensor in the Beaufort Sea acquired on 20 April 2015. (a) Sea ice thickness [m] is represented by the colorbar. The red rectangle outlines a region with deformed first year ice (FYI) while the black rectangle outlines a region with thinner and smoother first-year ice. These regions were determined from visual analysis of a SAR image acquired on 19 April 2015. (b) Sequential representation of the AEM data shown in panel (a).

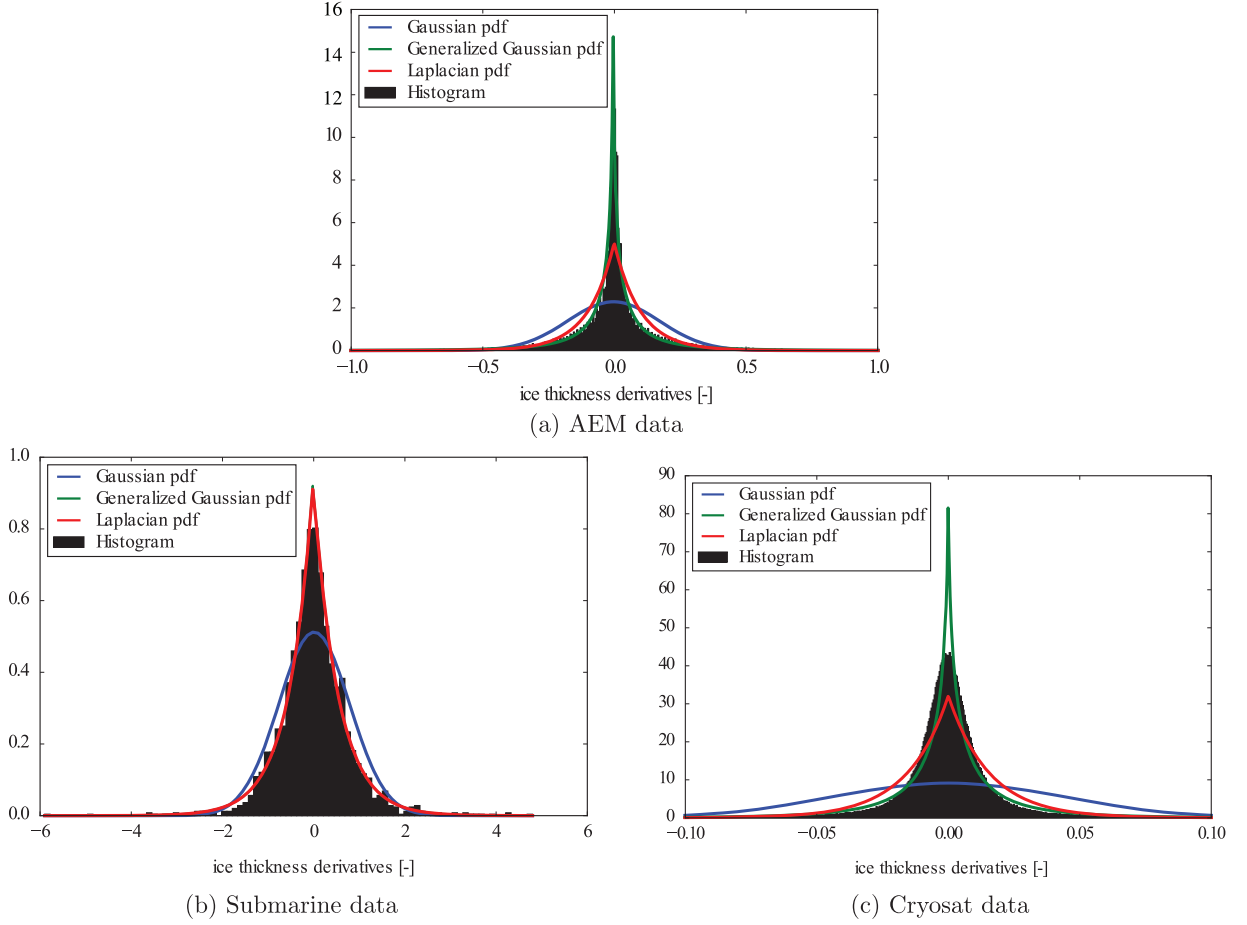


Fig. 2. Normalised histograms of the ice thickness spatial derivative fields of AEM data (a), submarine data (b), and Cryosat data (c) with the fitted Gaussian, Laplacian and generalised Gaussian distributions. All histograms show more similarity to generalised Gaussian and Laplacian distributions rather than a Gaussian distribution. For the submarine data (b), the fitted generalised Gaussian distribution and the Laplacian distribution overlap.

Table 1. Kullback-Leibler divergence (D_{KL}) between different ice thickness datasets and their fitted distributions on the derivative field.

Distribution	AEM_all	AEM_smooth_FYI	AEM_deformed_FYI	Cryosat	Submarine
Gaussian	0.44	0.47	0.3	0.81	0.13
Laplacian	0.15	0.12	0.07	0.23	0.03
Generalised Gaussian	0.02	0.08	0.04	0.06	0.03

A lower number indicates a closer fit to the specified distribution. AEM_all refers to all AEM thickness measurements shown in Fig. 1, while AEM_smooth_FYI and AEM_deformed_FYI refer to the data in the black and red boxes in Fig. 1, respectively.

distribution are better fits to the data in comparison with a Gaussian distribution.

3. Data assimilation method

The assimilation method used here is a form of a VDA method, formulated as an inverse problem with

regularisation, that will be discussed in detail in the following sections.

3.1. Variational data assimilation

The VDA scheme can be defined as the following minimisation problem,

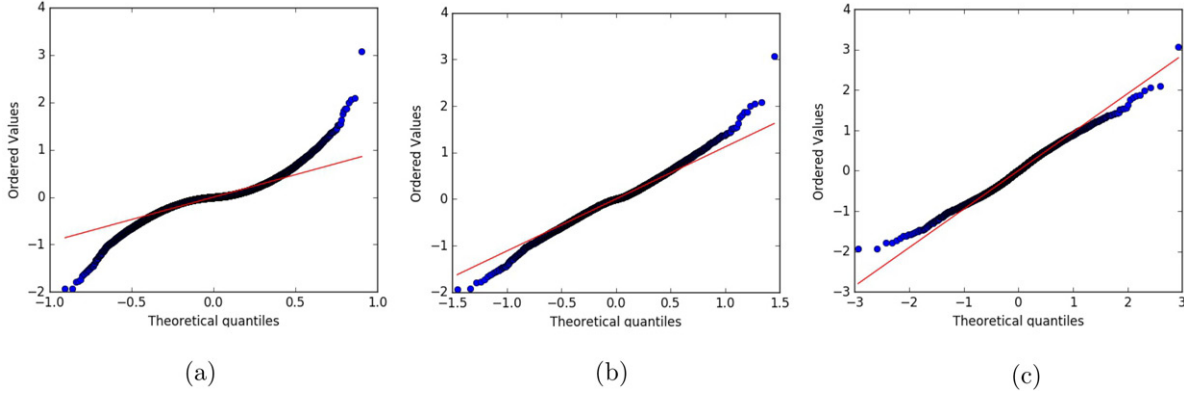


Fig. 3. Q-Q plot of the data derivative field against Gaussian, Laplacian and generalised Gaussian distributions. The horizontal axes are the quantiles of the fitted distributions and the vertical axes are the ordered values of the derivative field of AEM data. Departure from the straight line indicates departure from the specified distribution. (a) Gaussian distribution; (b) Laplacian distribution; (c) Generalised Gaussian distribution.

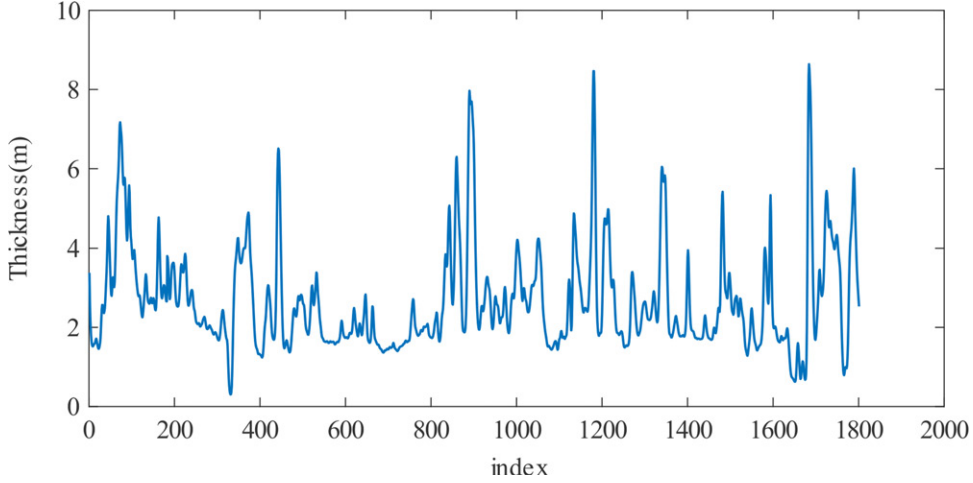


Fig. 4. The deformed FYI sequence of the AEM thickness data outlined in Fig. 1 with red rectangle. The data points are spaced at 7 m.

$$\begin{cases} J(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) + (\mathbf{y} - H(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{y} - H(\mathbf{x})) \\ \mathbf{x}_a = \operatorname{argmin}_{\mathbf{x}} J(\mathbf{x}) \end{cases} \quad (3)$$

In (3), $\mathbf{x} \in \mathbb{R}^n$ is the current best approximation to the true state, $\mathbf{x}_b \in \mathbb{R}^n$ is the background state, $\mathbf{y} \in \mathbb{R}^m$ is the observation vector and $H \in \mathbb{R}^{m \times n}$ is the observation operator that maps the state vector to the observation space. The background and observation error covariance matrices are given by $\mathbf{B} \in \mathbb{R}^{n \times n}$ and $\mathbf{R} \in \mathbb{R}^{m \times m}$, respectively, and the analysis state, $\mathbf{x}_a \in \mathbb{R}^n$ is the state that corresponds to when the objective function, $J(\mathbf{x})$ reaches a minimum.

3.2. Regularisation in variational data assimilation

In a study conducted by Johnson et al. (2005) the classic VDA problem was reformulated to the standard form of

Tikhonov regularisation (Tikhonov et al., 1977). Accordingly, by letting $\mathbf{B} = \sigma_b^2 \mathbf{C}_B$ and $\mathbf{R} = \sigma_o^2 \mathbf{C}_R$ where $\mathbf{C}_B$ and $\mathbf{C}_R$ are the error correlation matrices and σ_b^2 and σ_o^2 are the background and observation error variances, and using a change of variable $\mathbf{z} = \mathbf{C}_B^{-1/2} (\mathbf{x} - \mathbf{x}_b)$, the objective function in (3) can be reformulated as (Freitag et al., 2010)

$$J(\mathbf{z}) = \|\mathbf{C}_R^{-1/2} (\mathbf{y} - H\mathbf{x}_b) - \mathbf{C}_R^{-1/2} H \mathbf{C}_B^{1/2} \mathbf{z}\|_2^2 + \mu^2 \|\mathbf{z}\|_2^2, \quad (4)$$

where $\mu^2 = \sigma_o^2 / \sigma_b^2$ is the regularisation parameter, $\mu^2 \|\mathbf{z}\|_2^2$ is the regularisation term, and $\mathbf{H}$ is the linearised observation operator. If we define $\mathbf{f}$ and $\mathbf{G}$ as $\mathbf{f} = \mathbf{C}_R^{-1/2} (\mathbf{y} - H\mathbf{x}_b)$ and $\mathbf{G} = \mathbf{C}_R^{-1/2} H \mathbf{C}_B^{1/2}$, the objective function can be written as

$$J(\mathbf{z}) = \|\mathbf{f} - \mathbf{G}\mathbf{z}\|_2^2 + \mu^2 \|\mathbf{z}\|_2^2. \quad (5)$$

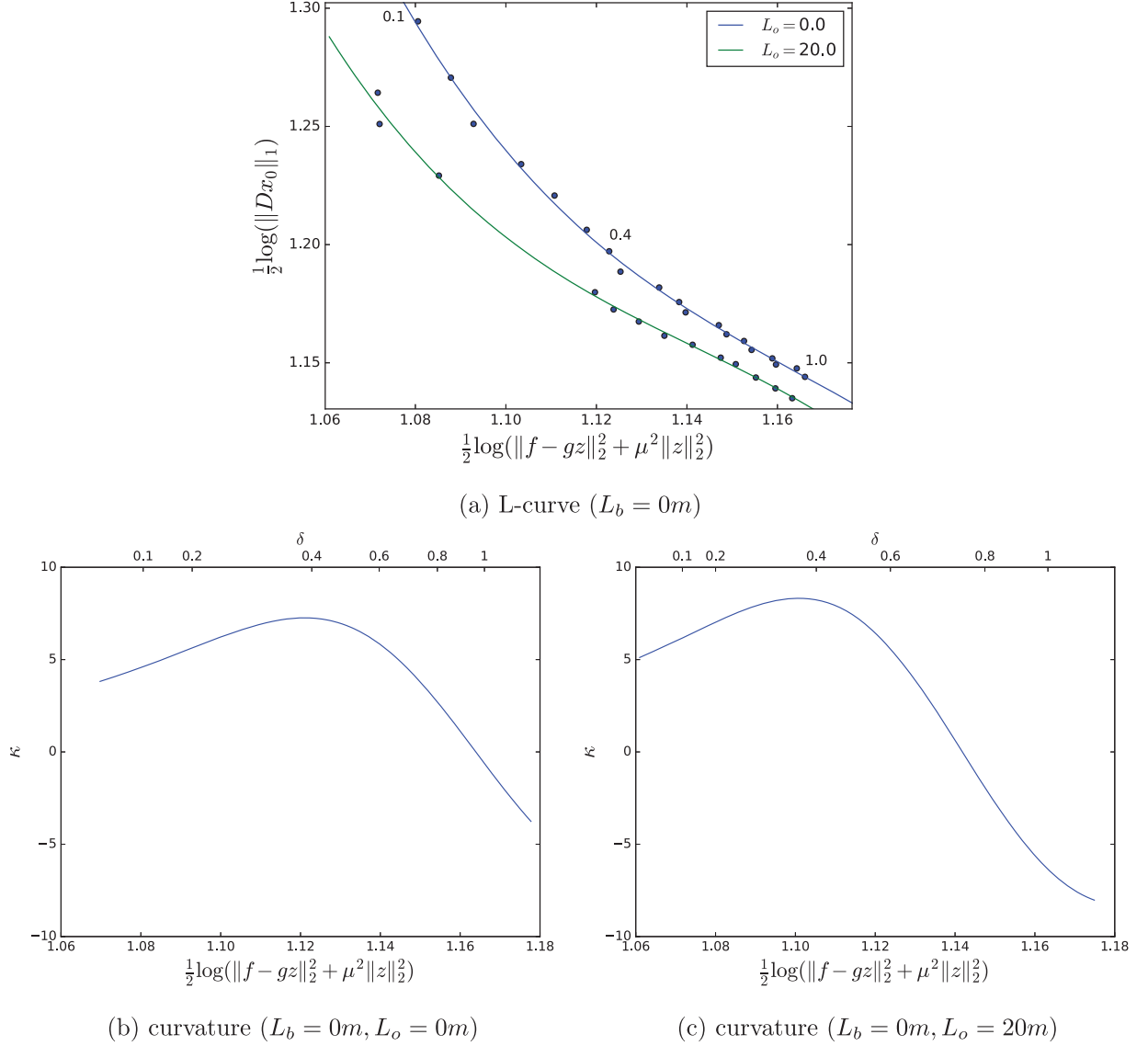


Fig. 5. Using maximum curvature of L-curve to find regularisation parameter, δ . Panel (a) shows L-curve of two cases for which $L_b = 0m$. In panel (b) $L_o = 0m$ and in panel (c) $L_o = 20m$. The top axis of panels (b) and (c) shows the corresponding δ values for curvature points.

Therefore, by solving

$$\mathbf{z}_a = \underset{\mathbf{z}}{\operatorname{argmin}} J(\mathbf{z}), \quad (6)$$

the analysis can be obtained as

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{C}_B^{1/2} \mathbf{z}_a. \quad (7)$$

We note the objective function in [Equations \(4\) and \(5\)](#) uses the scalar parameter μ , which assumes spatially homogeneous error variances. An alternative formulation that does not require this assumption is given in [Appendix A](#).

3.3. Mixed l_1 - l_2 -norm regularisation

Freitag et al. (2013) proposed their mixed regularisation framework with an objective function as

$$J(\mathbf{z}) = \|\mathbf{f} - \mathbf{G}\mathbf{z}\|_2^2 + \mu^2 \|\mathbf{z}\|_2^2 + \delta \|\mathbf{D}\mathbf{x}_0\|_1, \quad (8)$$

where $\mathbf{x}_0 = \mathbf{x}_b + \mathbf{C}_B^{1/2} \mathbf{z}$, $\|\cdot\|_1$ indicates l_1 -norm (Laplacian distribution), and $\delta > 0$ is another regularisation parameter and has to be chosen empirically. As δ increases, the gradient of the solution will be more sparse. Details about how to set this parameter are discussed in [Section 4.3](#).

In (8) the first order derivative can be approximated by applying matrix $D \in \mathbb{R}^{(n-1) \times n}$ to the state vector, where $\mathbf{D}$ can be written as

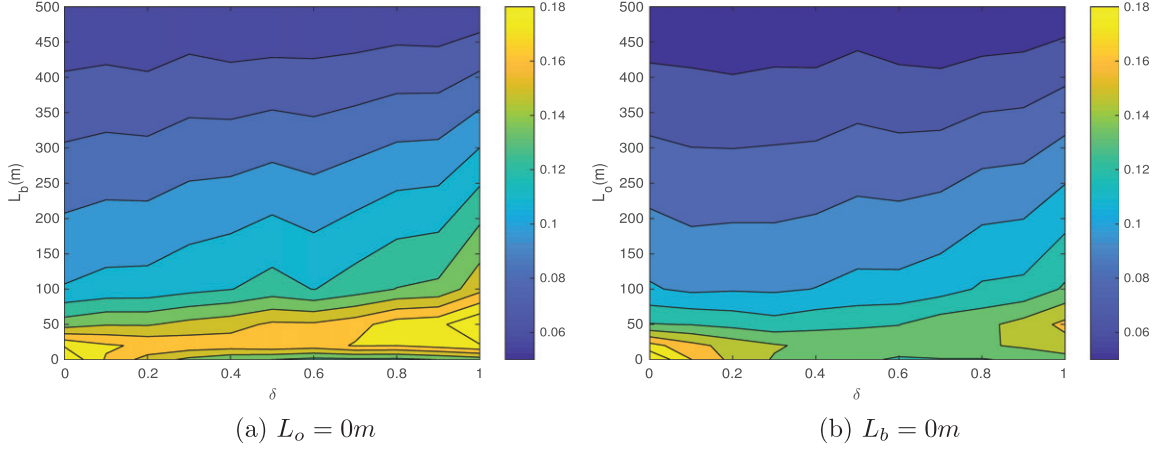


Fig. 6. Analysis RMSE of the data fusion experiment over a range of δ values when (a) $L_o = 0m$ and L_b is changing and (b) $L_b = 0m$ and for a range of L_o values.

$$\mathbf{D} = \begin{bmatrix} -1 & 1 & 0 & \dots & \\ 0 & -1 & 1 & 0 & \dots \\ & \ddots & \ddots & \ddots & \\ & & \dots & 0 & -1 & 1 \end{bmatrix}_{(n-1) \times n} \quad (9)$$

This operator corresponds to a first-order differencing, with equal spacing between data points. Other approximations to the first order derivative operator were tested, but the results were not found to be sensitive to this choice. This representation in Equation (9) is chosen for simplicity.

As discussed in Section 2 the ice thickness observations exhibit sparsity in the derivative domain and have a closer fit to both the Laplacian and generalised Gaussian distribution as compared with the Gaussian distribution. Even though the generalised Gaussian distribution with $p=0.8$ is the best fit for these observations, assuming this distribution for the regularisation results in a non-convex optimisation, which might not converge easily. Therefore, assuming the Laplacian distribution as the prior model for the gradient field is the approach taken in this study.

4. Experimental method

To evaluate the advantage of using the mixed regularisation framework, two different regularised VDA systems are evaluated. The first system, l_2 regularisation, minimises the objective function given by (5), while the second system, l_1 - l_2 regularisation, minimises the objective function described by (8). Note that the l_1 - l_2 regularisation method when $\delta=0$ is equivalent to the l_2 method. The minimisation problems are reformulated to a quadratic programming using an approach introduced by Fu et al. (2006) and solved by Python CVXOPT package (Andersen et al., 2013). These two systems were analysed by first carrying out a set of data fusion experiments,

where the analysis is not cycled through a model. The objective of the first set of experiments is to look at the impact of the mixed regularisation on the analyses, without the complications introduced by model physics and resampling of the AEM data to the model grid. The data fusion experiments are followed by a set of data assimilation experiments where a toy (1-D) sea ice model is used (Appendix B).

For both the data fusion and data assimilation experiments the background error covariance matrix is defined as

$$\mathbf{B} = \Sigma \mathbf{C}_B \Sigma \quad (10)$$

where Σ is a diagonal matrix with diagonal elements that are the background error standard deviation, σ_b , and $\mathbf{C}_B$ is the background error correlation matrix. The error correlation function is calculated based on the compactly supported fifth-order piecewise polynomial correlation function proposed by Gaspari and Cohn (1999). Note that in this formulation the error correlation function is an explicit function of the specified length scale and the spacing between points. The observation error covariance matrix, $\mathbf{R}$, is defined in the same way as the $\mathbf{B}$ matrix using error standard deviation σ_o and the same correlation function. Both the l_1 and l_1 - l_2 systems require calculation of the square root of the background error correlation matrix. This is a computationally expensive operation, that scales as $O(n^3)$, where n is the dimension of the problem. For this study, n is small ($n \approx 400$) and the square root is computed using Cholesky decomposition, which takes about 5% of the total time of an analysis cycle. For a larger, operational sea-ice data assimilation system, a method to compute the background error correlation matrix square root using a diffusion equation is described in Caya et al. (2010). A similar

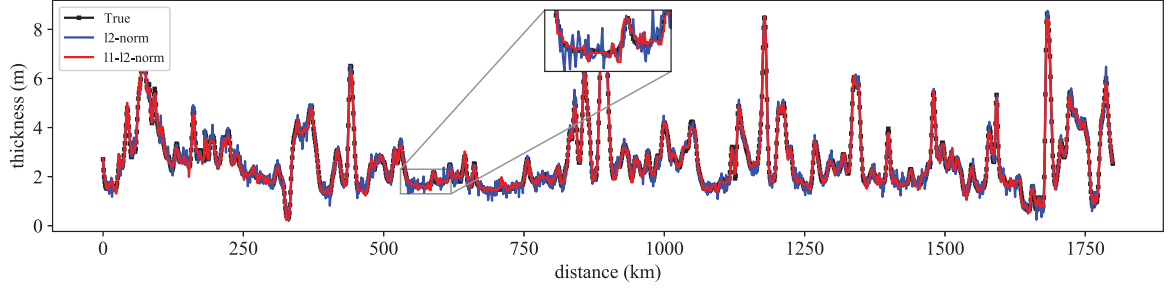
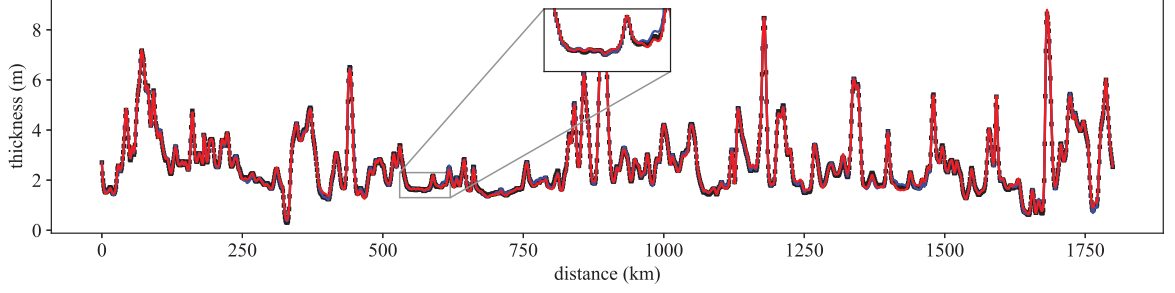
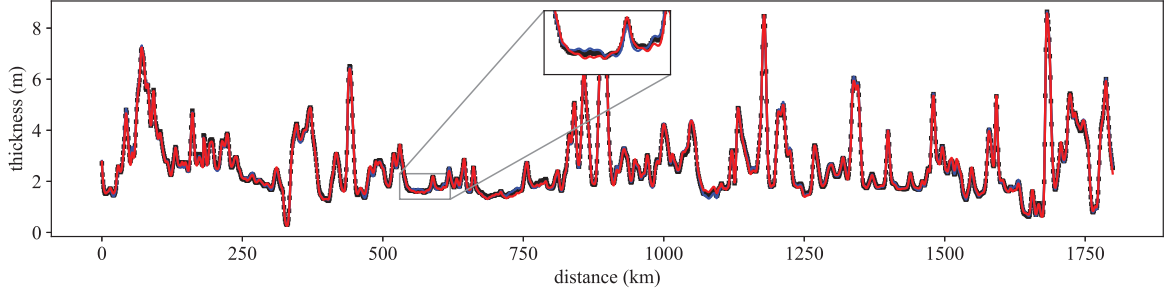
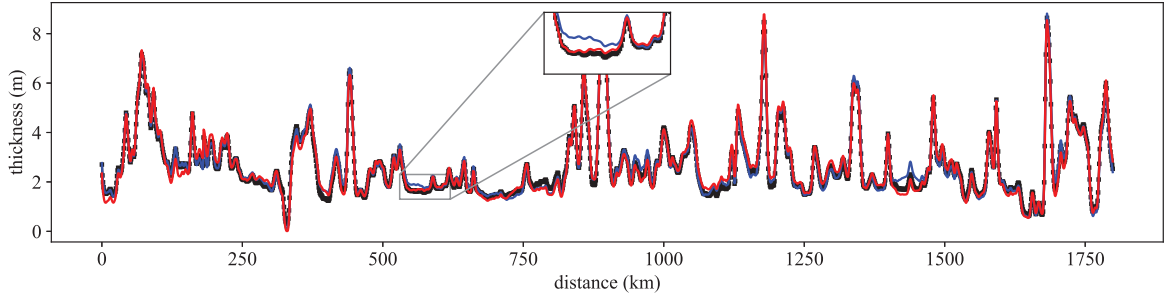
(a) $L_b = 0m, L_o = 0m$ (b) $L_b = 0m, L_o = 500m$ (c) $L_b = 500m, L_o = 0m$ (d) $L_b = 500m, L_o = 500m$

Fig. 7. Data fusion analysis states for l_2 -norm (blue), l_1 - l_2 -norm (red) and true state (black) for different background and observation error correlation length scales for $\delta = 0.4$. (a) $L_b = 0m, L_o = 0m$; (b) $L_b = 0m, L_o = 500m$; (c) $L_b = 500m, L_o = 0m$; (d) $L_b = 500m, L_o = 500m$.

method could be used for the l_1 - l_2 method described here. An additional, potentially expensive, part of the algorithm is the inversion of the square root of the

observation error correlation matrix. Large scale systems typically assume a diagonal observation error correlation matrix, which makes this operation trivial. Methods for

Table 2. Data fusion analysis errors for different background and observation error correlation length scales (L_b and L_o) when $\delta = 0.4$, $\text{KURT}[\mathbf{x}_i] = 3.12$ and $\text{KURT}[D\mathbf{x}_i] = 10.58$.

$L_b(m)$	$L_o(m)$	Regularisation method	MAE	RMSE	$\text{KURT}[\mathbf{x}_a]$	DIFF_MAE	DIFF_RMSE	$\text{KURT}[D\mathbf{x}_a]$
0	0	l_2 -norm	0.16	0.20	3.00	0.22	0.28	1.69
		l_1 - l_2 -norm	0.10	0.14	3.06	0.11	0.15	8.46
	20	l_2 -norm	0.14	0.18	3.02	0.15	0.19	3.8
		l_1 - l_2 -norm	0.11	0.14	3.08	0.09	0.12	8.67
	500	l_2 -norm	0.05	0.06	3.11	0.01	0.01	10.54
		l_1 - l_2 -norm	0.04	0.05	3.11	0.01	0.01	10.60
20	0	l_2 -norm	0.15	0.19	3.01	0.17	0.21	2.95
		l_1 - l_2 -norm	0.13	0.17	3.04	0.15	0.19	4.08
	20	l_2 -norm	0.16	0.20	2.94	0.15	0.18	3.71
		l_1 - l_2 -norm	0.14	0.18	2.95	0.13	0.16	5.11
	500	l_2 -norm	0.06	0.07	3.13	0.00	0.01	10.60
		l_1 - l_2 -norm	0.06	0.07	3.15	0.00	0.01	10.58
500	0	l_2 -norm	0.05	0.06	3.14	0.00	0.00	10.59
		l_1 - l_2 -norm	0.05	0.06	3.10	0.00	0.00	10.51
	20	l_2 -norm	0.06	0.08	3.08	0.00	0.00	10.55
		l_1 - l_2 -norm	0.06	0.08	3.08	0.01	0.00	10.60
	500	l_2 -norm	0.15	0.19	2.96	0.01	0.00	10.50
		l_1 - l_2 -norm	0.15	0.19	2.93	0.01	0.00	10.57

The best results are in bold font.

large scale systems utilising non-diagonal observation error correlation matrices are still a topic of research in the operational data assimilation community.

The accuracy of the analyses is evaluated using six measures: (1) mean absolute error (MAE), (2) root mean squared error (RMSE), (3) kurtosis, $\text{KURT}[\mathbf{x}_a] = \frac{E[(\mathbf{x} - E[\mathbf{x}])^4]}{(E[(\mathbf{x} - E[\mathbf{x}])^2])^2}$, where $E[\cdot]$ represents the expectation operator, (4) MAE of the derivative field (DIFF_MAE) (5) RMSE of the derivative field (DIFF_RMSE) and (6) kurtosis of the derivative field ($\text{KURT}[D\mathbf{x}_a]$). Kurtosis of the state is used to show tail of the state's probability distribution so that, an analysis with high kurtosis has retained more sharp features. Reported results are averaged over 40 simulations where the additive random noise for background and observation states are different for each simulation. Note that the background and observation error standard deviation ratio (μ) was held constant at $\mu = 1$ for all experiments. Details specific to the data fusion and data assimilation experiments are now given.

4.1. Data fusion experiments

The deformed FYI subset of the AEM data shown in Fig. 1 is selected as the true state ($\mathbf{x}_t \in \mathbb{R}^n$ where $n = 1800$). This sample of the data, shown in Fig. 4, was found to include more sharp features as compared to the smooth FYI sample. The background ($\mathbf{x}_b \in \mathbb{R}^n$) and observations ($\mathbf{y} \in \mathbb{R}^n$), consist of the true state with additive random noise sampled from a Gaussian distribution with mean zero and covariance matrix $\mathbf{B}$ and $\mathbf{R}$,

respectively. For both background and observation error covariance matrices, experiments were carried out over a range of correlation length scales, measured in meters, corresponding to $L = \{0, 20, 50, 100, 500\}$ and a range of δ values. It was found the values of the error measures (given below) increased with increasing δ for $\delta > 1$. Hence the performance of the l_1 - l_2 -norm regularisation method is reported for a range of δ values between 0 and 1.

4.2. Data assimilation experiments

For the data assimilation experiments the initial true state was based on the entire set of AEM ice thickness measurements acquired on 20 April, shown in Fig. 1b. The data points were averaged over segments of 1 km in length. This was done to obtain a state at a spatial resolution closer to that of interest in operational sea ice forecasting (Carrieres et al., 2017; Batrak and Müller, 2018; Mohammadi-Aragh et al., 2018), in addition to one where the continuum assumption used in the ice model should be valid Flato (1998). When this true state was used to initialise the sea ice model it was found the ice did not move very significantly over the 72 h period. This was likely because the ice concentration was close to 100% over the entire domain and the model was run with thermodynamics off to simplify analysis of the results. This means the ice concentration can only change by openings in the ice cover, which may be difficult to obtain due to the high ice pressure when the ice concentration is close to 100% and the lack of solid boundaries

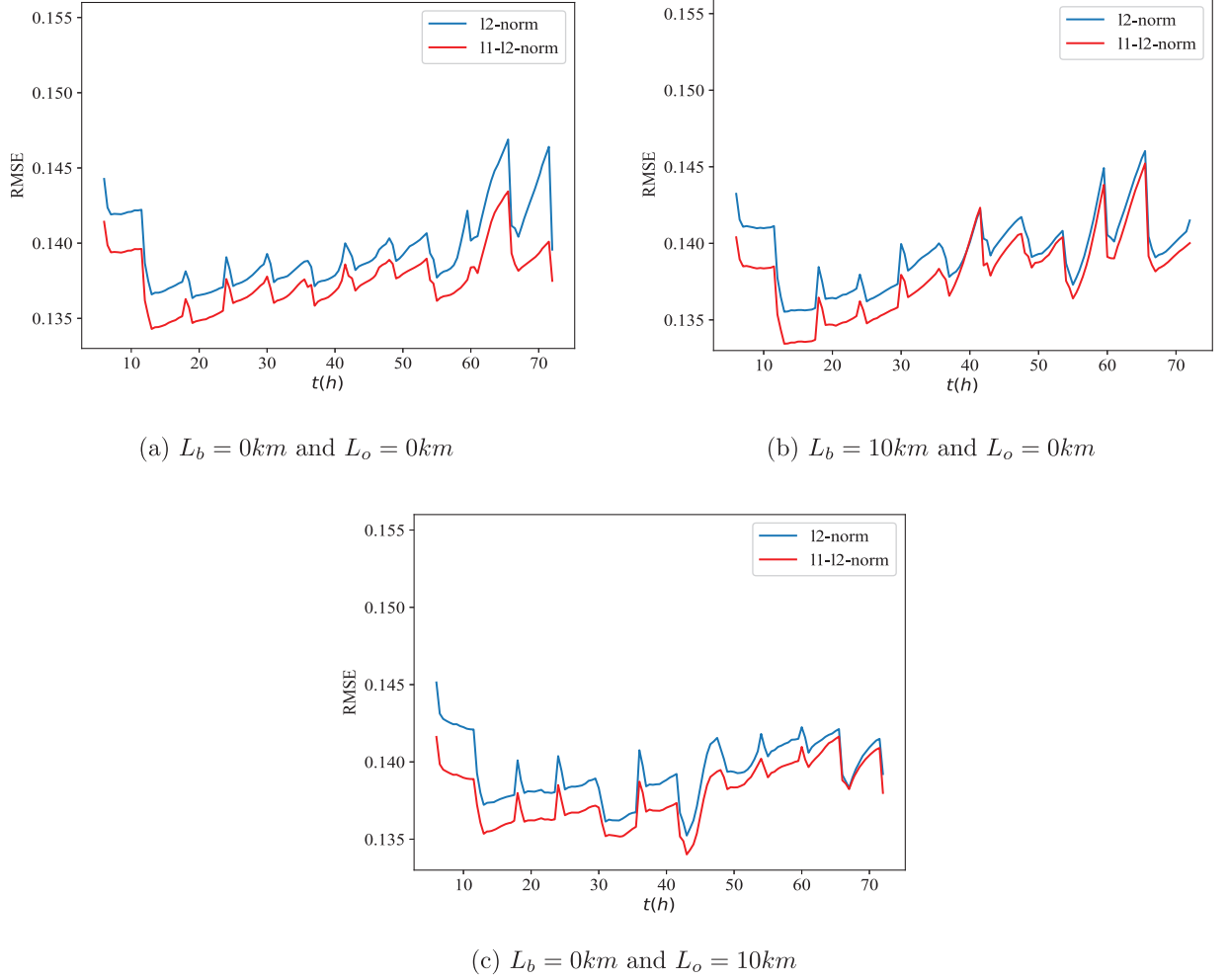


Fig. 8. RMSE of the data assimilation ice thickness [m] states when (a) $L_b = 0\text{ km}$ and $L_o = 0\text{ km}$, (b) $L_b = 10\text{ km}$ and $L_o = 0\text{ km}$ and (c) $L_b = 0\text{ km}$ and $L_o = 10\text{ km}$ for the 72 h data assimilation experiment. Data are assimilated every 6 h.

in the domain. To create open water regions in the ice cover the thickness everywhere was reduced by 0.5 m and the points with negative thickness were set to zero. This state was used to initialise the ice model which was run forward in time to provide true states at $t > 0$, referred to in this paper as the true model run (Appendix B). While this state differs from the actual truth in part due to model error, such an approach should be sufficient for evaluating the difference between l_2 and l_1 - l_2 regularisation methods. Each of the experiments (truth, l_1 - l_2 and l_2) was run with identical atmosphere and ocean states, which provide the forcing to the sea ice model.

The data assimilation experiments were initialised with a background state consisting of $\mathbf{x}_b = \mathbf{x}_t + \varepsilon_b$ where ε_b is a sample generated from the $\mathbf{B}$ matrix described in the previous section. Every 6 h the model state was updated by assimilating a set of observations generated by perturbing

the current state of the true model run at that time according to $\mathbf{y} = \mathbf{x}_t + \varepsilon_o$ where ε_o is a sample generated from the $\mathbf{R}$ matrix described in Section 4. Data assimilation experiments were carried out for background and observation correlation length scales corresponding to 0 and 10 km.

4.3. Choice of regularisation parameter

Identifying an appropriate value of the regularisation parameter is an important task in a regularisation problem. The method used to choose the parameter is typically problem-dependent. There are several methods that find the value of the regularisation parameter assuming that the norm of the noise in observations is known (Golub and Von Matt, 1997; Calvetti et al., 2000). Conversely, there are approaches that do not need the observation noise to be known explicitly (Akaike, 1980; Wahba, 1990).

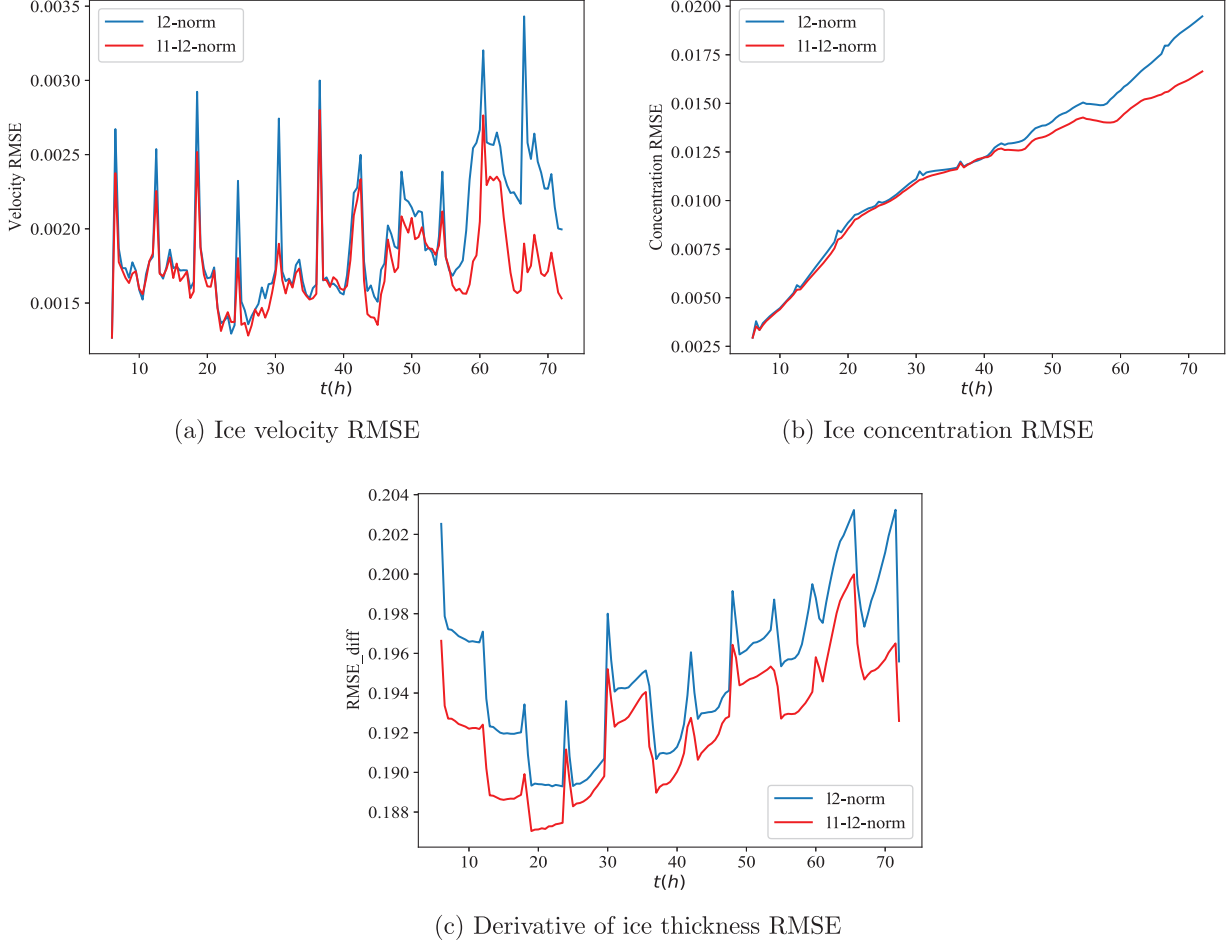


Fig. 9. RMSE of the (a) ice velocity, (b) ice concentration and (c) ice thickness derivative when $L_b = 0$ km and $L_o = 0$ km for the 72 h data assimilation experiment. Data are assimilation every 6 h.

Here, we follow one of the latter approaches, known as the L-curve method. In this method it is assumed that the objective function can be written in the form $\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_2^2$ with regularisation parameter λ . The L-curve is the log-log plot of the norm of the regularised solution, $\|\mathbf{x}_a\|_2^2$, versus the norm of the corresponding residual vector, $\|\mathbf{y} - \mathbf{H}\mathbf{x}_a\|_2^2$, evaluated at the minimisation solution, $\mathbf{x}_a$. In this method, the value corresponding to the point of maximum curvature in the obtained L-shape curve is selected as the optimal regularisation parameter (Hansen, 1992; Hansen and O’Leary, 1993; Lawson and Hanson, 1995) because at this point there is a balance between the norm of the solution vector and the fit to the observations.

For the problem studied here, a different choice of axes is made for the L-curve due to the fact that the objective function has both l_1 and l_2 terms. Recalling the minimisation problem in (8), the corresponding L-curve is defined as the plot of $\frac{1}{2} \log \|D\mathbf{x}_0\|_1$ versus $\frac{1}{2} \log (\|\mathbf{f} - \mathbf{G}\mathbf{z}\|_2^2 + \mu^2 \|\mathbf{z}\|_2^2)$. In this case, the point of maximum curvature

represents the balance between minimising the l_1 -norm of the solution in derivative space, and the l_2 -norm for the Tikkonov regularisation problem.

The L-curve is shown in Fig. 5a for $L_b = 0$ m and for two different observation error correlation length scales, $L_o = 0$ m, $L_o = 20$ m where the analysis used to evaluate the functions comprising the L-curve were obtained following the data fusion experimental set-up described in Section 4.2. The corresponding curvature plots are shown in Fig. 5b,c. Based these results, it was decided $\delta \approx 0.4$ must be chosen for the l_1 - l_2 regularisation method for the data fusion experiments, since this value corresponds to the maximum curvature in the L-curve.

Since in these experiments we have a known true state, the validity of the L-curve results could be verified by comparing the analysis states resulting from different δ values. Using the RMSE between the true state (taken as the data from the deformed FYI in Fig. 1) and the analyses as the error measure (see Section 5.1 for further details). Two sets of experiments were carried out. In the

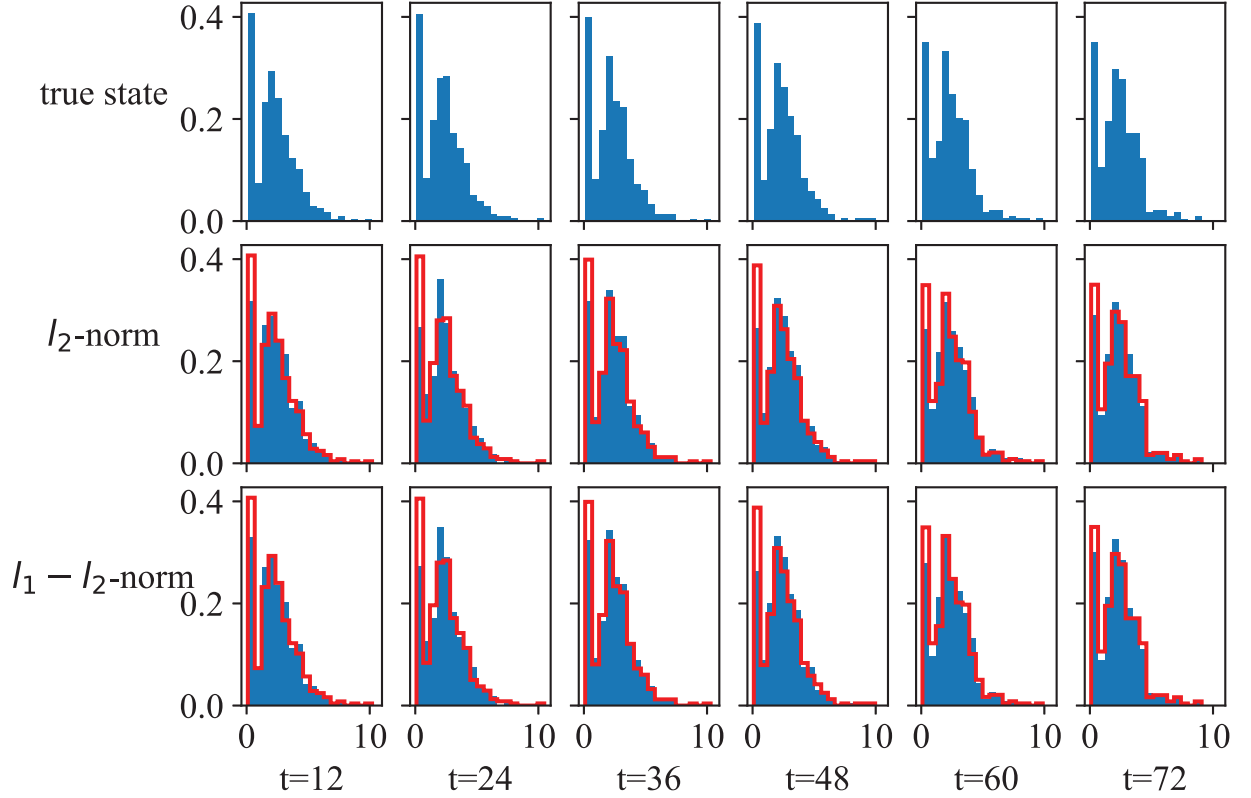


Fig. 10. Histograms of the analysis states (ice thickness [m]) from the data assimilation when $L_b = 0$ km and $L_o = 0$ km. Each column shows the analysis state at the indicated time. The top row shows the true model, and the second and third rows show l_2 -norm and l_1 - l_2 -norm results, respectively. Histogram from the true state is overlayed by red colour in the second and third rows.

first set, the observation error covariance matrix was diagonal ($L_o = 0$ m) while L_b was varied over the range of 0 m–500 m, and in the second set, the background error covariance matrix was diagonal ($L_b = 0$ m) and L_o was varied over the range 0 m–500 m. Note that the spacing of the analysis grid is 7 m, hence these length scales range from 0 to 70 times the analysis grid spacing. Results are represented in Fig. 6. In both cases the benefit of using the l_1 - l_2 approach as compared to l_2 can be seen for the shorter length scales (i.e. when L_b or L_o is lower than 50 m) for δ values in the range of $0.3 < \delta < 0.7$, consistent with the results from the L-curve.

For the data assimilation experiment, similar to the data fusion experiments, δ values of $\delta = 0.01, 0.05, 0.08, 0.1, 0.2, 0.4, 0.5$ were tested. The optimal δ value was found to be 0.05, hence results using this value are shown here.

5. Results

5.1. Results from data fusion experiments

The analysis for the data fusion experiments corresponding to the case when both background and observation

errors are uncorrelated is shown in Fig. 7a. It can be seen the l_1 - l_2 analysis is less noisy than that from l_2 and the sharp features in the true state are captured more accurately. The analysis for the case when the background error is uncorrelated and $L_o = 500$ m is represented in Fig. 7b. Again, in Fig. 7b, the analysis from the l_1 - l_2 method is less noisy than that from the l_2 method and the sharp features are captured more accurately. However, a benefit is no longer visibly apparent when the background error correlation length scale is $L_b = 500$ m, as in Fig. 7c,d.

Table 2 represents the analysis errors measures described in Section 5 when $\delta = 0.4$. Note the kurtosis of the true state, $\text{KURT}[\mathbf{x}_t]$, and its derivative, $\text{KURT}[D\mathbf{x}_t]$, were 3.12 and 10.58, respectively. Kurtosis is higher when the tails of the probability distribution are fatter which can be interpreted as sharp features in the state when kurtosis is computed on the derivative field of the state. If kurtosis of the analysis derivative is less than the true state derivative kurtosis, it shows that sharp features have been damped. Results presented in Table 2 confirms that when the background error is uncorrelated, the l_1 - l_2 -norm method outperforms the l_2 -norm method in all terms of error measures. This is similar to what has been

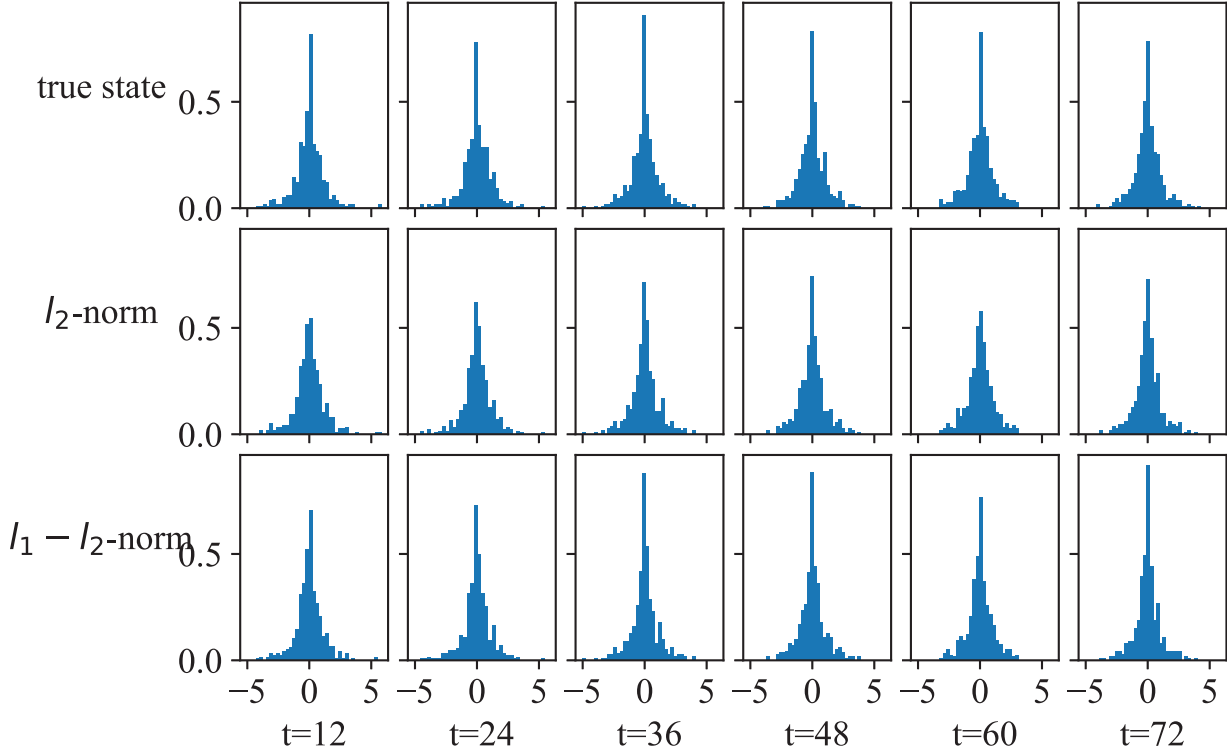


Fig. 11. Histograms of the derivative of analysis states (ice thickness derivatives $[-]$) from data assimilation when $L_b = 0$ km and $L_o = 0$ km. Each column shows a different time step of the model. The top row shows the true model, and the second and third rows are showing l_2 -norm and l_1 - l_2 -norm results, respectively.

illustrated in Fig. 7. This improvement for the MAE and RMSE measures is about 25% while the improvement is more than 40% for the ice thickness derivative fields (DIFF_MAE and DIFF_RMSE). In addition, the kurtosis shows that when the background error is uncorrelated, the l_1 - l_2 method retains sharp features in the analysis better than l_2 method. When the observation error length scale increases, the kurtosis of both methods approaches the true state however, the l_1 - l_2 method is always better. When $L_b = 20$ m, the l_1 - l_2 method is slightly better than the l_2 method regardless of the observation error correlation length scale. However, when the background error correlation length scale is $L_b = 500$ m, the resulting analysis states are relatively smooth and DIFF_MAE and DIFF_RMSE are almost zero for all cases. Finally, it should be noted that while the kurtosis of the analysis states for both methods are similar, in all cases kurtosis of the analysis derivative is always better for the l_1 - l_2 -norm method, which shows the ability of this method in retaining the sharp features in ice thickness.

5.2. Results from data assimilation experiments

The root-mean-squared (RMS) ice thickness errors from the data assimilation experiments are shown in Fig. 8 for

various combinations of background and observation error correlation length scales. In all cases the errors are reduced when l_1 - l_2 is used as compared with l_2 . Time traces of the data assimilation states (see Figs. 13 and 14) showed that this growth in errors, which is also seen in the ice concentration and ice velocity (Fig. 9) is related to the development of openings in the ice cover, which are better captured with l_1 - l_2 than l_2 .

Histograms of the analysis states (ice thickness) and their derivative fields at each analysis time are presented in Figs. 10 and 11, respectively for the case where $L_b = 0$ km and $L_o = 0$ km. In both figures, the first row illustrates the histogram of the true state. In addition, for better comparison, the histogram of the true state is overlayed on the histograms of the data assimilation method. In Fig. 10 (top row) it is difficult to see substantial differences between the ice thickness histograms, although the derivative fields in Fig. 11, show that for the l_1 - l_2 method, there is a higher level of sparsity on the derivative field of analysis as compared to the l_2 method. This is consistent with the expected results, since the presence of the l_1 term on the ice thickness derivative should result in more zeros in this field.

A representative result from the data assimilation experiments is shown in Fig. 12. This result is obtained at

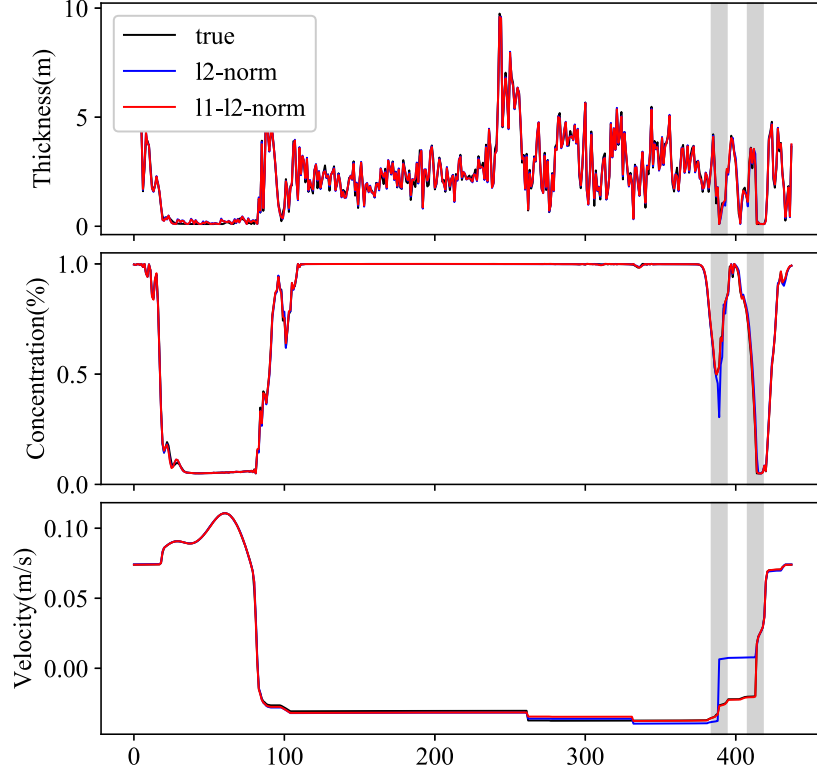


Fig. 12. An example of data assimilation states at $t=64.0(h)$ when $L_b = 0 \text{ km}$ and $L_o = 0 \text{ km}$. The shaded regions indicate the locations at which time traces (shown in Figs. 13 and 14) are taken.

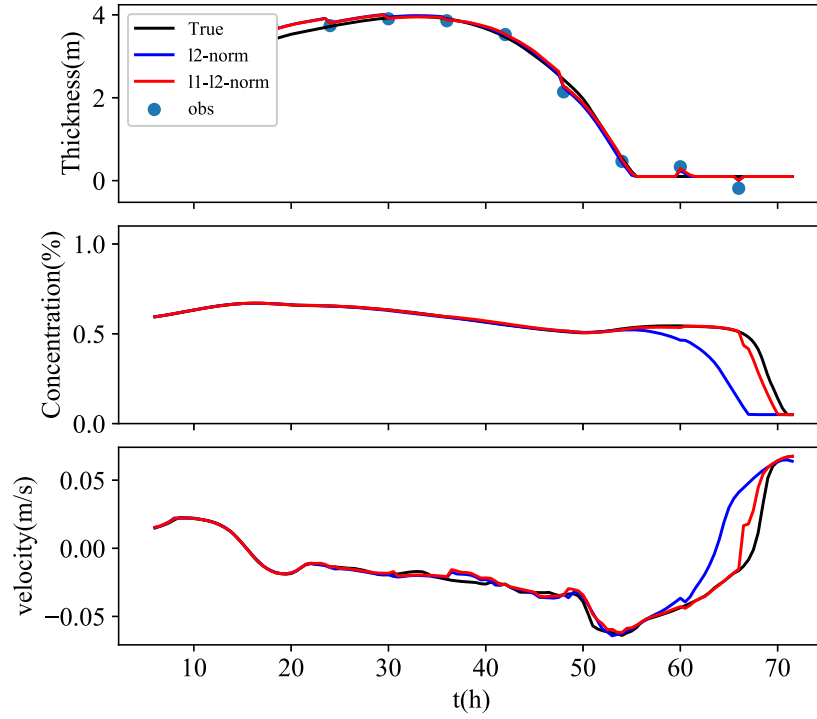


Fig. 13. Model states from a single realisation of the data assimilation experiment for $L_b = 0 \text{ km}$ and $L_o = 0 \text{ km}$ at index = 389 (indicated by vertical line in Fig. 12). The decrease in ice concentration and thickness is overestimated for l_2 .

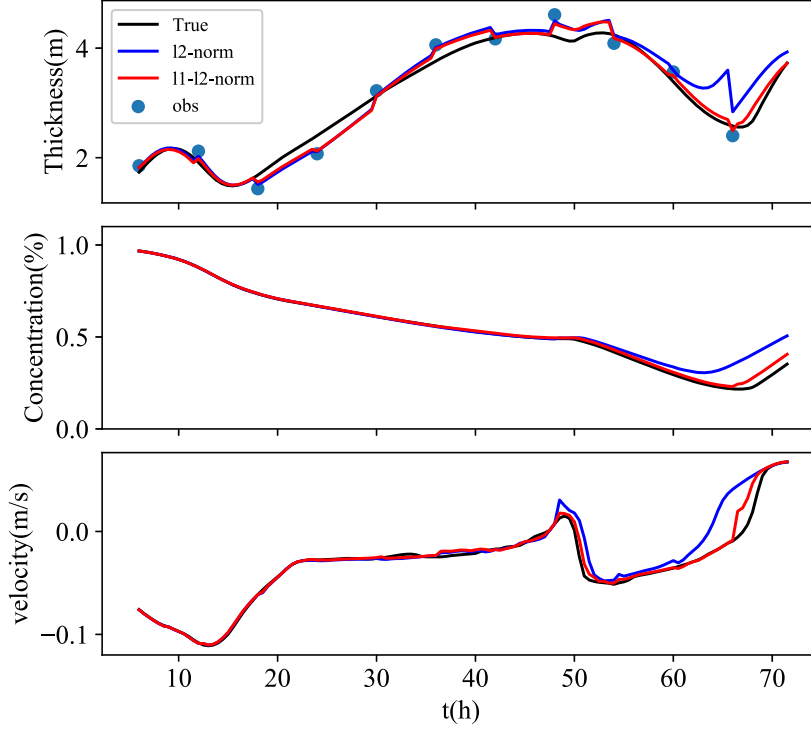


Fig. 14. Model states from a single realisation of the data assimilation experiment for $L_b = 0$ km and $L_o = 0$ km at index = 245 (indicated by vertical line in Fig. 12). The opening in the ice cover is predicted better for the l_1 - l_2 method than for the l_2 method.

$t = 64$ h for $L_b = 0$ km and $L_o = 0$ km. It can be seen in Fig. 12 that the reduction in ice concentration at spatial index 389 is captured more accurately for l_1 - l_2 as compared to l_2 . The time trace of model states for this spatial index is shown in Fig. 13. Note the states are in agreement when the ice is thick and the concentration is high, but differences develop as the ice thins and the concentration drops. A similar result is shown in Fig. 14, where the reduction in ice concentration and thickness, is better captured with the l_1 - l_2 regularisation. Note also that both time traces show the ice velocity is also in better agreement with the true state for l_1 - l_2 .

Similar results were found for other background and observation error length scales and also for the case when the ice cover was consolidated (although these events were observed less frequently for the consolidated ice cover). In general, openings in the ice cover and related changes in the ice velocity are better captured with l_1 - l_2 than l_2 . This may be due to the relationship between ice thickness, ice strength and bulk viscosity, and the role this plays in the momentum equation (see Appendix B). The exact details of this interaction are complicated and will depend on the chosen parameters and numerical methods used. This will be investigated in a future study.

6. Concluding remarks

This study has demonstrated for the first time that sea ice thickness exhibits a sparse representation in the derivative domain. This has been demonstrated using (1) sea ice thickness measurements from an AEM sensor over the Beaufort Sea; (2) submarine upward looking sonar data; and (3) sea ice thickness from Cryosat. To retain this feature when using sea ice thickness data in a data fusion or data assimilation scheme, the use of an additional term in the objective function to constrain sparsity on the derivative of the ice thickness state, is evaluated. This l_1 - l_2 formulation is compared with the standard l_2 regularisation first using data fusion, and then by carrying out data assimilation experiments using a toy sea-ice model.

For data fusion a clear benefit to the l_1 - l_2 formulation is seen when the background correlation error length scale is small (on the order of twice the analysis grid spacing). Based on previous studies (Caya et al., 2010), it can be expected that in the vicinity of a sharp feature (e.g. ice edge) the background error correlation length scale may be in this range. This data fusion result could be relevant for the generation of merged sea ice products, where sharp features are desired (see for example the ice concentration used at the lower boundary in (Batrak and Müller, 2018)).

For data assimilation a clear benefit is also seen to the l_1 – l_2 regularisation, although the impact of the error correlation length scales on the difference between the l_1 – l_2 method and the l_2 method is less clear. This may be due to the spatial averaging of ice thickness that was required to increase the scale of the data, or it could be due to the model dynamics. For example, changes in the ice thickness will impact the bulk viscosity in the momentum equation, which would interact with the numerical methods used in the model. This will be investigated in a later study. However, based on the preliminary results shown here, it can be noted that the l_1 – l_2 method is superior with regards to capturing openings in the ice cover than the conventional l_2 method. This was observed for a variety of error correlation length scales, values of the regularisation parameter (μ), and model initial conditions.

Acknowledgements

The authors would like to thank Mark Buehner and Jean-Francois Lemieux for helpful discussions, and Christian Haas for the use of the AEM data. The authors would also like to thank the reviewers whose constructive comments greatly improved the quality of this article.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work has been funded by Environment and Climate Change Canada, and the Marine Environmental Observation Prediction and Response Network Centre of Excellence.

Note

1. <http://www.cpom.ucl.ac.uk/csopr/seaice.html>

References

- Akaike, H. 1980. Likelihood and the Bayes procedure. *Trab. Estadística Invest. Oper.* **31**, 143–166. doi:10.1007/BF02888350.
- Andersen, M. S., Dahl, J. and Vandenberghe, L. 2013. Cvxopt: A python package for convex optimization, version 1.1. 6. Online at: cvxopt.org.
- Auclair, J.-P., Lemieux, J.-F., Tremblay, B. and Ritchie, H. 2017. Implementation of Newton's method with an analytical Jacobian to solve the 1d sea ice momentum equation. *J. Comput. Phys.* **340**, 69–84. doi:10.1016/j.jcp.2017.02.065.
- Batrak, Y. and Müller, M. 2018. Atmospheric response to kilometer-scale changes in sea ice concentration within the marginal ice zone. *Geophys. Res. Lett.* **45**, 6702–6709.
- Budd, C., Freitag, M. A. and Nichols, N. 2011. Regularization techniques for ill-posed inverse problems in data assimilation. *Comput. Fluids* **46**, 168–173. doi:10.1016/j.compfluid.2010.10.002.
- Calvetti, D., Reichel, L., Sgallari, F. and Spaletta, G. 2000. A regularizing Lanczos iteration method for underdetermined linear systems. *J. Comput. Appl. Math.* **115**, 101–120. doi:10.1016/S0377-0427(99)00298-8.
- Carrieres, T., Buehner, M., Lemieux, J.-F. and Pedersen, L. 2017. *Sea Ice Analysis and Forecasting*. Cambridge University Press, Cambridge, UK.
- Caya, A., Buehner, M. and Carrieres, T. 2010. Analysis and forecasting of sea ice conditions with three-dimensional variational data assimilation and a coupled ice-ocean model. *J. Atmos. Oceanic Technol.* **27**, 353–369. doi:10.1175/2009JTECHO701.1.
- Chambers, J. M. 1983. *Graphical Methods for Data Analysis*. CRC Press, Boca Raton, FL.
- Daley, R. 1993. *Atmospheric Data Analysis. Number 2*. Cambridge University Press, Cambridge, UK.
- Ebtehaj, A. M. and Foufoula-Georgiou, E. 2013. On variational downscaling, fusion, and assimilation of hydrometeorological states: A unified framework via regularization. *Water Resour. Res.* **49**, 5944–5963.
- Ebtehaj, A. M., Foufoula-Georgiou, E., Lerman, G. and Bras, R. L. 2015. Compressive earth observatory: An insight from AIRS/AMSU retrievals. *Geophys. Res. Lett.* **42**, 362–369.
- Ebtehaj, A. M., Zupanski, M., Lerman, G. and Foufoula-Georgiou, E. 2014. Variational data assimilation via sparse regularisation. *Tellus Dyn. Meteorol. Oceanogr.* **66**, 21789. doi:10.3402/tellusa.v66.21789.
- Flato, G. M. 1998. The thickness variable in sea-ice models. *Atmos. Ocean* **36**, 29–36. doi:10.1080/07055900.1998.9649604.
- Foufoula-Georgiou, E., Ebtehaj, A., Zhang, S. and Hou, A. 2014. Downscaling satellite precipitation with emphasis on extremes: A variational l_1 -norm regularization in the derivative domain. *Surv. Geophys.* **35**, 765–783. doi:10.1007/s10712-013-9264-9.
- Freedman, D. and Diaconis, P. 1981. On the histogram as a density estimator: L_2 theory. *Z. Wahrscheinlichkeitstheorie Verw. Gebiete* **57**, 453–476. doi:10.1007/BF01025868.
- Freitag, M. A., Nichols, N. K. and Budd, C. J. 2010. L_1 -regularisation for ill-posed problems in variational data assimilation. *Proc. Appl. Math. Mech.* **10**, 665–668. doi:10.1002/pamm.201010324.
- Freitag, M., Nichols, N. and Budd, C. 2013. Resolution of sharp fronts in the presence of model error in variational data assimilation. *Q. J. Royal Meteorol. Soc.* **139**, 742–757. doi:10.1002/qj.2002.
- Fu, H., Ng, M. K., Nikolova, M. and Barlow, J. L. 2006. Efficient minimization methods of mixed l_2 – l_1 and l_1 – l_1

- norms for image restoration. *SIAM J. Sci. Comput.* **27**, 1881–1902. doi:10.1137/040615079.
- Gaspari, G. and Cohn, S. E. 1999. Construction of correlation functions in two and three dimensions. *Q. J. Royal Meteorol. Soc.* **125**, 723–757. doi:10.1002/qj.49712555417.
- Girard, L., Bouillon, S., Weiss, J., Amitrano, D., Fichet, T. and co-authors. 2011. A new modeling framework for sea-ice mechanics based on elasto-brittle rheology. *Ann. Glaciol.* **52**, 123–132. doi:10.3189/172756411795931499.
- Golub, G. H. and Von Matt, U. 1997. *Tikhonov regularization for large scale problems*. Workshop on Scientific Computing, Springer, New York, pp. 3–26.
- Haas, C., Hendricks, S., Eicken, H. and Herber, A. 2010. Synoptic airborne thickness surveys reveal state of Arctic sea ice cover. *Geophys. Res. Lett.* **37**, L09501.
- Haas, C. and Howell, S. E. 2015. Ice thickness in the Northwest Passage. *Geophys. Res. Lett.* **42**, 7673–7680.
- Haas, C., Lobach, J., Hendricks, S., Rabenstein, L. and Pfaffling, A. 2009. Helicopter-borne measurements of sea ice thickness, using a small and lightweight, digital EM system. *J. Appl. Geophys.* **67**, 234–241. doi:10.1016/j.jappgeo.2008.05.005.
- Hansen, P. C. 1992. Analysis of discrete ill-posed problems by means of the l-curve. *SIAM Rev.* **34**, 561–580. doi:10.1137/1034115.
- Hansen, P. C., Nagy, J. G. and O’Leary, D. P. 2006. *Deblurring Images: matrices, spectra, and Filtering*, Vol. 3. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.
- Hansen, P. C. and O’Leary, D. P. 1993. The use of the l-curve in the regularization of discrete ill-posed problems. *SIAM J. Sci. Comput.* **14**, 1487–1503. doi:10.1137/0914086.
- Hibler, W. D. 1979. A dynamic thermodynamic sea ice model. *J. Phys. Oceanogr.* **9**, 815–846. doi:10.1175/1520-0485(1979)009<0815:ADTSIM>2.0.CO;2.
- Johnson, C., Nichols, N. and Hoskins, B. 2005. Very large inverse problems in atmosphere and ocean modelling. *Int. J. Numer. Meth. Fluids* **47**, 759–771. doi:10.1002/flid.869.
- König Beatty, C. and Holland, D. M. 2010. Modeling landfast sea ice by adding tensile strength. *J. Phys. Oceanogr.* **40**, 185–198. doi:10.1175/2009JPO4105.1.
- Kubat, I., Sayed, M. and Babaei, M. H. 2013. Analysis of besetting incidents in Frobisher Bay during 2012 shipping season. In: *Proceedings of the International Conference on Port and Ocean Engineering Under Arctic Conditions*, Sweden, POAC.
- Kullback, S. and Leibler, R. A. 1951. On information and sufficiency. *Ann. Math. Statist.* **22**, 79–86. doi:10.1214/aoms/1177729694.
- Lahoz, W., Khattatov, B. and Menard, R. 2010. Data assimilation and information. In: *Data Assimilation: Making Sense of Observations* (eds. W. Lahoz, B. Khattatov, and R. Menard). Springer, Berlin Heidelberg, pp. 3–12.
- Lawson, C. L. and Hanson, R. J. 1995. *Solving Least Squares Problems*. SIAM, Philadelphia, PA.
- Laxon, S. W., Giles, K. A., Ridout, A. L., Wingham, D. J., Willatt, R. and co-authors. 2013. Cryosat-2 estimates of Arctic sea ice thickness and volume. *Geophys. Res. Lett.* **40**, 732–737.
- Lemeux, J.-F., Knoll, D. A., Losch, M. and Girard, C. 2014. A second-order accurate in time Implicit-EXplicit (IMEX) integration scheme for sea ice dynamics. *J. Comput. Phys.* **263**, 375–392. doi:10.1016/j.jcp.2014.01.010.
- Lemeux, J., Tremblay, B., Thomas, S., Sedláček, J. and Mysak, L. 2008. Using the preconditioned Generalized Minimum RESidual (GRMES) method to solve the sea-ice momentum equation. *J. Geophys. Res.* **113**, 1–12.
- Lipscomb, W. H., Hunke, E. C., Maslowski, W. and Jakacki, J. 2007. Ridging, strength, and stability in high-resolution sea ice models. *J. Geophys. Res. Oceans* **112**, C03S91.
- Loptien, U. and Axell, L. 2014. Ice and AIS: ship speed data and sea ice forecasts in the Baltic Sea. *Cryosphere* **8**, 2409–2418. doi:10.5194/tc-8-2409-2014.
- Lorenc, A. C. 1986. Analysis methods for numerical weather prediction. *Q. J. Royal Meteorol. Soc.* **112**, 1177–1194. doi:10.1002/qj.49711247414.
- Meier, W., Hovelsrud, G., van Oort, B., Key, J., Kovaks, K. and co-authors. 2014. Arctic sea ice in transformation: A review of recent and observed changes and impacts on biology and human activity. *Rev. Geophys.* **51**, 185–217.
- Mohammadi-Aragh, M., Goessling, H., Losch, M., Hutter, N. and Jung, T. 2018. Predictability of Arctic sea ice on weather time scales. *Sci. Rep.* **8**, 6514. doi:10.1038/s41598-018-24660-0.
- NSIDC (National Snow and Ice Data Center). 1998. Submarine upward looking sonar ice draft profile data and statistics, version 1. doi:10.7265/N54Q7RWK.
- Pizzolato, L., Howell, S. E. L., Dawson, J., Laliberté, F. and Copland, L. 2016. The influence of declining sea ice on shipping activity in the Canadian Arctic. *Geophys. Res. Lett.* **43**, 12,146–12,154. 2016GL071489.
- Rainwater, S., Bishop, C. and Campbell, W. 2015. The benefits of correlated observation errors for small scales. *Q. J. Royal Meteorol. Soc.* **141**, 3439–3445. doi:10.1002/qj.2582.
- Renner, A. H., Gerland, S., Haas, C., Spreen, G., Beckers, J. F. and co-authors. 2014. Evidence of Arctic sea ice thinning from direct observations. *Geophys. Res. Lett.* **41**, 5029–5036.
- Rothrock, D. and Thorndike, A. 1980. Geometric properties of the underside of sea ice. *J. Geophys. Res.* **85**, 3955–3963.
- Stonebridge, G., Scott, K. A. and Buehner, M. 2018. Impacts on sea ice analyses from the assumption of uncorrelated ice thickness observation errors: Experiments using a 1D toy model. *Tellus Dyn. Meteorol. Oceanogr.* **70**, 1445379. doi:10.1080/16000870.2018.1445379.
- Sturges, H. A. 1926. The choice of a class interval. *J. Am. Statistic. Assoc.* **21**, 65–66. doi:10.1080/01621459.1926.10502161.
- Thomas, M., Geiger, C. and Kambhamettu, C. 2008. High resolution (400 m) motion characterization of sea ice using ERS-1 SAR imagery. *Cold Reg. Sci. Technol.* **52**, 207–223. doi:10.1016/j.coldregions.2007.06.006.
- Tikhonov, A. N., Arsenin, V. I. and John, F. 1977. *Solutions of Ill-Posed Problems*, Vol. 14. Winston, Washington, DC.
- Wahba, G. 1990. *Spline Models for Observational Data*. SIAM, Philadelphia, PA.
- Wahba, G. and Wendelberger, J. 1980. Some new mathematical methods for variational objective analysis using splines and

cross validation. *Mon. Wea. Rev.* **108**, 1122–1143. doi: 10.1175/1520-0493(1980)108<1122:SNMMFV>2.0.CO;2.

Appendix A

Generalised objective function

In Section 3, the objective functions in Equations (5) and (8) are formulated assuming the background and observation error variances are spatially homogeneous. A more general formulation allowing nonhomogeneous error variances is explained in this Appendix.

We start by considering the l_2 objective function given earlier in Equation (3),

$$J(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_b)^T \mathbf{B}^{-1} (\mathbf{x} - \mathbf{x}_b) + (\mathbf{y} - H(\mathbf{x}))^T \mathbf{R}^{-1} (\mathbf{y} - H(\mathbf{x})) \quad (\text{A.1})$$

Following Caya et al. (2010) we introduce a change of variable $\xi = \mathbf{B}^{-\frac{1}{2}}(\mathbf{x} - \mathbf{x}_b)$, which means $\mathbf{x} = \mathbf{x}_b + \mathbf{B}^{\frac{1}{2}}\xi$. By substituting $\mathbf{x}$ into J we arrive at the following objective function:

$$J(\xi) = \xi^T \xi + \|\mathbf{R}^{-\frac{1}{2}}(\mathbf{y} - H(\mathbf{x}_b)) - \mathbf{R}^{-\frac{1}{2}}\mathbf{H}\mathbf{B}^{\frac{1}{2}}\xi\|_2^2 \quad (\text{A.2})$$

By defining $\mathbf{f} = \mathbf{R}^{-\frac{1}{2}}(\mathbf{y} - H(\mathbf{x}_b))$ and $\mathbf{G} = \mathbf{R}^{-\frac{1}{2}}\mathbf{H}\mathbf{B}^{\frac{1}{2}}$ we can write Equation (A.2) as

$$J(\xi) = \|\mathbf{f} - \mathbf{G}\xi\|_2^2 + \|\xi\|_2^2, \quad (\text{A.3})$$

For the l_1 – l_2 -norm method an objective function can be defined that is the same as that for l_2 but with the additional regularisation term,

$$J(\xi) = \|\mathbf{f} - \mathbf{G}\xi\|_2^2 + \|\xi\|_2^2 + \delta \|\mathbf{D}\mathbf{x}_0\|_1. \quad (\text{A.4})$$

This objective function has the same form as that used in the present study, although optimal values of δ would be different since the scaling of the problem is different.

Appendix B

Toy sea ice model

The sea ice model used in the present study is based on the model introduced by Hibler (1979), and uses a viscous-plastic (VP) rheology. Although some authors have recently proposed new sea ice rheologies (e.g. Girard et al. (2011)), almost all current sea ice models are based on the VP rheology. With the VP constitutive law, the ice behaves as a very viscous fluid when the state of stress is inside the yield curve while it deforms plastically when the state of stress reaches the yield curve. The viscous coefficients are capped when the ice is in the viscous regime. The minimum delta value is $2 \times 10^{-9} \text{ s}^{-1}$. This value is so small that the ice is basically modelled as an ideal plastic material.

In 2-D, the standard VP elliptical yield curve leads to strong resistance in compression, significant resistance in shear and small resistance in tension. As in Lipscomb et al. (2007); König Beatty and Holland (2010); Lemieux et al. (2014); Auclair et al. (2017), the 1-D model is obtained by assuming that the v component of the ice velocity and spatial gradients $\partial/\partial y$ are zero. The 1-D model is a simplification of the 2-D model as one can not simulate failure in shear. However, it still allows one to simulate and study interesting aspects of the VP rheology such as free drift, landfast ice and failures in compression and tension.

The toy sea ice model used in this study solves a momentum equation that can be written:

$$\rho_{ice} a h \frac{\partial u}{\partial t} = \rho_a a C_a u_a |u_a| + \rho_o a C_o (u_o - u) |u_o - u| + \frac{\partial}{\partial x} \left((\zeta + \eta) \frac{\partial u}{\partial x} - \frac{P}{2} \right), \quad (\text{B.1})$$

where ρ_{ice} represents the ice density, C_a represents the ice-atmosphere drag coefficient, u_a represents the wind velocity, C_o represents the ice-ocean drag coefficient, u_o represents the ocean current velocity and ρ_a , ρ_o are the air of air and water, respectively. The velocity of the water and air were obtained from shallow water models that are coupled to the ice model. The variables in the ice stress term, which is a 1-D version of the stress tensor defined in Hibler (1979) are

$$\begin{aligned} \zeta &= \frac{P}{2\Delta} \\ \eta &= \zeta e^{-2} \\ \Delta &= \left| \frac{\partial u}{\partial x} \right| (1 + e^{-2}) \\ P &= P^* h \exp(-C(1 - a)), \end{aligned} \quad (\text{B.2})$$

where P represents the ice strength, and P^* , C , and e are empirically-derived constants. In addition to the ice momentum equation, there are transport equations for ice thickness and concentration (Hibler, 1979):

$$\begin{aligned} \frac{\partial a}{\partial t} &= -\frac{\partial u a}{\partial x} + S_a \\ \frac{\partial h}{\partial t} &= -\frac{\partial u h}{\partial x} + S_h \end{aligned} \quad (\text{B.3})$$

where S_a and S_h are thermodynamic terms governing the ice growth. For the current application both S_a and S_h are zero.

The sea ice equations were discretised using a central differencing scheme with velocity grid points at locations staggered from the thickness, concentration and viscosity. Atmospheric and oceanic boundary conditions were provided by one-layer shallow water models, tuned to produce qualitatively realistic velocity fields. Periodic

boundary conditions were used for the sea ice momentum, concentration and thickness equations as well as for the forcing. Values used for the constants are $P^* = 8000 N/m$, $C_a = 0.0015$, $C_w = 0.0015$, $e = 2$, $\rho_a = 1.3 \text{ kg/m}^3$ and $\rho_i = 910 \text{ kg/m}^3$ and $\rho_o = 1035 \text{ kg/m}^3$. Note that ζ was constrained to be within the range of

$\zeta_{min} = 4 \times 10^8 \text{ kgs}^{-1}$ and $\zeta_{max} = 2.8 \times 10^8 P$ to avoid singularities that can occur for small Δ values and reduce numerical instabilities that can occur for small viscosities (Hibler, 1979; Lemeiux et al., 2008). For further details about the ice model see (Stonebridge et al., 2018).