Simultaneous inference of sea ice state and surface emissivity model using machine learning and data assimilation

Alan Geer[1]

[1]European Centre for Medium-range Weather Forecasts

Key Points:

- A new hybrid of machine learning and data assimilation can infer an unknown state and an unknown model simultaneously.
- This empirical state method could help other chicken and egg earth system problems that are not suitable for supervised machine learning.
- This work finds a new sea ice surface emissivity model and makes global all-season maps of sea ice concentration using inverse methods.

Corresponding author: Alan Geer, `alan.geer@ecmwf.int`

Abstract

Satellite microwave radiance observations are strongly sensitive to sea ice, but physical descriptions of the radiative transfer of sea ice and snow are incomplete. Further, the radiative transfer is controlled by poorly-known microstructural properties that vary strongly in time and space. A consequence is that surface-sensitive microwave observations are not assimilated over sea ice areas, and sea ice retrievals use heuristic rather than physical methods. An empirical model for sea ice radiative transfer would be helpful but it cannot be trained using standard machine learning techniques because the inputs are mostly unknown. The solution is to simultaneously train the empirical model and a set of empirical inputs: an "empirical state" method, which draws on both generative machine learning and physical data assimilation methodology. A hybrid physical-empirical network describes the known and unknown physics of sea ice and atmospheric radiative transfer. The network is then trained to fit a year of radiance observations from Advanced Microwave Scanning Radiometer 2 (AMSR2), using the atmospheric profiles, skin temperature and ocean water emissivity taken from a weather forecasting system. This process estimates maps of the daily sea ice concentration while also learning an empirical model for the sea ice emissivity. The model learns to define its own empirical input space along with daily maps of these empirical inputs. These maps represent the otherwise unknown microstructural properties of the sea ice and snow that affect the radiative transfer. This "empirical state" approach could be used to solve many other problems of earth system data assimilation.

Plain Language Summary

One way to learn about the earth system would be through machine learning, but typical 'supervised learning' already requires good knowledge of the geophysical variables of interest. Quite often this 'geophysical state' is not well known and the main observations are from satellites, which measure earth-emitted radiation. In many cases the links between this observed radiation and the geophysical state are poorly known. This work illustrates a new method which allows both the geophysical state and its links to the satellite observations to be learnt at the same time. The specific application is finding the sea ice concentration, which is a difficult problem because there is poor knowledge of the properties of sea ice, such as the air bubbles within it, which can strongly affect the radiation measured by satellites, and poor knowledge of exactly how this affects the satellite observations. The new approach solves this double problem by making daily maps of the sea ice and its properties, along with learning a new model to simulate the satellite measured radiation from the sea ice properties. Similar approaches could improve knowledge of both models and geophysical state in other areas of earth system science.

1 Introduction

Huge progress has been made in earth system observation and prediction using physically-based methods (Bauer et al., 2015). Underpinning this is data assimilation, which combines a physical model, which propagates the geophysical state forwards in time, with observations, which are used to synchronize the evolving geophysical state with reality. For global observational coverage, we rely on direct assimilation of satellite radiance observations (e.g. Eyre et al., 2020) which require an 'observation model' to make the link between the radiances and the geophysical state (e.g. Gettelman et al., 2022). In the early development of radiance assimilation, the focus was on temperature-sounding radiances in clear sky conditions, where physical models for state and observations were most accurate. More recently, physical models have become good enough to allow significant impact on the quality of forecasts through the assimilation of radiances sensitive to humidity, cloud and precipitation (e.g. Peubey & McNally, 2009; Geer et al., 2017, 2018). To further improve atmospheric predictability and to give a better representation of surface

parameters, a new focus is to improve the use of models and observations of the ocean, land surface and cryosphere (de Rosnay et al., 2022). However, particularly when it comes to the land surface and cryosphere, neither the modelling of the state, nor the use of satellite observations, is accurate enough to be able to assimilate satellite radiances using entirely physical approaches.

The aim of this work is to build tools and techniques for direct assimilation of satellite microwave radiances for the cryosphere and in particular, the sea ice. This work is ultimately intended for use in an operational weather forecasting system, to allow the assimilation of microwave radiances over sea ice surfaces and to contribute to analyses of the sea ice concentration. Microwave radiances have a strong sensitivity to sea ice, particularly its fractional coverage but also ice age, snow depth and snow microstructure. Operational methods to retrieve the sea ice concentration use largely heuristic ('tie point') approaches (e.g. Comiso et al., 2003; Spreen et al., 2008; Sandven et al., 2023) and an experimental method using optimal estimation used a simplified forward model (e.g. Scarlat et al., 2020). This is because physical modelling of snow and ice radiative transfer at microwave frequencies is difficult, with better results below 20 GHz than above (e.g. Hirahara et al., 2020). A fundamental issue is the importance of centimetre to micron scale details of the snow and ice, including air inclusions in ice or grain shapes and their layout in snow, that are relevant to the radiative transfer. Physical models of snow radiative transfer (e.g. Picard et al., 2018) are starting to provide convincing results up to around 250 GHz over land, for example linking measurements from snow pit profiles to aircraft radiance measurements with reasonable accuracy in late winter in the Canadian Arctic (Sandells et al., 2023). Physical models for sea ice have also shown potential (Kang et al., 2023), though with the necessity of empirically adjusting the snow and sea ice optical properties to fit each location. For global assimilation of snow and ice surfaces, observation models will need to be reliable across all seasons and hemispheres, and will need to handle many different types of ice and snow, including wet and refrozen snow in the warmer seasons. There is also the problem of defining the input state, since it is impossible to provide snow pits and ice cores globally.

An alternative path to improving our modelling frameworks is to use the vast amount of existing observational data to improve models in an empirical way (e.g. Schneider et al., 2017; Geer, 2021; Gettelman et al., 2022). The idea of training models from observations has become more feasible in recent years following the advent of easy-to-use machine learning tools such as Keras (Chollet et al., 2015) and rapid progress has been made in machine learning alternatives to weather forecasting models (e.g. Lam et al., 2022). A partly empirical approach is taken in the current work, with the aim to train a new observation model for sea ice using global data through the whole year, in order to handle all seasons, hemispheres and snow and ice types. However, empirical geophysical models are normally built using 'supervised learning' strategies that require the inputs and outputs of the empirical model to be known and available in large quantities. For example, in the case of ML-based weather forecasting models, this relies on existing atmospheric re-analysis datasets (e.g. Hersbach et al., 2020). But to train an empirical observation model to link the geophysical state to the satellite radiances for the land surface or the cryosphere, the inputs, in other words the state of the soil, snow and ice, are not well enough known on a global scale. In fact, we will likely only know the state parameters well enough on a global scale once an observation model is available to help derive them from satellite observations, so training an empirical model for surface radiative transfer can be seen as a chicken and egg problem.

If supervised learning strategies are not possible, alternative 'unsupervised learning' strategies are known in the wider field of machine learning. An example is generative adversarial networks, which in the geophysical world have been used to generate plausible snowflake shapes (Leinonen et al., 2021). The forward model in a generative adversarial network defines its own 'latent space' of empirical input variables which de-

scribe underlying statistical properties of, for example, the output snowflake shapes. A similar principle is used in the current work, which allows the empirical observation model to define its own latent space as its input, in order to represent the mostly unknown microphysical structures of sea ice and snow that are needed to simulate microwave radiances. This will be referred to as an 'empirical state', which will be trained simultaneously with the empirical model to go from the state to the observations. The meaning of the empirical state variables is learnt along with the empirical model.

However, a purely empirical framework is of no use if we want to infer geophysical properties, not empirical variables. To solve this, we can impose known physical relationships, such as the radiative transfer of the atmosphere. Reichstein et al. (2019) described how empirical models could be used and trained in networks with physical models. These networks can be represented in a Bayesian way that is mathematically equivalent to data assimilation (Geer, 2021). The Bayesian viewpoint brings the insight that physical models represent parts of the network where there is good prior physical knowledge, whereas machine learning models can represent areas where there is little or no prior knowledge. In the current work, the empirical modelling is limited to the surface emissivity of the sea ice, whereas physical modelling is used for the surface emissivity of the ocean and for the radiative transfer of the atmosphere. In particular, it is the inclusion of physical equations that allows the sea ice concentration to be derived within the system, but the equations would have been incomplete without the empirical contribution. The entire network can be referred to as a hybrid empirical-physical model.

A number of other aspects of geophysical modelling and observation are more well-developed in data assimilation frameworks than in typical machine learning methods (Geer, 2021). One of these is mapping between geophysical fields on a regular time and space grid to the irregular arrangement of satellite observations in time and space, which is handled by an interpolation operator. Also used is background error modelling (to control the weight of prior physical knowledge against the new knowledge from observations) and observation errors (to account for the varying difficulty of forward modelling at different frequencies measured by the satellite). All these techniques are also employed in the current work, making a mix of data assimilation and machine learning, following much work merging the two fields (e.g. Hsieh & Tang, 1998; Bocquet et al., 2020; Brajard et al., 2020).

The framework created to solve these problems is introduced in the methods section (Sec. 2), which starts with an overview of the framework, followed up by subsections on the different scientific and technical aspects. The results (Sec. 3) are presented for a version of this framework which learns daily sea ice maps for the Arctic and Antarctic for an entire year at 40 km spatial resolution, along with the new hybrid surface emissivity model and daily maps of the empirical properties of the sea ice and snow. Since the framework was created through a rapid prototyping approach, its design decisions are explored after the fact using sensitivity tests which are described in the appendices. Both the training dataset and the Python-based framework are also available to complete the documentation of this work.

2 Method

2.1 Overview

Figure 1 shows a simplified overview of the hybrid physical-empirical network that describes radiative transfer over sea ice and polar ocean at microwave frequencies. The aim of this work is to find daily maps of the sea ice concentration $\mathbf{C}_{\text{ice}}$ and its physical state, $\mathbf{X}_{\text{ice}}$ simultaneously with a new empirical model for the sea ice surface emissivity (neural network weights $\mathbf{w}$), in order to generate simulated observations $\mathbf{y}$ to best fit real AMSR2 observations. The trainable variables are given dotted outlines on the

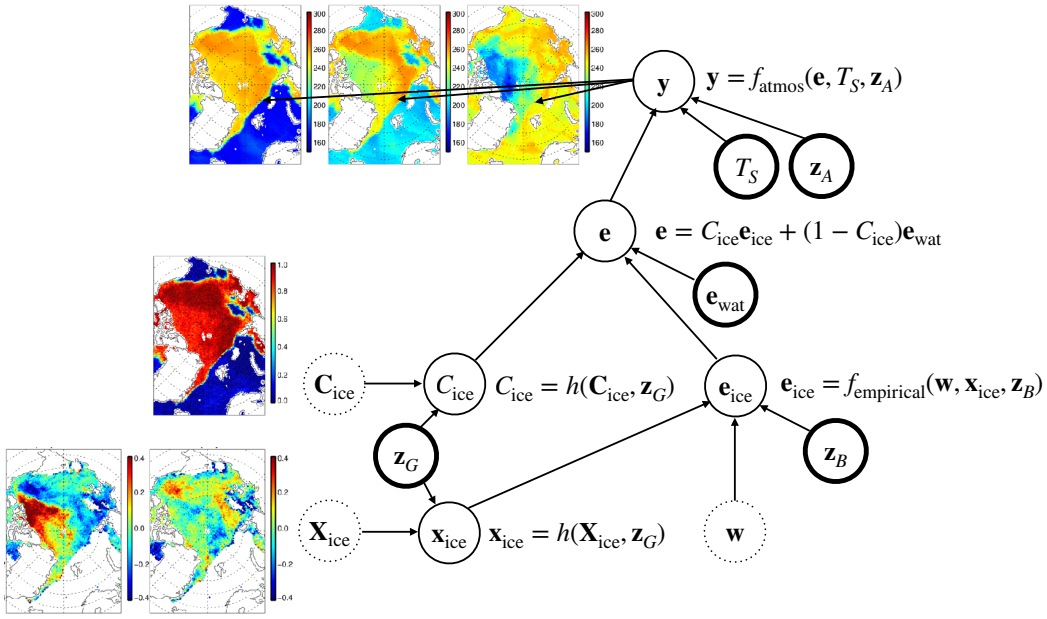

Figure 1. Simplified overview of the hybrid empirical-physical training network, for a single day and a single multi-channel observation. Circles represent variables that are trainable (dotted), dependent (thin solid), or fixed (thick solid). Arrows indicate dependencies and equations give the functional form of these dependencies. The meaning of the variables is explained in the text. Colour maps illustrate, for the Arctic, the observations (top); sea ice concentration (middle) and empirical sea ice properties (bottom) for 7th November 2020; full size extended versions, along with full explanatory details, can be found later in Figs. 2, 9 and 6 respectively.

figure. The design and purpose of the network, and the meanings of other variables in the figure, will be overviewed in the current section, with further detail provided in the relevant subsections that follow.

The network is implemented in Keras (Chollet et al., 2015), part of Tensorflow (Abadi et al., 2015), using Python. The inputs to the network are fixed parameters shown with bold outlines on Fig. 1, such as the observation locations $\mathbf{z}_{\text{G}}$ and estimated state of the atmosphere from the European Centre for Medium-range Weather Forecasts (ECMWF) model $\mathbf{z}_{\text{A}}$. The physical and empirical relations between parameters are described by equations with dependent output variables, indicated by thin solid outlines on the figure. These equations are encoded in custom network layers, so that parameters of the network can be defined as weights (akin to neural network weights) within these custom layers, which may be trainable or fixed.

Training the network can be seen as either a data assimilation process, or roughly equivalently, the training of a generative machine learning framework, that aims to best fit the observations by learning the geophysical state at each location and day over an entire year. The network is not intended to directly predict the observations for another year because the sea ice locations and surface properties will be different. Those parts of the network would have to be re-trained to fit another year. This means there is no possibility using separate training and test datasets, as is typical in standard supervised machine learning practice. Instead, the performance of the network is judged by how well it fits the observations, and how well it estimates the geophysical state (e.g. the sea ice concentration). Out of training set performance has been demonstrated through the use

of the trained hybrid surface emissivity model to fit another year's observations in a weather forecasting context (Geer, 2023c).

The hybrid network is trained by minimising a loss function. The aim is to be able to replicate as closely as possible a year of Advanced Microwave Sounding Radiometer-2 (AMSR2) observations over ocean and sea ice areas poleward of 45° latitude. The simulated observation $\mathbf{y}$ (Fig. 1) represents an observation at one location comprising measurements at different frequencies and polarisations, known as channels. The notation here is that variables in lower-case bold represent vectors. The measured quantity is radiance and this is given in terms of equivalent black body brightness temperature (TB). Further detail on the microwave observations is in Sec. 22.3. The observed AMSR2 observations are $\mathbf{y}_{\rm obs}$ and the discrepancy with the simulations is measured by a mean squared error loss function $J_{\rm obs}$, which weights observations by an observation error $\mathbf{r}$ following standard DA practice:

$$J_{\rm obs} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m}\frac{(y_{{\rm obs},ij} - y_{ij})^2}{r_j^2} \tag{1}$$

Here, i is the index over all observations, with $n \simeq 600$ million of them covering the full year. For simplicity the observation index has not been shown in Fig. 1. The elements of vectors $\mathbf{y}_i$ and $\mathbf{r}$ are scalars y_{ij} and r_j with the channel index j. Note that the division by n is not typically done in DA (e.g. Geer, 2021) and instead reflects the way batch-based loss terms in Keras are accumulated as a per-observation average. As in data assimilation, this is not the only loss function used to constrain the results of the network, but a number of other losses are used for constraining the estimated parameters (similar to background error constraints in data assimilation) and are described in Sec. 2.2.

The physical details of the network are now briefly described by following it back from its outputs. The observed AMSR2 radiances are affected by the atmospheric absorption and scattering from gases and clouds, as well as the emission, scattering and reflection of the surface. Hence a physical description of the atmospheric radiative transfer is encapsulated in $\mathbf{y} = f_{\rm atmos}(\mathbf{e}, T_S, \mathbf{z}_A)$, where the surface inputs are surface temperature T_S and emissivity vector $\mathbf{e}$ (one element for each channel) and the atmospheric state is represented in the vector $\mathbf{z}_A$. The atmosphere and skin temperature are treated as known and fixed and are extracted from the ECMWF 12 h background forecast. Further details of the atmospheric radiative transfer are in Sec. 2.4. One detail omitted from Fig. 1 is the use of a trainable bias correction for the observations, which needs 20 trainable weights; however in practice this is mostly constrained to prior knowledge and is not a major aspect of the problem (Sec. 2.3).

The mixed surface emissivity $\mathbf{e}$ must be estimated within the network to provide the surface boundary conditions to the atmospheric radiative transfer model. The emissivity is described as a linear combination of the sea ice and ocean water surface emissivities $\mathbf{e}_{\rm ice}$ and $\mathbf{e}_{\rm wat}$, weighted by the sea ice concentration in the satellite's field of view, $C_{\rm ice}$. The contrast in the surface emissivity between ocean and ice is strongest and easiest to detect at microwave wavelengths, and is fundamentally what allows the inference of the sea ice concentration from from satellite observations. This equation also defines the meaning of sea ice concentration in this work, which is the fraction of the ocean surface covered with ice or snow, which excludes melt ponds but can include ice bergs and ice shelves. The ocean surface emissivity is described physically and is treated as a known quantity. Not included in the simplified diagram is a windspeed error correction that is applied to the ocean surface emissivity, which uses 10 trainable weights. This correction is not a major aspect of the problem and is further described in Sec. 2.5.

The sea ice concentration and its emissivity are the key unknowns in this work. To describe the latter, the network includes a time-evolving geographical map of sea ice concentration $\mathbf{C}_{\rm ice}$. Here the capitalised bold notation denotes a geographical map, possibly with multiple layers, although for the sea ice concentration there is just one. The in-

terpolation operator which goes from the map to the observation location is represented here as $h(\mathbf{C}_{\text{ice}}, \mathbf{z}_G)$. The position of the observation on the earth's surface and the time it was taken, in other words its geolocation, is represented by $\mathbf{z}_G$. The geographical map of sea ice is treated as an unknown and is hence one of the trainable variables, shown on Fig. 1 with a dotted outline.

The sea ice emissivity $\mathbf{e}_{\text{ice}}$ is described by an empirical function $f_{\text{empirical}}(\mathbf{w}, \mathbf{x}_{\text{ice}}, \mathbf{z}_B)$ where $\mathbf{w}$ are the unknown parameters of the function (e.g. trainable neural network weights), and $\mathbf{x}_{\text{ice}}$ are unknown empirical inputs describing the unknown micro and macro-physical properties of the sea ice and any snow lying on top of it. Finally, $\mathbf{z}_B$ represents any known inputs for the surface emissivity model. In the current work this just contains a function of the skin temperature, but in future it could be extended to include the relevant outputs of a sea ice and snow physical model, as these become more accurate. In the current work it has proven sufficient to represent $f_{\text{empirical}}()$ using a single dense linear neural network layer with 50 trainable weights. Multilayer and nonlinear ('deep') neural networks have also been tried out, but their complexity is unnecessary here and also they increase the possibilities for over-fitting (Appendix A1). The weights $\mathbf{w}$ of the empirical model are constant globally and through the year, with the intention of creating a compact and universally valid model for the sea ice emissivity. Further details of the surface emissivity modelling are in Sec. 2.5.

The unknown empirical state inputs to the sea ice emissivity model ($\mathbf{x}_{\text{ice}}$) are interpolated from a geographical map $\mathbf{X}_{\text{ice}}$ in a similar way to the sea ice concentration. This map represents the empirical micro and macro physical properties of the sea ice and is another set of trainable parameters. The number of layers in this map, and hence the number of empirical inputs to the model, is a hyperparameter chosen to be 3 (sensitivity tests on this are in Appendix A1). Further details on the empirical state parameters are in Sec. 2.6.

As is well-recognised (e.g. Hsieh & Tang, 1998; Bocquet et al., 2020; Geer, 2021), machine learning and data assimilation are ultimately Bayesian inverse methods. Hence, Fig. 1 also represents the problem as a Bayesian network (e.g. Ghahramani, 2015), in particular mapping onto the graphical representation of data assimilation and ML used by Geer (2021). The physical content of the network could be seen as a sophisticated way of applying physical constraints to training of an empirical model, in the line of physics-informed neural network techniques (e.g. Raissi et al., 2017). Alternatively, the inclusion of a trainable empirical model within a physical data assimilation framework could be seen as an extended form of parameter estimation for data assimilation (e.g. Bocquet et al., 2019). But ultimately, all these views can be united in the Bayesian framework.

2.2 Technical summary

Figure 1 represents the network for one single observation, but the training is done on approximately 0.6 billion observed radiances across one year and with approximately 4.9 billion pieces of supporting information extracted from ECMWF short-range forecasts relating to the atmospheric radiative transfer terms and the surface temperature. Table 1 summarises the technical details of the network and its training.

The loss function J used in training includes the basic observational loss function J_{obs} (Eq. 1) alongside additional regularisation terms:

$$J = J_{\text{obs}} + J_{\text{seaice\_bounds}} + J_{\text{seaice\_tsfc}} + J_{\text{emis}} + J_{\text{bias}}. \tag{2}$$

The sea ice concentration is constrained by two loss functions representing physical constraints. The first sea ice loss function, $J_{\text{seaice\_bounds}}$, imposes the limits zero and 1 and the second, $J_{\text{seaice\_tsfc}}$, imposes the climatological probability of sea ice as a function of surface temperature (Sec. 2.7). The bias correction requires a loss function J_{bias} (Sec. 2.3) and finally there is a loss function for applying physical constraints to the sea ice emis-

Table 1. Overview of the hybrid network and its training details.

Aspect	Parameter	Value
Training data	Time period start	1st July 2020
	time period end	30th June 2021
	Number of observations	64,184,021
	Channels per observation	10
	TOTAL (observations by channels)	641,840,210
Geographical grid	Resolution	40 km
	Number of spatial points	62,499
	Number of time points (see Sec. 2.6 and 22.7)	365 or 366
Trainable parameters	Sea ice map $\mathbf{C}_{\text{ice}}$ (366 days, see Sec. 2.7)	22,874,634
	Empirical micro- and macro-physical properties $\mathbf{X}_{\text{ice}}$ (3 variables, 365 days)	68,436,408
	Empirical sea ice emissivity model weights $\mathbf{w}$	50
	Observation bias correction coefficients $\mathbf{b}_{\text{ice}}$, $\mathbf{b}_{\text{wat}}$	20
	Ocean surface emissivity windspeed correction coefficients $\mathbf{b}_{\text{emis}}$	10
	TOTAL	91,311,120
Fixed parameters	Atmospheric radiative transfer $\mathbf{z}_A$	61 per observation
	Surface (skin temperature ($T_S, \mathbf{z}_B$), surface windspeed, ocean emissivity $\mathbf{e}_{\text{wat}}$)	13 per observation
	Geolocation $\mathbf{z}_G$ (grid point, day, observation number)	3 per observation
	TOTAL (multiplied by number of observations)	4,942,169,617
Loss functions	Observation fit J_{obs}, Eq. 1	
	Observational bias J_{bias}, Eq. 4	
	Ice emissivity J_{emis}, Eq. 9	
	Sea ice physical bounds $J_{\text{seaice\_bounds}}$, Eq. 11	
	Sea ice probability $J_{\text{seaice\_tsfc}}$, Eq. 12	

sivity J_{emis} (Sec. 2.5). All these additional loss terms are important to impose physical behaviour and to prevent the problem being under constrained. In practice the bias and ice emissivity terms J_{bias} and J_{emis} were set so tightly that almost no change from the prior was allowed, but these terms provide a way of carrying out sensitivity tests on the number of parameters that can be constrained within the framework (see Appendix). Note also that the maps of empirical state parameters are not constrained at all. In Bayesian terms this means we impose no prior knowledge on what these values should be (this is often termed 'equal priors') following normal practice in ML.

The geographical maps of sea ice $\mathbf{C}_{\text{ice}}$ and empirical properties $\mathbf{X}_{\text{ice}}$ are estimated on an irregular spatial grid with a consistent resolution of about 40 km, containing 62499 points, and with a time resolution of 1 day. Precisely, the grid is derived from an N256 reduced Gaussian grid (Hortal & Simmons, 1991) from which all land points and latitudes less than 45° have been removed. A typical single map layer therefore contains 62499 locations × 365 days = 23 million parameters to be estimated.

Training of the hybrid model and geographical maps was done on the ECMWF supercomputer with a single process allowing 64 CPU threads, 128 GB of memory, and a

Table 2. AMSR2 channels in this work: (top) details from Okuyama and Imaoka (2015); (bottom) usage details.

Original channel number	5	6	7	8	9	10	11	12	13	14
Polarisation	v	h	v	h	v	h	v	h	v	h
Frequency [GHz]	10.65		18.7		23.8		36.5		89	
Footprint [km]	24 × 42		14 × 22		15 × 26		7 × 12		3 × 5	
Noise [K]	0.55	0.47	0.56	0.54	0.51	0.41	0.89	1.01	1.18	0.91
Local channel number j	1	2	3	4	5	6	7	8	9	10
Short name	10v	10h	19v	19h	24v	24h	37v	37h	89v	89h
Observation error [K]	2.5	4.0	2.5	4.5	2.5	5.0	4.0	7.0	4.5	10.0
RMS error initial [K]	19.31	20.61	8.41	41.25	18.02	20.80	24.74	23.60	23.85	34.94
RMS error analysis [K]	2.69	4.32	2.58	4.92	2.66	5.22	3.88	7.63	4.71	9.93

maximum 48 h of wallclock time (no GPUs were available). This allowed up to 8 training epochs, which was sufficient in the current work. To train the model, the loss function (Eq. 2) was minimised using the Adam variant of the mini-batch stochastic gradient descent approach (Kingma & Ba, 2014) at its default settings, including its learning rate set to 0.001, with the exception of the use of a batch size of 1024 (Appendix A1). From the point of view of Keras training, the fixed parameters were treated as input values (features), the observations were treated as output values (labels). Further technical aspects of the training and hyperparameters are explored in Appendix A1.

2.3 Observations

Observations are provided by AMSR2 (Okuyama & Imaoka, 2015), a conical scanning microwave radiometer on the polar-orbiting GCOM-W2 satellite observing at frequencies from 6.7 GHz to 89 GHz, each separately measuring vertically (v) and horizontally (h) polarised radiances (Tab. 2). The original radiance observations were obtained from the Japan Aerospace Exploitation Agency (JAXA, https://gportal.jaxa.com) and averaged onto a 40 km reduced Gaussian grid (Hortal & Simmons, 1991). This averaging standardises the measurement footprint, which varies with frequency (the footprint varies both in size, Table 2, but also in central location by up to around 4 km). The spatial locations of the resulting 'superobservations' or superobs are taken as those of the centres of the Gaussian grid points. The times of the superobs are the mean of the times of the original observations, which span only a few seconds, so the superob remains an almost instantaneous observation of the earth. AMSR2 also has channels at 6 – 7 GHz with excellent sea ice sensitivity but these have been left out due to their footprint being larger than the 40 km grid. Superobs are based on an average of 20 raw observations; those based on less than 6 raw observations are discarded. Following current practice in atmospheric data assimilation (e.g. Kazumori et al., 2016; Geer et al., 2018), the observations are used in all-sky conditions, i.e. clear, cloudy and precipitating.

Figure 2 illustrates these observations over the Arctic ocean. At 10 GHz, v-polarised, (10v) the sea ice areas are relatively easily distinguished from ocean by the large contrast in brightness temperature, with sea ice showing a fairly uniform value around 250 K. But at higher frequencies (19v – 89v) and in the horizontally-polarised channels (10h – 89h) the sea ice areas show more strongly variable brightness temperature, driven by the micro and macrophysical characteristics of the sea ice and snow cover. The influence of the atmosphere (e.g. clouds and water vapour) is also important, particularly at 89 GHz.

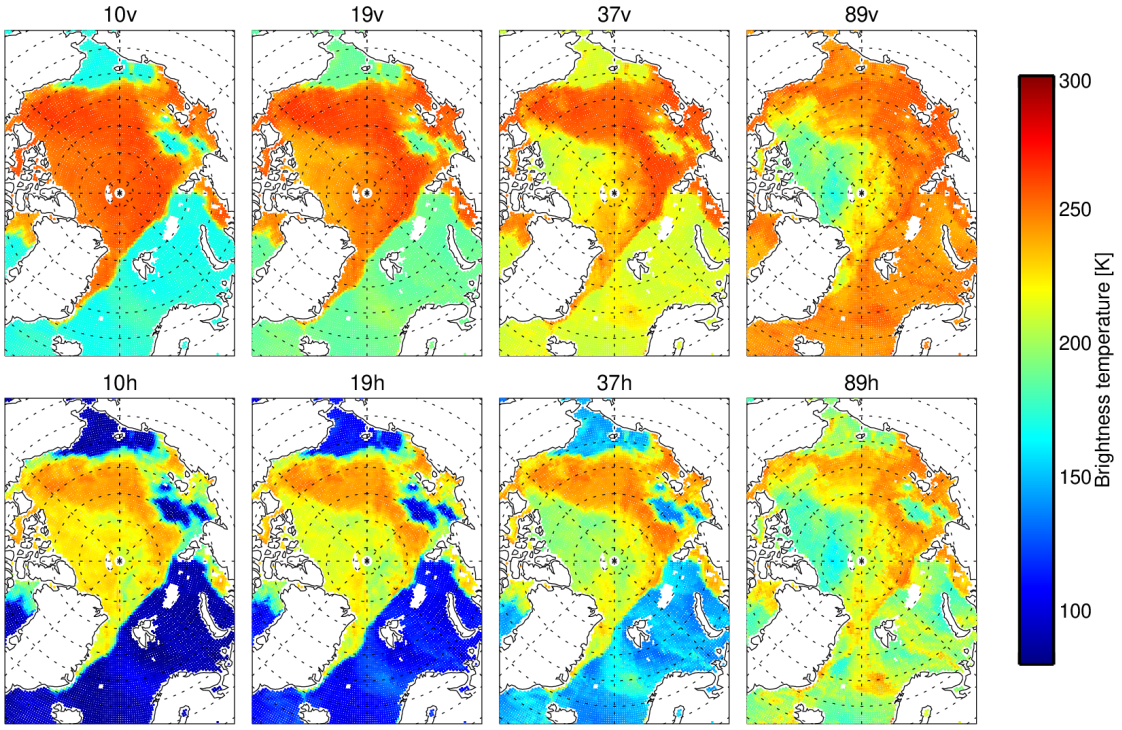

Figure 2. Observed AMSR2 brightness temperatures on the 40 km grid over the Arctic on 7th November 2020, showing only pure ocean scenes, and allowing observations from multiple orbits to overlay each other. Channels 24v and 24h are also used in this work, but are omitted from the figure to save space.

The ocean surface generally has much lower brightness temperatures than the sea ice and is more strongly polarised, so that h-polarisation measurements have much lower brightness temperatures than v-polarisation measurements, down to around 85 K at 10h. Existing heuristic sea ice algorithms use higher frequencies to obtain better spatial resolution, and are based on empirically observed characteristics of the polarisation and frequency dependence of ocean and sea ice surfaces. These include methods using 19v, 37v, 19h and 37h (e.g. Comiso et al., 2003; OSI-SAF, 2016) and those using 89v and 89h (e.g. Spreen et al., 2008) but in general no existing algorithm uses a physical description of the problem like the hybrid network used here (Fig. 1). This is mainly due to the lack of knowledge of the sea ice emissivity and its variations with underlying micro- and macrophysical ice and snow characteristics. The heuristic approaches can be vulnerable to atmospheric effects such as cloud and water vapour along with variations in sea ice and ocean surface characteristics, which could incorrectly be interpreted as variations in sea ice concentration.

Because of its orbit, AMSR2 crosses the polar regions every 100 minutes or so, taking measurements across a sub-satellite swath of 1450 km. Each overpass produces separate superobs, so up to around 8 of these are available on the 40 km grid in a 24 h period, and Fig. 2 has allowed multiple observations to overlay each other. Figure 3 shows the average number of superobs per day, counted on the daily 40 km grid on which the sea ice concentration and empirical properties ($\mathbf{C}_{\text{ice}}$, $\mathbf{X}_{\text{ice}}$) are estimated. The interpolation operator from Fig. 1, $h(\cdot, \mathbf{z}_G)$ is responsible for mapping between the sea ice daily grid and the location of the observations, whose availability changes from day to day, primarily due to orbital precession. The mapping is made easy because the superobs and

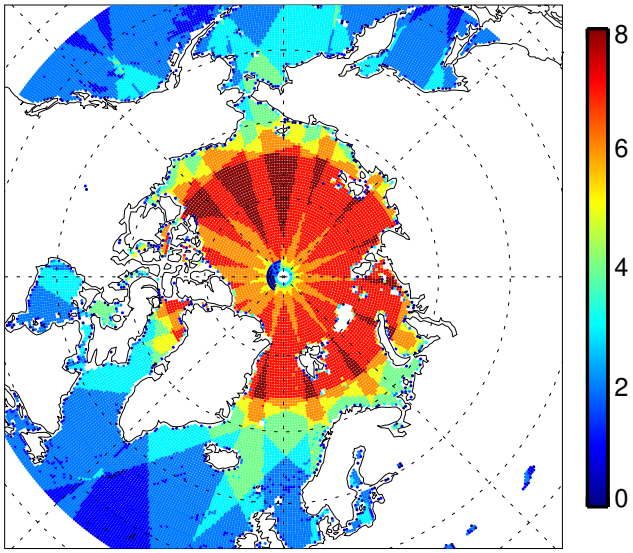

Figure 3. Number of AMSR2 pure ocean superobs available on the 40 km grid over the Arctic on 7th November 2020, in the 24 h period from 2100 UTC on 6th November to 2100 UTC on 7th November. Dashed lines indicate latitude every 10° and longitude every 45°.

the sea ice grid are on the same standardised locations. The result is that the daily estimates of sea ice concentration, and its the empirical properties, typically have to fit between 4 and 8 AMSR2 superobs and represent, in rough terms, a daily average.

AMSR2 has measurement biases which vary as a function of scene brightness temperature and can reach 5 K (Berg et al., 2016). It is routine and necessary to remove these biases when the data are assimilated. Hence in practice (and omitted from Fig. 1 for simplicity) a bias-corrected brightness temperature $\mathbf{y}_{\text{corr}}$ is used in the training (precisely, in the observation loss function Eq. 1) in place of the uncorrected $\mathbf{y}$:

$$\mathbf{y}_{\text{corr}} = \mathbf{y} + C_{\text{ice}} * \mathbf{b}_{\text{ice}} + (1 - C_{\text{ice}}) * \mathbf{b}_{\text{wat}} \tag{3}$$

The bias corrections are a function of the sea ice concentration, and are per channel, so there is a vector of 10 bias corrections for for sea ice, $\mathbf{b}_{\text{ice}}$, and similarly for ocean water, $\mathbf{b}_{\text{wat}}$. These are also trainable parameters, though in practice they are not allowed to vary much. The bias correction is initialised with the background values $b_{\text{ice,bg}}$ and $b_{\text{wat,bg}}$, respectively set to 2.5 K and 5.0 K (in all channels) using prior estimates of AMSR2 biases over ocean and land surfaces (Geer et al., 2022) and assuming that sea ice surfaces will have similar biases to those seen over land surfaces. Given that the bias corrections are trainable values, they are constrained by a loss term

$$J_{\text{bias}} = \frac{1}{n} \sum_{j=1}^{m} \frac{(b_{\text{ice},j} - b_{\text{ice,bg}})^2 + (b_{\text{wat},j} - b_{\text{wat,bg}})^2}{b_{\text{bgerr}}^2}. \tag{4}$$

Here, the per-channel bias corrections are $b_{\text{ice},j}$ and $b_{\text{wat},j}$ with a background error of $b_{\text{bgerr}} = 0.001$ K, and the division by n, the number of training observations, standardises the loss function to the Keras approach (Eq. 1). Given the very small chosen background error, in practice the bias is forced to stay extremely close to prior values, rather than being allowed to evolve to fit the observations. Hence the bias loss term exists mainly to explore the possibility of relaxing this constraint and evolving the bias correction in the sensitivity tests described in A2. These show that if biases are allowed to evolve away

from the prior values, it makes the problem under-constrained, so in practice it is necessary to estimate any instrument biases in advance.

A final aspect of using observations in DA is the observation error $\mathbf{r}$, which comes into the observation loss term (Eq. 1). The uncertainty in the observations themselves is indicated by the instrument noise, which is around 0.5 to 1.0 K (Table 2). However, in a DA system that does not otherwise account for forward modelling error, this must be represented in the observation error, and hence the total observation error is often much larger than the instrument noise (e.g. Geer & Bauer, 2011). In the current work, there is substantial forward modelling error not represented within the hybrid network (Fig. 1). This is because the network is not a perfect description of the physical processes and also the fixed parameters (e.g. $\mathbf{z}_{\text{atmos}}$) are not perfectly known. The assigned observation errors are given in Table 2 and have been inspired by the size of the residuals after training earlier prototype versions; estimating these errors is recognised as an iterative problem (Desroziers et al., 2005). The assigned errors range from around 2.5 K in channels 10v, 19v and 24v, up to 10.0 K in channel 89h. The larger observation errors reflect the increased difficulty in modelling the complex brightness temperature patterns over sea ice at higher frequencies and in h-polarised channels, as well as any errors in the fixed parameters for cloud and water vapour that mainly affect higher frequencies (see also Fig. 2).

2.4 Atmospheric radiative transfer

Atmospheric radiative transfer terms $\mathbf{z}_{\text{atmos}}$ have been computed prior to training, using the background 12 hour forecast from the ECMWF data assimilation system, which assimilates AMSR2 observations for their atmospheric information content in all-sky conditions (Kazumori et al., 2016) as well as many other satellites and observation types (e.g. Geer et al., 2017). Atmospheric radiative transfer is simulated by the physical scattering radiative transfer model RTTOV-SCATT (Radiative transfer for TOVS Scattering module, Bauer et al., 2006). This uses two independent sub-columns, one clear and one cloudy. The clear sub-column simulates the surface interaction and absorption by gases, primarily water vapour. The cloudy sub-column also includes the effect of cloud and precipitation using a delta-Eddington scattering solver. In each sub-column $k\epsilon[\text{clear}, \text{cloudy}]$, and for one channel j, the top-of-atmosphere brightness temperature is described by:

$$y_{jk} = e_j T_S \Gamma_{jk} + (1 - e_j) T^{\Downarrow}_{jk} \Gamma_{jk} + T^{\Uparrow}_{jk}. \tag{5}$$

Here, $T^{\Downarrow}_{jk}$ is the downwelling radiation (TB) at the surface, $T^{\Uparrow}_{jk}$ is the component of upwelling radiation at the top of the atmosphere coming from the atmosphere itself, and Γ_{jk} is the atmospheric transmittance. The surface is represented using the approximation of specular reflection, where the surface emits radiation according to the surface temperature T_S multiplied by the surface emissivity e_j, and reflects downwelling radiation modulated by a reflectivity $1-e_j$. This is not a perfect assumption, because microwave radiation can penetrate centimetres to metres into snow and ice surfaces, and future work will use a more physical representation of the radiative transfer within the sea ice and snow (e.g. Picard et al., 2018). The final 'all-sky' brightness temperature is obtained by weighting the two sub-columns according to the effective cloud fraction C_{eff} (Geer et al., 2009):

$$y_j = (1 - C_{\text{eff}}) y_{j,\text{clear}} + C_{\text{eff}} y_{j,\text{cloudy}} \tag{6}$$

Given that there are 10 channels, two subcolumns, and Eq. 5 requires three atmospheric variables to be prescribed, this means that $\mathbf{z}_{\text{atmos}}$ is composed of 60 radiative transfer terms plus the cloud fraction C_{eff}. These are required at the locations of every observation in the training set (Table 1). The most important atmosphere-related approximation is that $\mathbf{z}_{\text{atmos}}$ is based on an ECMWF 12 h forecast but is treated as a fixed truth; the impact of this is explored later.

These equations are already used in a different way in a dynamic surface emissivity retrieval at ECMWF (Baordo & Geer, 2015, 2016). This traditional technique for estimating the surface emissivity of land, snow and ice surfaces attempts to directly invert Eqs. 5 and 6 to obtain e_j, given fixed estimates of all other parameters. This can fail in many circumstances including when the surface becomes invisible due to heavy cloud or high water vapour amounts ($\Gamma_{jk} \to 0$) though this is mainly a problem of higher frequency channels (Baordo & Geer, 2016). An advantage of the Bayesian inversion of the physical forward modelling framework used here (Fig. 1) is that it naturally handles this situation, and does not attempt to extract information from observations where there is none to be had.

2.5 Surface emisivity model

The mixed-surface emissivity $\mathbf{e}$ (which is a vector over the 10 frequencies and polarisations of AMSR2 being used here) depends on the fractional cover of sea ice within the scene $C_{\rm ice}$ and on the emissivities of sea ice and open ocean $\mathbf{e}_{\rm ice}$ and $\mathbf{e}_{\rm wat}$:

$$\mathbf{e} = C_{\rm ice}\mathbf{e}_{\rm ice} + (1 - C_{\rm ice})\mathbf{e}_{\rm wat}. \tag{7}$$

If the sea ice and water emissivity were both known then it would be possible to retrieve the sea ice concentation from the total surface emissivity using this equation. Even if the ocean emissivity is treated as fully known, the problem of estimating the unknown sea ice concentration and ice emissivity from this equation is ill-posed (e.g. Rodgers, 2000). There are 11 unknowns but only 10 simultaneous equations. One way to make it well-posed would be to constrain the frequency variation of ice emissivity $\mathbf{e}_{\rm ice}$. Hence, one aim of the current work is to generate a sea ice emissivity model with significantly fewer input parameters than the number of frequencies in $\mathbf{e}_{\rm ice}$, in order to facilitate sea ice concentration retrievals.

The ice surface emissivity model being trained in this work is:

$$\mathbf{e}_{\rm ice} = f_{\rm empirical}(\mathbf{w}, \mathbf{x}_{\rm ice}, z_B) = \mathbf{w}_b + \mathbf{w}_0 z_B + \sum_{l=1}^{p} \mathbf{w}_l x_{{\rm ice},l}. \tag{8}$$

Here, $\mathbf{w}_b$ and $\mathbf{w}_0$ to $\mathbf{w}_p$ are each a vector across the 10 microwave channels and they are extracted from the 50-element weights vector $\mathbf{w}$ (Table 1). With $p = 3$ empirical variables to represent the micro- and macrophysical properties of the sea ice and snow, there are $l = 1, p$ components of $\mathbf{x}_{\rm ice}$, written $x_{{\rm ice},l}$ as inputs to the emissivity model, plus a scalar z_B. This model is implemented as shown in the equation as a single neural network layer with a linear activation function, although more complex and nonlinear neural networks have also been tried (Appendix A1). The appendix also justifies the choice of $p = 3$ empirical variables.

The trained values of $\mathbf{w}_b$ and $\mathbf{w}_0$ to $\mathbf{w}_3$ are illustrated in Fig. 4 as well as being tabulated in Appendix B. Here, $\mathbf{w}_b$ broadly represents a mean ice surface emissivity and other vectors describe variability around this as a function of the input values. Of these inputs, z_B is the only physical one: it is a scalar transformation of the surface temperature $z_B = \max(273.0 - T_S, 0.0)/30.0$. The factor 30.0 in the denominator is for normalisation and is chosen so that maximum values do not much exceed 1, as is common practice in machine learning. As mentioned in Sec. 2.4, the representation of the surface using an emissivity and a skin temperature is a big approximation. Quite often the radiation may be coming from within the sea ice or snow layer, where it is warmer, and the effective radiating temperature of the snow and sea ice (in Eq. 5) should be higher than the given skin temperature from the ECMWF model. To partly compensate this, the model is designed to allow an increase in the surface emissivity as the skin temperature gets colder. Figure 4a shows the modelled surface emissivity at $T_S = 273.0\,{\rm K}$ ($z_B = 0$) and $T_S = 261.0\,{\rm K}$ ($z_B = 0.4$), assuming all other inputs are zero. Going to the lower

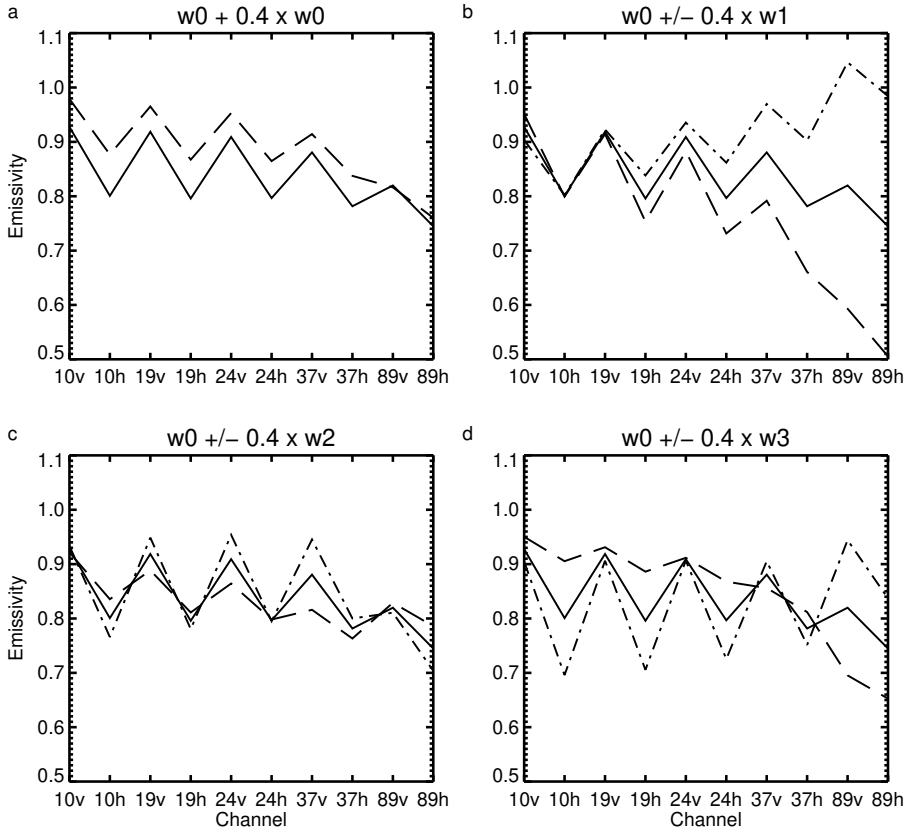

Figure 4. Illustrating the trained sea ice emissivity model. The ice emissivity with all inputs zero is $\mathbf{w}_b$ which is shown in all panels (solid line). If just one input parameter is perturbed to +0.4 (long dash) or -0.4 (dot dash) then the emissivity changes according to weights $\mathbf{w}_0$ to $\mathbf{w}_3$.

temperature increases the modelled surface emissivity by up to 0.07. The trained model describes this effect in a physically plausible way, being strongest at lower frequencies, where radiation typically penetrates deeper into the snow and ice, and weakest at 89 GHz, where the radiation is unlikely to be penetrating more than a few centimetres into the surface.

Figure 4b – d show the response of the trained emissivity model to changes in the empirical inputs, those variables whose meaning is defined by the model itself. Positive values of the first empirical value, $x_{\mathrm{ice},1}$, are able to make the surface emissivity decrease more strongly with frequency (panel b) although negative values can generate unphysical emissivity values outside the range 0 to 1. The second empirical value ($x_{\mathrm{ice},2}$, panel c) seems primarily to control the polarisation (the sawtooth shape on these plots) making it smaller when positive and larger when negative. Positive values of the third empirical value ($x_{\mathrm{ice},3}$, panel d) reduce polarisation and reduce emissivity mainly at 89 GHz. This represents a compact model of the polarisation and frequency dependence of the surface emissivity of ice and snow.

At the start of training, the sea ice emissivity model weights were initialised with the default initialiser (Glorot uniform; Glorot & Bengio, 2010) except for $\mathbf{w}_\mathrm{b}$ which was initialised to a background value of $w_{\mathrm{b,bg}} = 0.93$. With the aim to make the problem well-posed, the emissivity model was constrained, but only for the first element of the

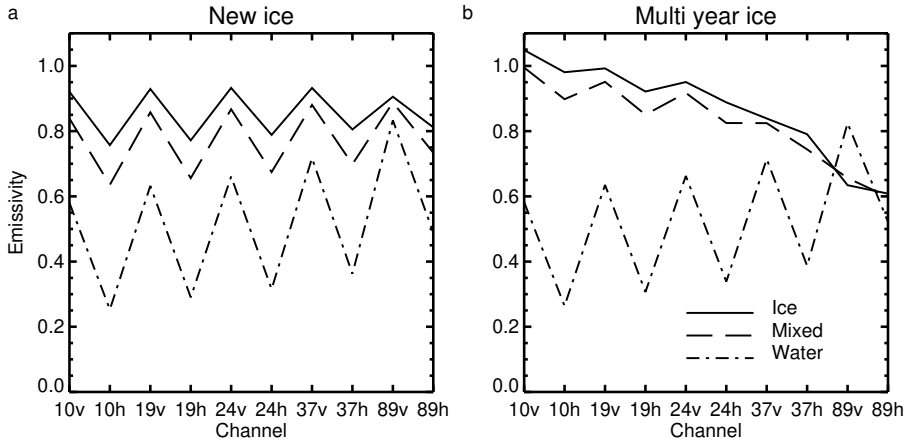

Figure 5. Emissivity as a function of channel at illustrative locations in the Arctic on 7th November (a) in the multi year ice and (b) in newly formed sea ice. The mixed emissivity (dashed) is generated from the sea ice concentration and the ice and water emissivities (solid and dot-dash) according to Eq. 7.

vector $\mathbf{w}_\mathrm{b}$, which gives the baseline surface emissivity for channel 10v:

$$J_\mathrm{emis} = \frac{1}{n}\frac{(w_{\mathrm{b},1} - w_{\mathrm{b,bg}})^2}{w_\mathrm{bgerr}^2}. \tag{9}$$

In practice, the size of the background error here, $w_\mathrm{bgerr} = 0.00001$, was chosen to keep $w_{\mathrm{b},1}$ extremely close to its background value. An emissivity of 0.93 is within the typical values for channel 10v, and further, the emissivity in this channel is thought to be relatively invariant between first year and multi year ice (Lee et al., 2017, their Table 4). The posterior sensitivity tests in Appendix A2 show that this constraint on 10v emissivity could in practice have been relaxed or removed. In early testing before the addition of sea ice concentration loss terms, the emissivity loss term was necessary to avoid completely non-physical sea ice concentrations being retrieved. However, it appears that the sea ice concentration loss terms are more effective and targeted, and would be sufficient on their own in future versions of this work. In any case, despite the constraint, the final trained model still has freedom to adjust the output emissivity at 10v down to at least 0.9 and up to at least 0.98 depending on the input parameters (Fig. 4).

Figure 5 illustrates sea ice surface emissivities generated by the empirical model (Eq. 8) at two locations in the Arctic, alongside ocean water emissivities (Eq. 10, to be described shortly) and the mixed-surface emissivity that is determined by the sea ice concentration (Eq. 7). The location with new ice (panel a) generates relatively strong polarisation but has little frequency dependence. The frequency dependence is in qualitative agreement with aircraft observations of new ice (e.g. Hewison & English, 1999). In contrast, a location in the multi-year ice is mostly unpolarised but its emissivity drops with frequency, as is generally observed (e.g. Baordo & Geer, 2015). Both sea ice emissivity spectra contrast strongly with the ocean water emissivity which is even more polarised and increases with frequency. These are locations where the sea ice concentration is close to 1, so the mixed-surface emissivity is relatively close to the ice emissivity.

To complete the description of surface emissivity modelling, ocean water emissivity is described as a function of skin temperature T_S, surface wind speed u, salinity and frequency by the FASTEM-6 model (Fast Emissivity, Kazumori & English, 2015). The

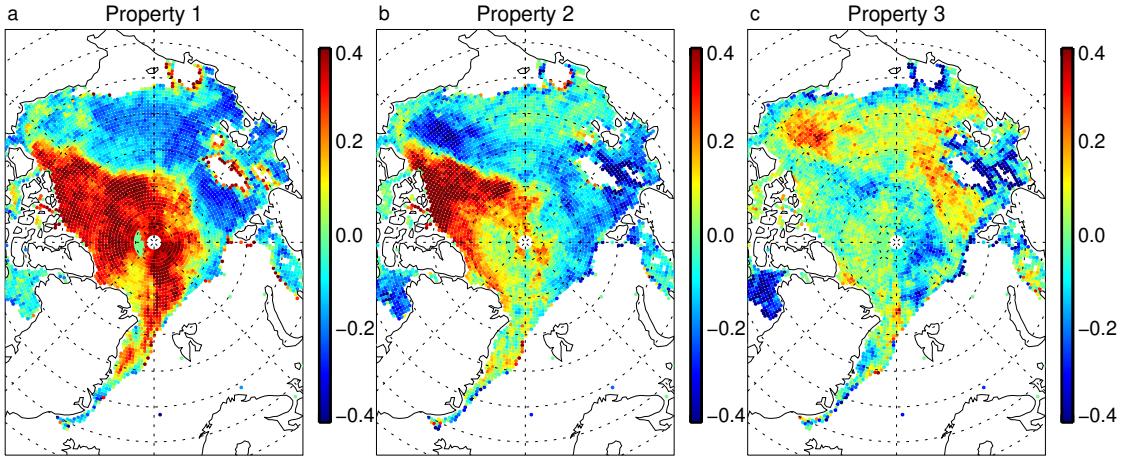

Figure 6. Estimated values of the three empirical sea ice properties in the Arctic on 7th November 2020. Values are only shown where the estimated sea ice concentration is greater than 0.2.

ECMWF 12 h forecast is used to provide surface temperature and wind speed at the observation time and location, and the salinity is fixed (35 in terms of practical salinity units). The ocean water cannot cool below around 271.35 K without freezing, but many sea ice locations have skin temperatures much lower than this. In these cases, the surface emissivity is set to an average of surface emissivities generated for surface temperatures between 271.0 and 273.0 K in nearby areas. FASTEM-6 also has biases as a function of windspeed u, coming from an imperfect representation of the sea state in high wind situations. Hence another alteration to the idealised network in Fig. 1 is to train a windspeed and channel-dependent bias correction $\mathbf{b}_{\text{fastem}}$ within the network:

$$\mathbf{e}_{\text{wat}} = \mathbf{e}_{\text{fastem}} + u\mathbf{b}_{\text{fastem}}. \tag{10}$$

Here, $\mathbf{e}_{\text{fastem}}$ is the ocean water surface emissivity coming from FASTEM-6. This windspeed bias correction requires 10 bias correction coefficients to be trained (one per channel, Table 1). The trained windspeed bias corrections appear to be well constrained, physically realistic, and they make only small adjustments to the simulated ocean emissivity (e.g. at $u = 20\text{ms}^{-1}$ at most +0.03 in emissivity in the 37h channel, and generally smaller). Hence, no loss function is used, and for brevity the corrections are not discussed further.

2.6 Empirical state representing micro- and macro-physical properties of sea ice

The initial value of the empirical parameters is zero before training, and as mentioned earlier, there is no attempt to constrain these values during training. Figure 6 illustrates the empirical sea ice properties after training. Positive values of properties 1 and 2 are found towards the W side of the Arctic sea ice and appear to correspond to multi-year ice, including a small tail of similar properties that has been advected around the Beaufort Gyre (near Alaska). The other parts of the Arctic sea ice have frozen during the autumn and have more variability, but properties 1 and 2 are typically zero or below, and property 3 is often zero or positive. Figure 5 has already illustrated typical corresponding sea ice emissivity spectra in Arctic multi-year and new ice. During the cold season, the empirical properties are often largely unchanged from one day to the next, but they can change rapidly, for example when new snow falls on the sea ice (not

shown). In the warm season, the empirical properties vary widely from day to day, likely depending on whether the snow cover is melting or frozen on the day in question (not shown).

2.7 Sea ice

In order to speed up the training, the sea ice maps $\mathbf{C}_{\text{ice}}$ were initialised with a monthly mean sea ice concentration computed from the existing ECMWF sea ice concentration analysis at the AMSR2 locations. However, apart from the physical constraints described below, the sea ice concentration can be freely adjusted to best fit the observations, and is not constrained in any way to the ECMWF sea ice concentration. The existing ECMWF sea ice analysis is based on a combination of an ocean model, a sea ice model (LIM2, Timmermann et al., 2005) and the assimilation of highly processed observations, using the ocean data assimilation framework (OCEAN5, Zuo et al., 2019; de Rosnay et al., 2022) and will be referred to by this name subsequently. The observations are the OSTIA sea ice dataset (Good et al., 2020) which is ultimately based on sea ice retrievals (OSI-SAF, 2016) obtained from a microwave sensor similar to AMSR2 using a traditional heuristic approach (e.g. Comiso et al., 2003). It is important that the new emissivity model should not try to fit these heuristic assumptions in any way. Further, due to the complex processing chain, the OCEAN5 sea ice can be at least 48 h behind reality once it has been mapped to observation locations (e.g. Baordo & Geer, 2015; Browne et al., 2019; de Rosnay et al., 2022, this work). As encountered when prototyping the current approach, if there is an incorrect sea ice concentration in Eq. 7, the sea ice emissivity model can take on characteristics of the ocean water surface emissivity, which would make the model essentially useless. For all these reasons, it was important not to allow the new sea ice analysis to be constrained by the OCEAN5 sea ice analysis in any way.

A more standard DA approach would have been to use the OCEAN5 sea ice as the background and to impose a loss term based on the misfit between this and the solution. This was tried in prototypes but it proved very difficult for the solution to move away from the OCEAN5 sea ice field and its known issues. Hence this was abandoned, but prototyping also showed that if the sea ice were not constrained at all, it could drift to unphysical values outside the range 0 to 1 (see also Appendix A2). Hence an alternative approach to constrain the sea ice was to impose physical bounds and climatological information on the likelihood of sea ice. The first of two sea ice loss functions impose a quadratically increasing penalty on sea ice concentrations that go outside the bounds 0 and 1:

$$J_{\text{seaice\_bounds}} = \frac{1}{n}\sum_{ab}\frac{\max(C_{ab}-1.0,0.0)^2}{c_{\text{bgerr}}^2} + \frac{1}{n}\sum_{ab}\frac{\min(C_{ab},0.0)^2}{c_{\text{bgerr}}^2} \tag{11}$$

Here, C_{ab} indicates one sea ice concentration on the grid, and a and b represent the time and space grid indices. The loss is computed over the entire grid of 22 million locations (Table 1). In the mini-batch minimisation technique, losses need to be computed once per batch, and summing repeatedly over the entire sea ice grid is a performance limitation that should ideally be avoided in future (see Appendix A1). The background error is set to $c_{\text{bgerr}} = 0.02$, with sensitivity experiments on this in Appendix A2.

The second sea ice loss penalises the presence of sea ice in warm locations and is primarily designed to prevent the aliasing of observed but not modelled cloud into trace sea ice features. This loss was made a function of the climatological probability of observing sea ice greater than 0.01 as a function of skin temperature, $P(C_{\text{ice}} > 0.01|T_S)$, estimated from the OCEAN5 data. Since in variational data assimilation the background loss term is -2 times the natural logarithm of the prior probability (see e.g. Geer, 2021) the relevant penalty function and its approximate piecewise linear fit to the OCEAN5 data were $-2\ln\left(P(C_{\text{ice}} > 0.01|T_S)\right) \simeq 4.0 \times \max(T_S - 273.2, 0.0)$. To apply this to all

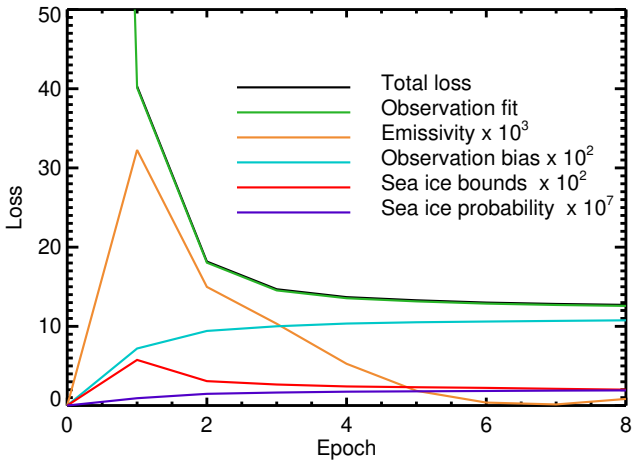

Figure 7. Loss terms during the training of the hybrid network (see Table 1) The total loss term is dominated by the observation fit, so it is mostly hidden behind the observation fit losses; the total loss at epoch zero is 340 and is off scale. The order of lines in the key follows the top-to-bottom order on the figure at Epoch 2.

sea ice values, the full loss term was:

$$J_{\text{seaice\_tsfc}} = \frac{1}{n} \sum_{d} 4.0 \times \max(T_{S,d} - 273.2, 0.0) \tag{12}$$

Here $T_{S,d}$ is the skin temperature at the subset of grid locations d with sea ice concentrations greater than 0.01, $C_{ab} > 0.01$. This subsetting means that the penalty applies only where the sea ice concentration is greater than 0.01.

A final constraint on the sea ice concentration is to impose a time smoothing, so that the interpolation operator $h(\mathbf{C}_{\text{ice}}, \mathbf{z}_G)$ (Fig. 1) takes a configurable weighted average of the sea ice at the location $\mathbf{z}_G$ from the current and previous day (this is why the sea ice grid covers 366 rather than 365 days). In this work the weights were 0.6 and 0.4 respectively, giving most weight to the current day's ice concentration. This is intended as a rough equivalent to using a persistence (i.e. constant) sea ice forecast model to constrain the evolution of the geophysical state from one day to the next. The model constraint is a particularly important part of data assimilation, and although a physical sea ice model would have been better here, the persistence approximation has been effective in sea ice data assimilation (e.g. Buehner et al., 2013). Sea ice persistence is useful for reducing the amount of spurious sea ice features generated by cloud over open ocean; broadly the assumption is that sea ice is longer-lived than cloud features.

3 Results

3.1 Training and fit to observations

Figure 7 shows the total loss J and its constituents (Table 1) as a function of the training epoch. Losses before any training (referred to as epoch 0) have been estimated using the initial conditions of the hybrid model, described further below. The model is mostly converged to the observations after 4 epochs. However, allowing the training to run to 8 epochs helps reduce the size of the sea ice emissivity loss term, keeping the 10v emissivity closer to the prescribed value of 0.93, and the sea ice bounds term, reducing

the occurrence of non-physical sea ice concentrations. The loss terms have a very wide range of magnitudes and hence all but the observation term have required re-scaling to be visible on this figure. Despite this, the results show that most of these additional smaller loss terms are effective (see appendix). The hybrid model appears mainly converged after 8 epochs, but Appendix A1 (using just a month of training data) investigates the use of larger numbers of epochs, and shows that the model can continue to evolve, but at the cost of some overfitting. Training on the yearly dataset is resource-limited to 8 epochs, but a beneficial, though involuntary, side effect has been that it is using 'early stopping' to avoid over-fitting.

Figure 8 illustrates the brightness temperatures generated by the network before ('initial') and after training ('analysis'), compared to observations. The initial simulated brightness temperatures are what would be obtained from the network before training, with all trainable parameters set to their initial values, with the exception that the emissivity weights $\mathbf{w}_0$ (the term sensitive to the skin temperature) have been set to zero. This is because the model weights are otherwise semi-randomly set by the Glorot initializer, making the true initial fit to observations much worse, and not particularly informative. With $\mathbf{w}_0 = 0$, the initial values of sea ice emissivity are all set to 0.93 and it is easy to see the brightness temperature errors caused by the monthly mean initial sea ice fields (panel a compared to c). Compared to the initial simulated brightness temperatures, the analysis replicates the observations very closely.

The RMS of the initial and final (or analysis) departures ($y_{\mathrm{obs},ij} - y_{ij}$) are given in Table 2. Initial RMS departures are of order 10 K to 40 K. Panel g illustrates these departures in channel 37v, with large discrepancies coming both from the incorrect surface emissivity and the incorrect sea ice initial field. By contrast, the analysis departures (e.g. panel h) are mostly within ±3 K and rarely larger than 10 K. The largest remaining departures are mainly found over open ocean and not sea ice, and come from errors in the location of cloud and precipitation in the ECMWF background fields, which cannot be adjusted in the current network. The RMS of the analysis departures ranges from 2.7 K in channel 10v to 9.9 K in 89h (Table 2). These values are close to the prescribed observation errors, as intended.

In Fig. 8a-c, at 10v, there is evidence of large adjustments in the sea ice field all around the sea ice edge, and particularly in the top right (Siberian) sector of the Arctic ocean, where the observations show some large holes in the sea ice (panel c), which are absent from the initial model (panel a) but which the analysis fits closely (panel b). Brightness temperatures over the sea ice are also increased by around 10 K in the 10v channel analysis, mainly due to the tuning of the skin temperature term ($\mathbf{w}_0$) of the ice emissivity model, in order to boost the surface emissivity in colder areas as intended (not shown).

At higher frequencies, the improvements between the initial model and the analysis (Fig. 8 d and e) are driven not just by the improved sea ice concentration, but also by the development of the empirical terms of the surface emissivity model, and the empirical state inputs. Empirical properties 1 and 2 (Fig. 6) are the main driver in the analysed brightness temperature at 37v, helping generate brightness temperatures down to 210 K over the multi-year ice (empirical property 1 values of up to 0.4) and up to 260 K over the first year ice (empirical property 1 down to -0.4). Hence the hybrid model appears to make a physically plausible set of adjustments in order to fit the observations.

3.2 Sea ice concentration

Figure 9 compares the Arctic sea ice obtained from the physical-empirical network to the OCEAN5 sea ice during a rapid freezing event. The new analysis (panel b) has identified a mostly complete sea ice cover over a region of the eastern Arctic ocean approximately 1000 by 3000 km where it scarcely exists in the existing ECMWF sea ice anal-

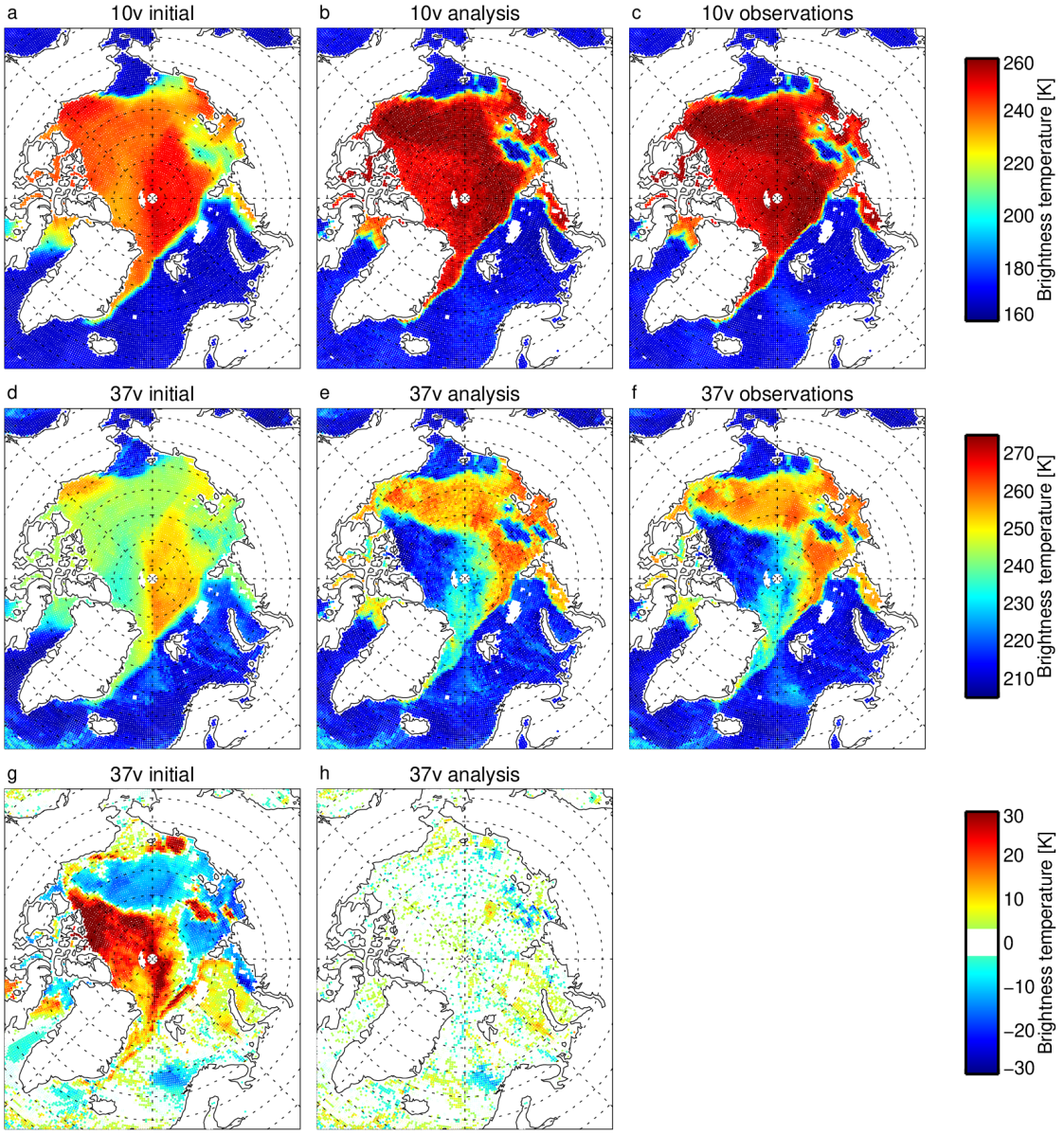

Figure 8. Comparison of simulated and observed brightness temperatures in channels 10v and 37v, over the Arctic on 7th November 2020: (a,d) initial guess; (b,e) after training, in other words the analysis; (c,f) observations; (g) the initial guess minus observation departure; (h) the analysis minus observation departure. Departures smaller than 3 K are not shown. The figure overlays mutiple orbits of AMSR2 observations and allows later observations to cover earlier osbervations.

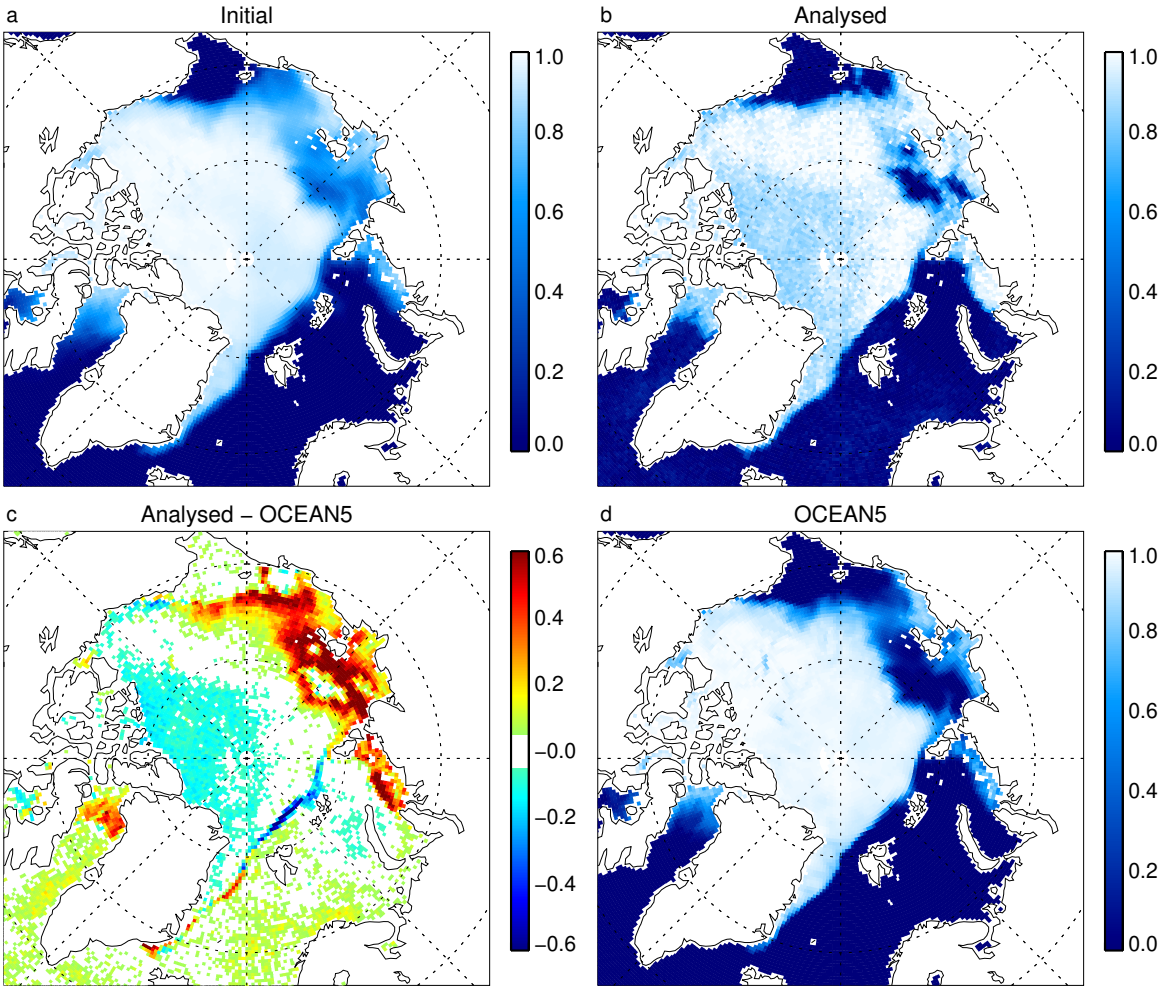

Figure 9. Comparison of sea ice concentration on 7th November 2020: (a) Initial monthly mean sea ice; (b) Analysed in this work; (c) Difference between analysed and OCEAN5; (d) OCEAN5 sea ice. Differences smaller than 0.05 are not shown. Latitude and longitude grid spacings are 10° and 45° respectively.

ysis (OCEAN5, panel d). There are also disagreements in the location of the ice edge by up to around 100 km in the region of Svalbard. The fundamental issue is the roughly 48 h time delay in producing the OCEAN5 analysis; eventually the OCEAN5 sea ice catches up and provides a similar picture (not shown). Table 3 shows that the new analysed sea ice agrees best with OCEAN5 if it is artificially delayed by 2 or 3 days, consistent with previous expectations (Baordo & Geer, 2015; Browne et al., 2019; de Rosnay et al., 2022). There is also some low concentration (< 0.2) sea ice incorrectly identified in the current analysis, for example off the N coast of Scandinavia, visible mainly in the difference plot (panel c). This spurious sea ice is generated when clouds are present in the observations but not in the ECMWF 12 h atmospheric forecast ($\mathbf{z_A}$ here). The network (Fig. 1) cannot adjust the atmosphere to add cloud and instead can only create sea ice to better fit the observations (an opposite effect would tend to create negative sea ice concentrations where the ECMWF forecasts have too much cloud, but this is suppressed by the sea ice bounds loss function).

Table 3. Standard deviation of differences between analysed and OCEAN5 sea ice concentration, aggregated across the Arctic and Antarctic for the full 365 day training period, with a variable lag applied to the analysed sea ice.

Lag [Days]	Std. dev.
0	0.0736
1	0.0692
2	0.0663
3	0.0666
4	0.0697
5	0.0738

Figure 10 shows a similar comparison for the Antarctic. Here, the interior sea ice amounts are close to 1 and agree well with OCEAN5. The main differences are along the ice edge, where the analysed sea ice has a much sharper boundary than the OCEAN5 sea ice, and also a significantly more detailed structure. The analyses also have a slightly greater extent, broadly consistent with the 2-3 day time delay in the OCEAN5 data and noting that the sea ice extent is typically growing at this time of year. An interesting detail on this day is in the top right of the plot (around 20 - 35° E) where the new analyses show a distinct series of waves along the sea ice edge in a location where the OCEAN5 analyses show only a shallow gradient. These waves have wavelengths of around 300 km and are seen developing over many days (not shown). They might be produced by a series of ocean surface eddies along the ice edge.

Figure 11 compares the sea ice probability density function (PDF) between the analyses and the existing ECMWF sea ice, OCEAN5. The OCEAN5 fields show clear bounds at 0 and 1 and a smooth variation in between. For the sea ice analysed in this work, the sea ice bounds loss function (Eq. 11) has mainly done its job of keeping the sea ice concentration within bounds, but tails do exist outside the physical range. There are also spikes in the analysed PDF, indicating some quantisation in the analysis. Based on the sensitivity tests in Appendix A1, this quantisation would likely have disappeared if resource constraints had not prevented the use of more epochs for training. A final issue with the current work is the excessively high probability of observing sea ice concentrations below 0.25, which comes from the cloud-aliasing issue discussed above. However, the retrieved sea ice can easily be cleaned up by capping the concentration at 1 and by setting any sea ice concentrations below 0.25 to zero. More study into loss functions for sea ice would be useful, or alternatively a physical sea ice model could be introduced into the network to better constrain the sea ice PDF, following typical practice in data assimilation.

Figure 12 shows the annual cycle of ice area, using 'cleaned up' sea ice concentrations as described in the previous paragraph. Compared to OCEAN5, this work gives slightly higher ice area in the Antarctic winter, and slightly lower ice area in the Arctic winter. The better timeliness of the new sea ice analysis is also clear. In early November in the Arctic, the current work identifies a rapid freezing event that is smoothed out and delayed in the OCEAN5 sea ice analysis, as also illustrated in Fig. 9. A similar picture is seen in the Antarctic, from late February to mid-May. Apart from these issues, there is good agreement between the annual cycles in the two ice products.

Figure 13 shows maps of the annual mean sea ice concentration (again using cleaned up values) and the difference between these and OCEAN5. In the Arctic (panel c) there is an underestimation of sea ice concentration compared to OCEAN5, mainly in the multi-

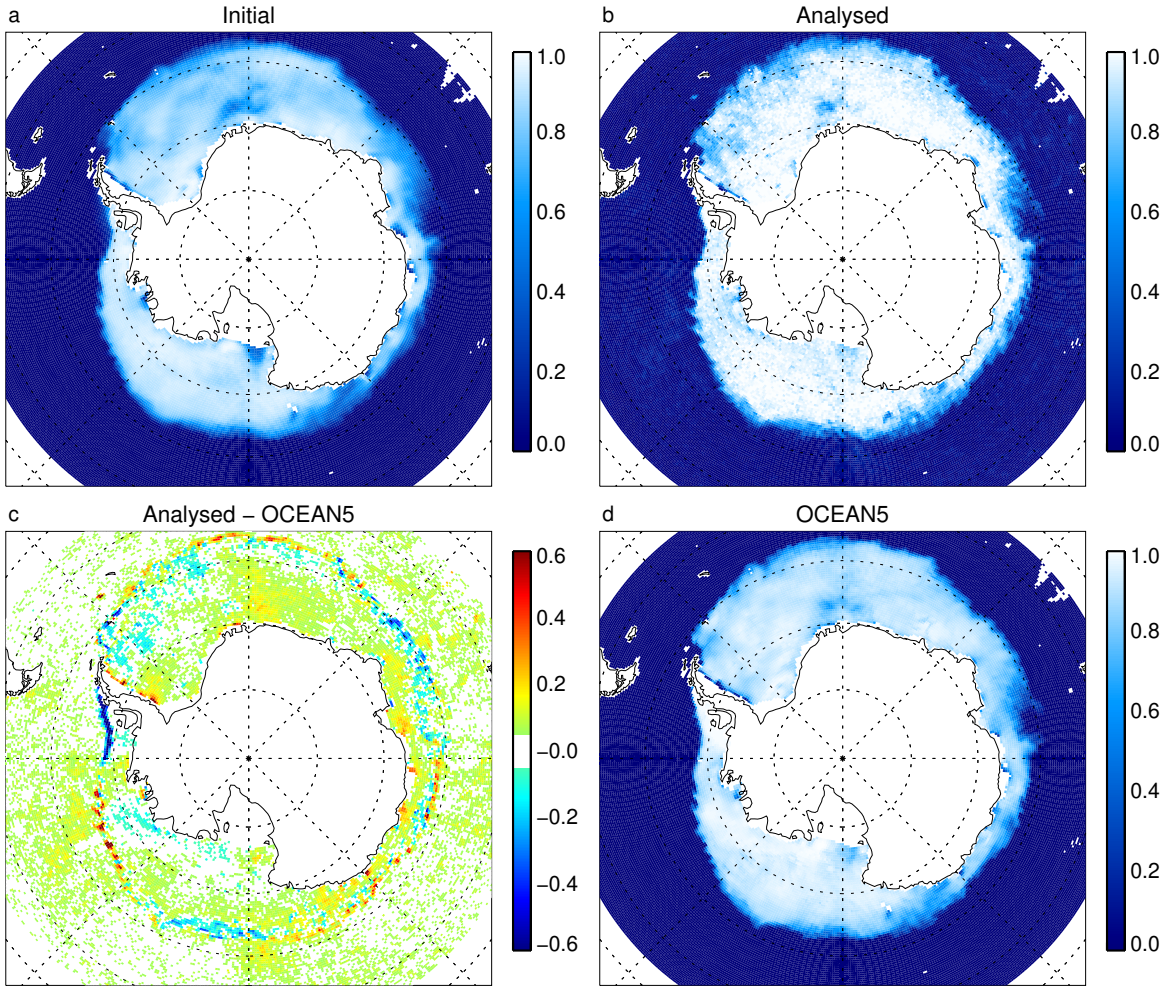

Figure 10. As Fig. 9 on 7th November 2020, but for the Antarctic. The Greenwich meridian is to the top.

year ice region (western Arctic ocean) and mostly by less than 0.1. This underestimation is visible even in the daily maps (e.g. Fig. 9). Assuming that the sea ice concentration in the multi-year ice zone should be close to 1, like in OCEAN5, this appears to be a defect in the current approach and likely relates to the constraint of 10v ice surface emissivity close to 0.93; Fig. 2a shows slightly lower TBs in this region even at 10v, compared to the eastern Arctic, which suggest that the modelled surface emissivity in multi-year ice should have been allowed to decrease more at 10v. In the Antarctic (panel b) there is typically a small overestimation (around 0.02) compared to OCEAN5, and much larger positive differences in patches close to the Antarctic coast. These can be traced to the Antarctic coastline in the summer, January and February, where the new analyses show considerably more ice extent (much of it fractional ice, e.g. concentrations around 0.5) along the coast than the OCEAN5 analyses. Given the careful treatment of the satellite field of view and land contamination in this work (Sec. 2.3) and the all-sky validity of the data this suggests there is a real defect in the OCEAN5 sea ice. There are small underestimations of the sea ice compared to OCEAN5 in the Weddel Sea which might also be due to the multi-year ice issue, since this is one of the few areas of the Antarctic where ice can persist from one year to the next. But apart from the areas of disagree-

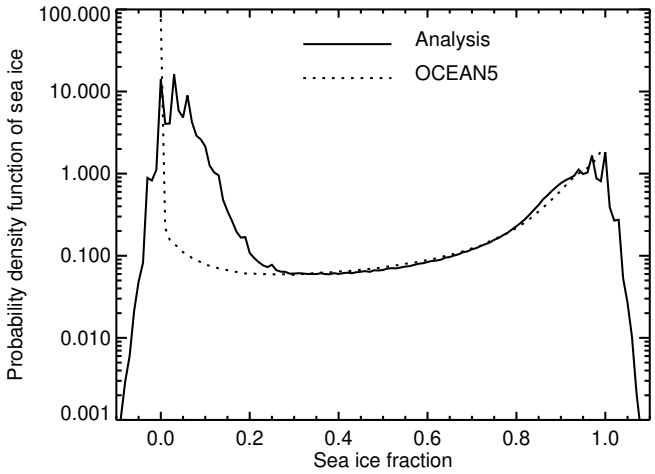

Figure 11. Probability density function for sea ice concentration for the new analysis and the existing OCEAN5, over the year and both hemispheres, using a log y-axis to better compare the full probability range.

ment already discussed, the annual mean ice concentrations agree very well across most of the Arctic and Antarctic, generally well within the bounds +0.05 and -0.05.

4 Conclusion

This work has trained a hybrid empirical-physical model to fit observed AMSR2 microwave radiances at 10 channels between 10 GHz and 89 GHz, over ocean and sea ice, based on a year-long training dataset. The atmospheric radiative transfer and the skin temperature are prescribed using 12-hour forecasts from the ECMWF atmospheric data assimilation system. The ocean surface emissivity is prescribed from a physically-based model. The primary unknowns have been the evolving sea ice concentration, the physical properties of the sea ice and its snow cover, and a model to determine the sea ice surface emissivity from those properties. These have been estimated simultaneously using a hybrid of machine learning and data assimilation. The AMSR2 observations have been fitted after training to within an RMS error of 2.6 K to 9.9 K, depending on channel. Most of the remaining error is over open oceans and is attributed to errors in the specification of clouds in the ECMWF 12-hour forecast. The analysis fits are much better than the initial RMS errors obtained using a flat sea ice surface emissivity model and monthly mean sea ice concentration, which are from 8.4 K to 35.0 K.

There is no perfect truth against which to evaluate the resulting global daily maps of sea ice, since existing global satellite retrievals of sea ice are based on heuristic methods which this work aims to replace, and in-situ measurements are very limited in coverage. The high quality of fit to observations suggests that the sea ice results are good, but comparisons are also made to the existing ECMWF sea ice analyses, generated by the OCEAN5 assimilation system. The new maps are 48 h to 72 h more timely than the sea ice concentration analysed in OCEAN5. The new maps have generally sharper resolution including plausible mesoscale features like some apparently eddy-generated 300 km wave features in the Antarctic sea ice edge. The new data also suggest there is a substantial underestimate of sea ice cover in the OCEAN5 analyses in the Antarctic summer. There are some limitations in the new data too, mainly an apparent underestimate

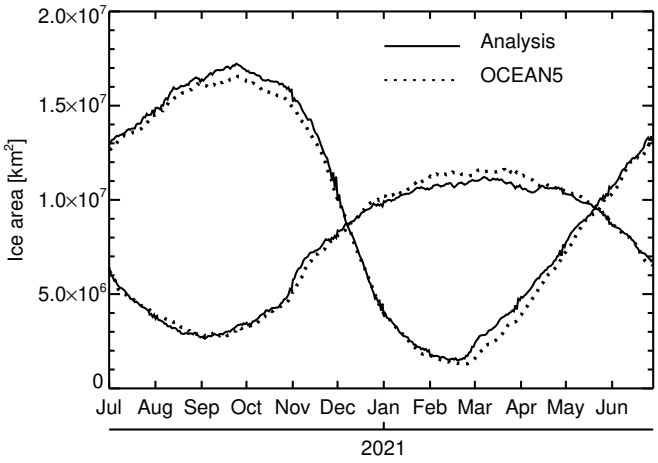

Figure 12. Estimated hemispheric sea ice area from the current work (solid) and OCEAN5 (dotted) from 1st July 2020 to 30th June 2021. The Antarctic ice extent is largest in September and the Arctic in March.

of the sea ice concentration in the Arctic winter multi-year ice, where it seems to be about 0.1 too low, in compensation for remaining inaccuracies in the sea ice emissivity model.

The hybrid model has also created daily maps of three empirical parameters that describe the sea ice and snow microphysical and macrophysical properties that affect microwave radiative transfer, along with an empirical model to convert these properties into the sea ice surface emissivity. The empirical parameters are linked to known properties of the sea ice such as differences between new ice and multi-year ice in the Arctic. These parameters are then used by the surface emissivity model to describe the surface emissivity as a function of frequency and polarisation. Examples show multi-year ice with a relatively flat but polarised surface emissivity spectrum, and the new ice with a less polarised spectrum that drops off towards higher frequencies. As shown by the globally and seasonally good fit to observations, the model is also able to handle all other conditions across the seasons and both hemispheres, such as thawing surfaces.

The broader goal of this work has been to demonstrate the feasibility of assimilating microwave radiances over sea ice areas in order to derive the sea ice concentration, as well as to provide an empirical surface emissivity model for use in those activities. In subsequent work, the trained sea ice surface emissivity model has been incorporated into the observation operator for all-sky microwave radiances in the ECMWF atmospheric data assmilation system (Geer, 2023c). The data assimilation system has been adapted so that it can estimate the sea ice concentration and the three empirical surface parameters at each observation location. This retrieves a good quality sea ice concentration as well as improving atmospheric forecasts through the ability to assimilate of observations in the vicinity of sea ice. Testing is done on different years to the training period used in this work, demonstrating the ability of the sea ice emissivity model to generalise outside the training data. As a result of this work, the assimilation of AMSR2 and GMI (GPM microwave imager) observations over sea ice will be activated in the next upgrade of the ECMWF operational weather forecasting system in 2024 (cycle 49r1).

In the future it is hoped to roll out similar hybrid empirical-physical methods for the assimilation of satellite radiances over the land surface, with the aim of inferring snow parameters, soil moisture and vegetation. There is also plenty that can be done to im-

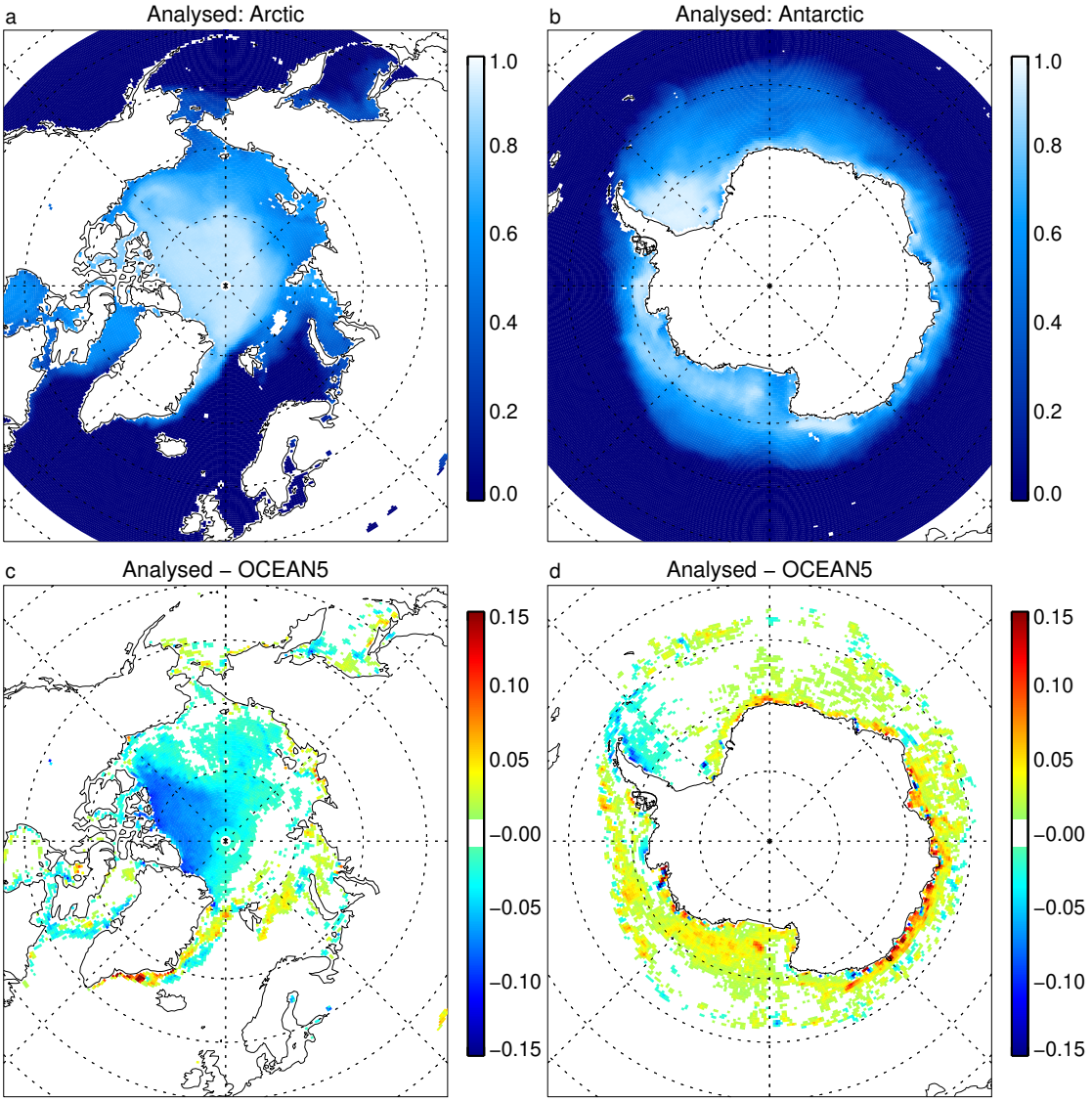

Figure 13. Annual mean sea ice concentration analysed in this work (a,b) and difference in annual mean between the current work and OCEAN5 (c,d). Differences smaller than 0.01 are not plotted.

prove this initial modelling of the sea ice radiative transfer. One aim is to extend the modelling to higher microwave and sub-mm frequencies using additional sensors. Another is to move beyond the initially crude description of the surface by an emissivity and a skin temperature, and instead to use a model which describes the known physics of radiative transfer within the snow and sea ice. In this approach, empirical state variables would still be required to describe the microphysical properties of the sea ice and snow, but the empirical model would have the more targeted responsibility of generating the optical properties that are required as input to such a model. Further, since many of the remaining errors in the network appear to come from cloud errors in the ECMWF forecast, it could be desirable to make clouds in some way a trainable parameter of the system.

On the more technical side, there are clearly many ways to improve the speed and quality of the network training. The training benefitted from early stopping to avoid overfitting, indicating that further physical constraints should ideally be added; alternatively a larger training dataset might help. The most likely routes for applying even stronger physical constraints are through adding more physics to the model (for example by including a prognostic model for sea ice), by adding observations with complementary sensitivities, and by improving the description of prior errors, particularly in the sea ice fields. There were also intriguing results (Appendix A1) that suggest big speedups could come from the use of much larger batch sizes in the training, though at the risk of overfitting. Improved access to hardware (e.g. GPUs) and the use of multi-process parallel training techniques could also help alleviate the resource problems that limited the current work.

The training of the hybrid physical-empirical model has demonstrated a number of innovations that have been facilitated by the availability of modern machine learning and differential programming tools such as Keras and TensorFlow (Abadi et al., 2015). Considering the network in Fig. 1, if the sea ice emissivity was a known parameter, the lowermost empirical part of the network (devoted to the empirical state parameters and the empirical surface emissivity model) would be unnecessary, and this work would have been a standard data assimilation problem of estimating the sea ice concentration maps from the satellite observations. However, there would still be the novelty that it was implemented in a machine learning framework. If the sea ice concentration, and the micro- and macro-physical parameters of the sea ice and snow cover, were known globally through modelling or observations, those parameters could have been treated as known input variables and the empirical sea ice emissivity model would be an ML component model trained inside an otherwise physical data assimilation network (e.g. Reichstein et al., 2019; Geer, 2021) but not achieved practically as yet, to this author's knowledge. But since the input variables are also unknown (the chicken and egg problem) the most novel aspect of this network is that it simultaneously trains an empirical model and works out what its inputs should be (the maps of sea ice parameters).

It is proposed to describe the simultaneous training of state and model as an 'empirical state' method, with the essential components being:

1. a spatially and temporally varying geophysical state that is represented statistically, using empirical parameters, here representing the macro- and micro-scale details of sea ice and its snow cover
2. an empirical model that generates a physical quantity from the empirical state, here sea ice surface emissivity at observation locations. The inputs to the empirical model define the meaning of the empirical state.

A similar approach could be extended to snow surfaces over land, soil moisture and vegetation analyses from microwave sensors, and possibly to many other 'chicken and egg' problems in the wider developments towards earth system assimilation. The approach can also be extended to a 'hybrid state', as in this work, where some of the inputs are physical and some empirical. By taking this mix of empirical and physical inputs, empirical methods can become progressively more physical, as models become able to supply more and higher quality input parameters. For example, snow grain size, depth and temperature may eventually be available within the ECMWF model, and these could be included as inputs to the empirical surface emissivity model, and the number of empirical parameters representing unknown aspects of the snow and sea ice state could hopefully be reduced. This means that empirical methods can be a quick way of getting started with a new assimilation domain, such as sea ice, but they can subsequently evolve towards more physical approaches as physical models become more capable within that domain.

Table A1. Sensitivity tests

Aspect	Tests	Standard setting
Number of epochs	Up to 300	20
Batch size	16384, 4096, 1024, 256 or 32	1024
Number of empirical variables	1 to 5, 7, 10	3
Nonlinear and deep neural networks	Nonlinear NN with 20 neurons and sigmoid activation, using 1, 2, 5 or 10 layers	1 linear layer
Sea ice fraction background error	20.0, 2.0, 0.2, 0.02, 0.002, 0.0002	0.02
Sea ice emissivity background error	1e-1,1e-2,1e-3,1e-4,1e-5,1e-6,1e-7	1e-5
TB bias background error	10.0, 1.0, 0.1, 0.01, 0.001, 0.0001, 0.00001	0.001 K

A last word goes to the Bayesian viewpoint encapsulated in the network diagram Fig. 1. All geophysical inference and forecasting problems could be represented in this way, allowing an optimal description of how observations can be used to improve our physical knowledge generally, as well as to direct that knowledge towards our goal of improved earth system forecasts. Current data assimilation generally assumes that all model components are perfectly known and only the geophysical state is unknown. There is great scope for relaxing that assumption by including empirical components, as done here, and by correctly describing the uncertainties in all the assumptions required by any physical model. On the other hand, pure machine learning techniques tend to throw away most prior knowledge and at most add back a few physical constraints. A more complete and formal description of both the known and unknown physics can come from using the Bayesian approach with mixed physical and empirical components illustrated here. Indeed Bayes' theorem suggests that the most accurate geophysical states and forecasts (the lowest posterior uncertainties) are only achievable by including as much prior knowledge as is available. This helps direct the informational power of the observations to the parts of the earth system that really need it, both the geophysical state, especially in less well-observed or chaotic parts of the system, and the model, in areas where physical models are not yet fully developed.

Appendix A Sensitivities - overview

The sensitivity tests listed in Tab. A1 were carried out to explore the robustness of the results. Since the full year's training dataset requires significant time and resources, the sensitivity tests were carried out using just the month of August 2020 for training. Settings were exactly the same as the yearly training except that 20 epochs were used by default, compared to 8 in the yearly training. The additional epochs may compensate for a training database that is roughly 12 times smaller, though on the other hand training is likely to be easier because a smaller range of geophysical conditions need to be fitted.

A1 Sensitivities - epochs, batch size, network complexity

Extending the number of epochs to 300 explores whether the training is fully converged and whether the constraints that have been imposed (such as regularisation) are complete enough. Figure A1a shows that only minor reductions in the loss term are avail-

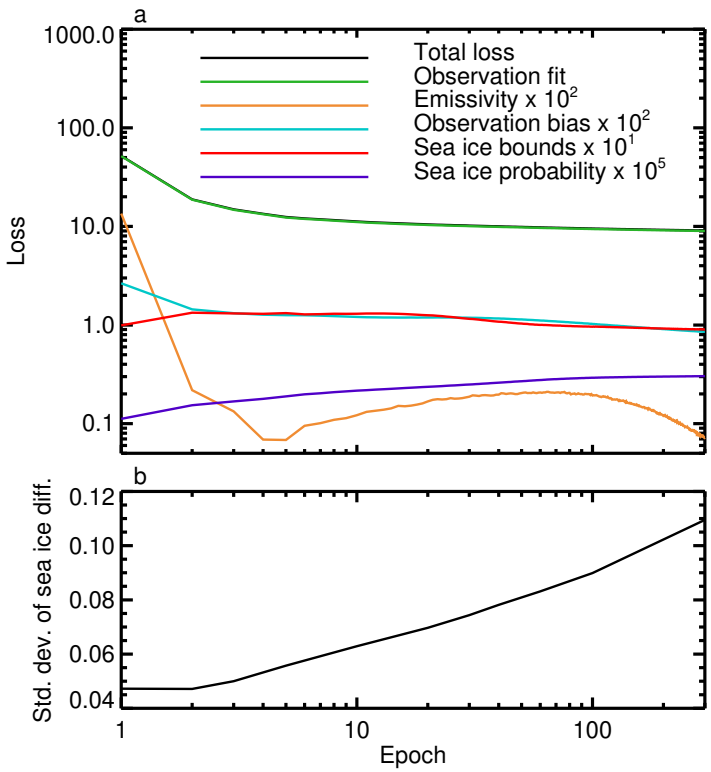

Figure A1. (a) Total loss and individual loss terms (scaled as noted in key) using an extended number of training epochs; (b) Standard deviation of the difference between analysed sea ice fraction with a 48 h delay and OCEAN5 sea ice fraction. Note the logarithmic x axis and, in panel a, also y axis, and that the total loss is hidden underneath the observation fit, which dominates.

able beyond 10 epochs. However, the sea ice probability loss term continues to increase. Further, Fig. A1b shows that the analysed sea ice fraction also continues to move away from the OCEAN5 sea ice fraction, even with a 48 h lag offset as indicated by Table 3. As shown in Sec. 3.1, some departure from the OCEAN5 results is necessary to get closer to the truth, but standard deviations heading beyond 0.1 are too large. This is easily seen in maps of sea ice fraction (not shown), and comes from a problematic increase in sea ice in areas where it is physically unlikely, as well as increasingly noisy and unrealistic looking sea ice over the Arctic and Antarctic. The problem is that if the training is left to continue too long, the network increasingly overfits cloud-related errors in the ECMWF background atmospheric state by making unphysical adjustments to the sea ice fraction. This shows that despite using two sea ice loss terms, the hybrid network is not fully constrained against creating unphysical sea ice. One fix could be to increase the weight of the sea ice probability term, in order to prevent formation of spurious sea ice over warmer seas, but a more general solution would be to allow the atmospheric terms to adjust to fit cloud errors in the ECMWF atmospheric background forecast. However, in the main results of this work, terminating the year-long training after 8 epochs has helped avoid such problems.

The impact of the number of empirical variables used to represent the sea ice and snow microphysical state is explored in Fig. A2. Adding more empirical variables always reduces the loss, with particularly significant reductions for up to 3. It is important not

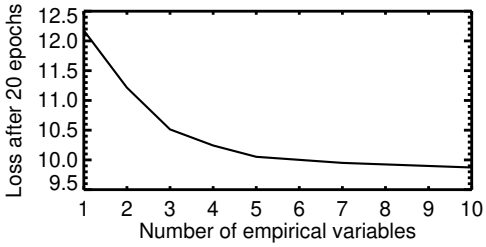

Figure A2. Total loss as a function of the number of empirical variables.

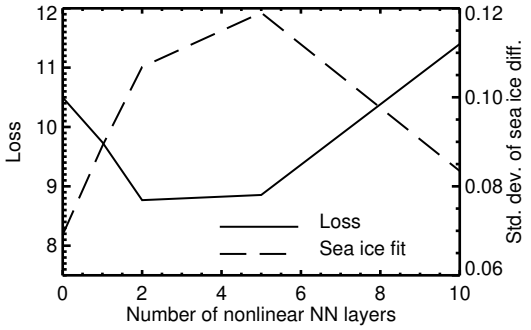

Figure A3. Total loss and sea ice fit to OCEAN5 with a 48 h offset, as a function of the number of neural network layers, based on 50 epochs of training and 20 neurons per layer. The data point for "0" layers is used to represent a single linear layer.

to allow too many variables, since by the time this reaches 10, the same as the number of satellite channels, the emissivity model becomes useless for sea ice retrievals (Sec. 2.5) and the empirical state could completely determine the surface emissivity required to fit each channel. Hence 3 appears to be a reasonable choice.

The chosen model for sea ice emissivity is linear, but the framework is perfectly capable of training a nonlinear model, including deep networks. Sensitivity tests were performed using a sigmoid activation function and between 1 and 10 fully connected layers using 20 neurons each. In these tests the number of epochs was set to 50 to ensure the deeper networks were converged. Fig. A3 shows the results in terms of the loss function at 50 epochs and the fit of the sea ice field to ECMWF sea ice (with the 2 day offset for best fit). The point marked 0 layers corresponds to the normal linear single layer model, but trained for 50 epochs. Going to a nonlinear activation function and adding up to 2 layers is capable of fitting the observations better, as indicated by the reductions in the loss function. However, this comes at the price of generating a poorer-quality sea ice field, as indicated by the increasing standard deviation of the difference with ECMWF / OCEAN5 sea ice becoming larger than 0.1. For 3 layers and greater, the picture reverses, but this is likely because 50 epochs is insufficient to fully train the deeper networks. In broad terms, adding multiple layers and nonlinearity to the surface emissivity model seems to give greater possibilities for over-fitting the data, similar to increasing the number of epochs.

A batch size of 1024 was used in this work because this was the smallest feasible batch size. Training was unfeasibly slow using any smaller batch size. Figure A4 illustrates the effect of using batch sizes from 32 (the Keras default) to 16384, noting that

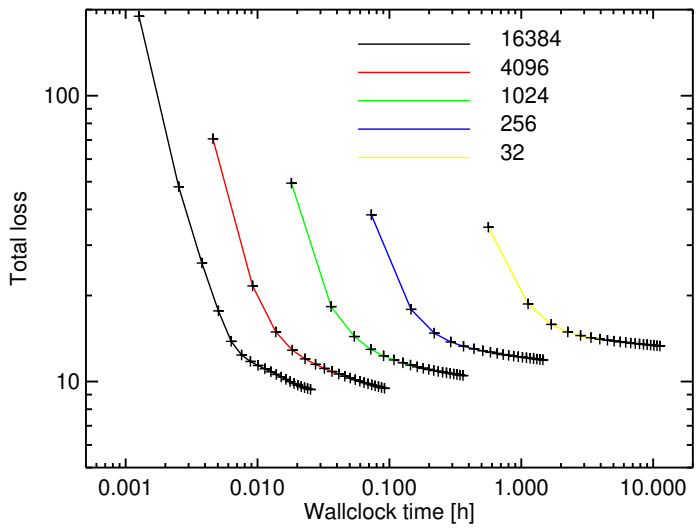

Figure A4. Total loss as a function of wallclock time for five different batch sizes, as indicated in the key. Crosses indicate the wallclock time and the corresponding loss after each of 20 epochs.

the Adam learning rate remained constant at the default 0.001 throughout. The resulting wall clock times range between 1.5 minutes and 11 hours to complete 20 epochs of training on the single-month dataset. The particularly poor performance for small batch sizes must in part be due to the need to evaluate the sea ice losses once per batch, each of which involves computations across the full month of sea ice data. Furthermore, the larger the batch size gets, the smaller the loss, and hence the better the fit to observations. The larger batch sizes also show surprisingly good fit to the OCEAN5/ECMWF sea ice, albeit with some signs of overfitting, such as some apparent inaccuracies in the PDF of sea ice fraction (no figures shown). For batch sizes between 32 and 1024, the empirical sea ice emissivity model has a broadly similar form to what is shown in Figure 4 but by the 16384 batch size it evolves to something quite different, with stronger correlations between the empirical state variables (not shown). There has not been time in the current work to fully explore the results with the larger batch sizes. In the wider machine learning community, larger (and variable) batch sizes have been advocated for speeding up the time to convergence, but with the danger that models can become less good at generalisation (e.g. Smith et al., 2017; Golmant et al., 2018). Smaller batch sizes lead to greater random variations in the solution, which may partly explain the way smaller batch sizes give larger loss values (Fig. A4) and which may act as a useful form of regularization. However, given the possible benefits in terms of time to solution and better fit to observations, an exploration of larger batch sizes could be worthwhile in future evolutions of this work. Further, adjustments to the learning rate would also be worth exploring.

A2 Sensitivities - prior knowledge loss terms

The background error setting for the sea ice PDF (c_{bgerr} in Eq. 11) controls the frequency of non-physical sea ice fraction values, as illustrated in Fig. A5. The fit to observations can be improved (and the loss function made smaller) by allowing a greater frequency of non-physical sea ice fractions (not shown) but clearly this is undesirable. In these tests based on a month of training data, it is possible to almost eliminate non-

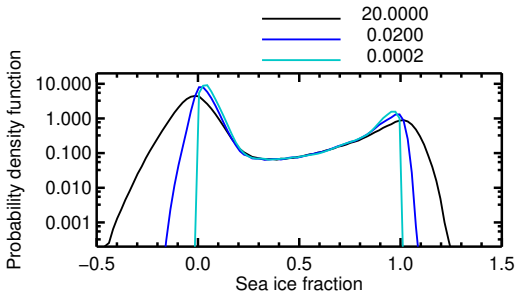

Figure A5. Probability density function of sea ice fraction for three settings of the sea ice background error, as given in the key.

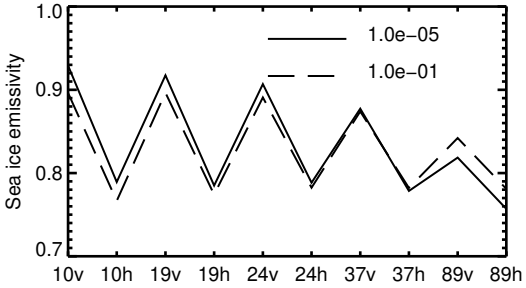

Figure A6. Sea ice surface emissivity baseline term $\mathbf{w}_b$, as a function of AMSR2 channel identifier, for values of the surface emissivity background error (w_{bgerr} in Eq. 9) as indicated in the key.

physical sea ice fractions by going to $c_{\mathrm{bgerr}} = 0.0002$ and there are no spikes in the PDF, unlike in the year-long training (Fig. 11). A more relaxed setting of $c_{\mathrm{bgerr}} = 0.02$ was needed in the year-long training because it created fewer problems with spikes in the PDF. If it had been possible to run more epochs of training, likely the spikes could have been eliminated and a tighter constraint on non-physical sea ice fractions could have been applied.

Figure A6 illustrates the effects of the sea ice emissivity background error w_{bgerr} in Eq. 9. The setting $w_{\mathrm{bgerr}} = 0.00001$ keeps the 10v emissivity very close to 0.93 as intended, but it can be relaxed with little effect on the fit to observations (not shown). The figure also shows the results for $w_{\mathrm{bgerr}} = 0.1$, which allows a drop of the baseline emissivity in most channels, to around 0.9 for 10v, and a small increase of the baseline emissivity for high frequency channels (89v and 89h). This suggests that the initial value of 0.93 for the 10v sea ice surface emissivity may have been slightly high. But these variations are within the range of available estimates (Lee et al., 2017). In the bigger picture, the results are surprisingly robust to relaxing the sea ice emissivity constraint, and future work could possibly even remove it.

Relaxing the bias background error b_{bgerr} in Eq. 4 to values larger than the chosen $b_{\mathrm{bgerr}} = 0.001\,K$ improves the fit to observations, and reduces the total loss from around 10.5 down to around 9.8 (no figure shown). However, this leads the network to generate unreasonably large bias corrections of up to -7 K over ocean and -16 K over sea ice (no figure shown). This likely shows that the problem is ill-posed when both TB bias and surface emissivity are allowed to vary. Hence the bias correction term is a very important constraint, and if prior estimates for the bias had not been available, it would

Table B1. Trained parameters of the sea ice surface emissivity model.

Channel	10v	10h	19v	19h	24v	24h	37v	37h	89v	89h
$\mathbf{w}_b$	0.9275	0.8006	0.9186	0.7958	0.9090	0.7966	0.8806	0.7816	0.8197	0.7448
$\mathbf{w}_0$	0.1286	0.1904	0.1163	0.1786	0.1097	0.1701	0.0841	0.1399	-0.0084	0.0384
$\mathbf{w}_1$	0.0598	-0.0035	-0.0105	-0.1060	-0.0660	-0.1628	-0.2223	-0.3022	-0.5666	-0.5981
$\mathbf{w}_2$	-0.0149	0.0871	-0.0752	0.0382	-0.1121	0.0038	-0.1617	-0.0461	0.0228	0.1025
$\mathbf{w}_3$	0.0569	0.2618	0.0311	0.2257	0.0062	0.1780	-0.0616	0.0756	-0.3111	-0.2304

not have been possible to estimate them simultaneously with the sea ice emissivity model. This underlines the importance of well-calibrated satellite observations and illustrates that empirical techniques still have mathematical limits that prevent them being able to infer all parameters of a system without the use of prior constraints.

A3 Sensitivities - reproducibility

The sensitivity tests described above were fully reproducible for reruns using the exact configuration of libraries and hardware described in Appendix C, by defining a fixed seed for random computations in Keras. The yearly results on which the main paper is based were not reproducible for rerun because the seed was mistakenly fixed too late, after the model creation (see the code, Geer, 2023a). However, if the seed is fixed in the right place, the yearly results can be made reproducible too. When re-run with a different seed, the output results are extremely similar in most respects (e.g. sea ice maps, output brightness temperatures) but the empirical sea ice emissivity model can change. This is mainly superficial, since similar sea ice emissivity structures are created but in a different order to what is seen in Fig. 4. But this highlights the fact that the meanings of the empirical variables are not entirely fixed and are naturally affected by random processes during the training.

Appendix B Surface emissivity model parameters

The trained sea ice surface emissivity model parameters are given in Tab. B1.

Appendix C Open Research

The data (Geer, 2023b) are available on Zenodo through the ECMWF open data license `https://apps.ecmwf.int/datasets/licences/general/` noting the conditions for the underlying AMSR2 data described below. Data is copyright 2023 ECMWF and is published under a Creative Commons Attribution 4.0 International (CC BY 4.0, `https://creativecommons.org/licenses/by/4.0/`). In applying this license, ECMWF does not accept any liability whatsoever for any error or omission in the data, their availability, or for any loss or damage arising from their use.

The code (Geer, 2023a) is on GitHub and is copyright 2023 ECMWF and is licensed under the Apache License, Version 2.0 (`http://www.apache.org/licenses/LICENSE-2.0`). In applying this licence, ECMWF does not waive the privileges and immunities granted to it by virtue of its status as an intergovernmental organisation nor does it submit to any jurisdiction.

Original AMSR2 data for this value added data product was provided by Japan Aerospace Exploration Agency (JAXA, 2023) via the G-Portal. The user is entitled to use JAXA G-Portal AMSR2 data free of charge without any restrictions (including commercial use) except for the condition about acknowledgement of data credit as stipulated in Article 7.(2) of the terms and conditions at `https://gportal.jaxa.jp/gpr/index/eula?lang=en`

The Python code was run on Python 3.8.8-01 (Python Software Foundation, 2021) including Tensorflow and Keras 2.8.0 (Abadi et al., 2015; TensorFlow, 2021) on the ECMWF ATOS supercomputer CPU nodes.

Acknowledgments

Niels Bormann, Fabrizio Baordo, Matthew Chantry and Tracy Scanlon are thanked for reviews and suggestions.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., ... Zheng, X. (2015). *TensorFlow: Large-scale machine learning on heterogeneous systems* [code]. Retrieved from `https://www.tensorflow.org/` (Software available from tensorflow.org) doi: 10.5281/zenodo.4724125

Baordo, F., & Geer, A. (2015). Microwave surface emissivity over sea-ice. *EUMETSAT NWP-SAF visiting scientist report NWPSAF_EC_VS_026*. Retrieved from `https://nwpsaf.eu/publications/vs\_reports/nwpsaf-ec-vs-026.pdf`

Baordo, F., & Geer, A. J. (2016). Assimilation of SSMIS humidity-sounding channels in all-sky conditions over land using a dynamic emissivity retrieval. *Quart. J. Roy. Meteorol. Soc.*, *142*, 2854–2866. Retrieved from `https://doi.org/10.1002/qj.2873`

Bauer, P., Moreau, E., Chevallier, F., & O'Keeffe, U. (2006). Multiple-scattering microwave radiative transfer for data assimilation applications. *Quart. J. Roy. Meteorol. Soc.*, *132*, 1259-1281. Retrieved from `https://doi.org/10.1256/qj.05.153`

Bauer, P., Thorpe, A., & Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*, *525*(7567), 47–55.

Berg, W., Bilanow, S., Chen, R., Datta, S., Draper, D., Ebrahimi, H., ... Yang, J. X. (2016). Intercalibration of the GPM microwave radiometer constellation. *Journal of Atmospheric and Oceanic Technology*, *33*(12), 2639–2654. Retrieved from `https://doi.org/10.1175/JTECH-D-16-0100.1`

Bocquet, M., Brajard, J., Carrassi, A., & Bertino, L. (2019). Data assimilation as a learning tool to infer ordinary differential equation representations of dynamical models. *Nonlinear Processes in Geophysics*, *26*(3), 143–162.

Bocquet, M., Brajard, J., Carrassi, A., & Bertino, L. (2020). Bayesian inference of chaotic dynamics by merging data assimilation, machine learning and expectation-maximization. *Foundations of Data Science*, *2*(1), 55–80. Retrieved from `https://doi.org/10.3934/fods.2020004`

Brajard, J., Carrassi, A., Bocquet, M., & Bertino, L. (2020). Combining data assimilation and machine learning to emulate a dynamical model from sparse and noisy observations: A case study with the lorenz 96 model. *Journal of Computational Science*, *44*, 101171.

Browne, P. A., de Rosnay, P., Zuo, H., Bennett, A., & Dawson, A. (2019). Weakly coupled ocean–atmosphere data assimilation in the ECMWF NWP system. *Remote Sensing*, *11*(3), 234. Retrieved from `https://doi.org/10.3390/rs11030234`

Buehner, M., Caya, A., Pogson, L., Carrieres, T., & Pestieau, P. (2013). A new environment canada regional ice analysis system. *Atmosphere-Ocean*, *51*(1), 18–34.

Retrieved from https://doi.org/10.1080/07055900.2012.747171

Chollet, F., et al. (2015). *Keras.* https://keras.io.

Comiso, J. C., Cavalieri, D. J., & Markus, T. (2003). Sea ice concentration, ice temperature, and snow depth using AMSR-E data. *IEEE Trans. Geosci. Remote Sens.*, *41*(2), 243–252. Retrieved from https://doi.org/10.1109/TGRS.2002.808317

de Rosnay, P., Browne, P., de Boisséson, E., Fairbairn, D., Hirahara, Y., Ochi, K., ... Rabier, F. (2022). Coupled assimilation at ECMWF: current status, challenges and future developments. *Quart. J. Roy. Meteorol. Soc.*, *148*, 2672—2702. Retrieved from https://doi.org/10.1002/qj.4330

Desroziers, G., Berre, L., Chapnik, B., & Poli, P. (2005). Diagnosis of observation, background and analysis-error statistics in observation space. *Quart. J. Roy. Meteorol. Soc.*, *131*, 3385–3396.

Eyre, J. R., English, S. J., & Forsythe, M. (2020). Assimilation of satellite data in numerical weather prediction. Part I: The early years. *Quart. J. Roy. Meteorol. Soc.*, *146*(726), 49–68. Retrieved from https:/doi.org/10.1002/qj.3654

Geer, A. J. (2021). Learning earth system models from observations: Machine learning or data assimilation? *Phil. Trans. R. Soc. A*, *379*, 20200089. Retrieved from https://doi.org/10.1098/rsta.2020.0089

Geer, A. J. (2023a). *Code for simultaneous inference of sea ice state and surface emissivity model using machine learning and data assimilation* [code]. Retrieved from https://doi.org/10.5281/zenodo.10013542

Geer, A. J. (2023b). *Data for simultaneous inference of sea ice state and surface emissivity model using machine learning and data assimilation* [dataset]. Retrieved from https://doi.org/10.5281/zenodo.10033377

Geer, A. J. (2023c). Joint estimation of sea ice and atmospheric state from microwave imagers in operational weather forecasting. , in preparation.

Geer, A. J., Baordo, F., Bormann, N., English, S., Kazumori, M., Lawrence, H., ... Lupu, C. (2017). The growing impact of satellite observations sensitive to humidity, cloud and precipitation. *Quart. J. Roy. Meteorol. Soc.*, *143*, 3189-3206. Retrieved from https:/doi.org/10.1002/qj.3172

Geer, A. J., & Bauer, P. (2011). Observation errors in all-sky data assimilation. *Quart. J. Roy. Meteorol. Soc.*, *137*, 2024–2037. Retrieved from https://doi.org/10.1002/qj.830

Geer, A. J., Bauer, P., & O'Dell, C. W. (2009). A revised cloud overlap scheme for fast microwave radiative transfer. *J. App. Meteor. Clim.*, *48*, 2257–2270. Retrieved from https://doi.org/10.1175/2009JAMC2170.1

Geer, A. J., Lonitz, K., Duncan, D., & Bormann, N. (2022). *Improved surface treatment for all-sky microwave observations* (Tech. Memo. No. 894). Reading, UK: ECMWF. Retrieved from https:/doi.org/10.21957/zi7q6hau

Geer, A. J., Lonitz, K., Weston, P., Kazumori, M., Okamoto, K., Zhu, Y., ... Schraff, C. (2018). All-sky satellite data assimilation at operational weather forecasting centres. *Quart. J. Roy. Meteorol. Soc.*, *144*, 1191–1217. doi: 10.1002/qj.3202

Gettelman, A., Geer, A. J., Forbes, R. M., Carmichael, G. R., Feingold, G., Posselt, D. J., ... Zuidema, P. (2022). The future of earth system prediction: Advances in model-data fusion. *Science Advances*, *8*(14), eabn3488.

Ghahramani, Z. (2015). Probabilistic machine learning and artificial intelligence. *Nature*, *521*(7553), 452–459.

Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics* (pp. 249–256).

Golmant, N., Vemuri, N., Yao, Z., Feinberg, V., Gholami, A., Rothauge, K., ... Gonzalez, J. (2018). On the computational inefficiency of large batch sizes for stochastic gradient descent. *arXiv preprint arXiv:1811.12941*.

Good, S., Fiedler, E., Mao, C., Martin, M. J., Maycock, A., Reid, R., ... others (2020). The current configuration of the OSTIA system for operational production of foundation sea surface temperature and ice concentration analyses. *Remote Sensing*, *12*(4), 720. Retrieved from https://doi.org/10.3390/rs12040720

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., ... Thépaut, J. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*. Retrieved from https://doi.org/10.1002/qj.3803

Hewison, T. J., & English, S. J. (1999). Airborne retrievals of snow and ice surface emissivity at millimeter wavelengths. *IEEE Trans. Geosci. Rem. Sens.*, *37*(4), 1871–1879.

Hirahara, Y., Rosnay, P. d., & Arduini, G. (2020). Evaluation of a microwave emissivity module for snow covered area with CMEM in the ECMWF integrated forecasting system. *Remote Sensing*, *12*(18), 2946. Retrieved from https://doi.org/10.3390/rs12182946

Hortal, M., & Simmons, A. (1991). Use of reduced Gaussian grids in spectral models. *Monthly Weather Review*, *119*(4), 1057–1074. Retrieved from https://doi.org/10.1175/1520-0493(1991)119<1057:UORGGI>2.0.CO;2

Hsieh, W. W., & Tang, B. (1998). Applying neural network models to prediction and data analysis in meteorology and oceanography. *Bulletin of the American Meteorological Society*, *79*(9), 1855–1870.

JAXA. (2023). *Global Change Observation Mission – Water Advanced Microwave Scanning Radiometer 2 level 1B* [dataset]. Retrieved from https://doi.org/10.57746/EO.01gs73ans548qghaknzdjyxd2h

Kang, E.-J., Sohn, B.-J., Tonboe, R. T., Noh, Y.-C., Kwon, I.-H., Kim, S.-W., ... Liu, C. (2023). Explicitly determined sea ice emissivity and emission temperature over the arctic for surface-sensitive microwave channels. *Quart. J. Roy. Meteorol. Soc.*. Retrieved from https://doi.org/10.1002/qj.4492

Kazumori, M., & English, S. J. (2015). Use of the ocean surface wind direction signal in microwave radiance assimilation. *Quarterly Journal of the Royal Meteorological Society*, *141*(689), 1354–1375. Retrieved from https://doi.org/10.1002/qj.2445

Kazumori, M., Geer, A. J., & English, S. J. (2016). Effects of all-sky assimilation of GCOM-W/AMSR2 radiances in the ECMWF numerical weather prediction system. *Quart. J. Roy. Meteorol. Soc.*, *142*, 721–737. Retrieved from https://doi.org/10.1002/qj.2669

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Pritzel, A., ... Battaglia, P. (2022). GraphCast: Learning skillful medium-range global weather forecasting. *arXiv preprint arXiv:2212.12794*.

Lee, S.-M., Sohn, B.-J., & Kim, S.-J. (2017). Differentiating between first-year and multiyear sea ice in the arctic using microwave-retrieved ice emissivities. *J. Geophys. Res.: Atmos.*, *122*(10), 5097–5112. Retrieved from https://doi.org/10.1002/2016JD026275

Leinonen, J., Grazioli, J., & Berne, A. (2021). Reconstruction of the mass and geometry of snowfall particles from multi-angle snowflake camera (MASC) images. *Atmos. Meas. Tech.*, *14*(10), 6851–6866.

Okuyama, A., & Imaoka, K. (2015). Intercalibration of advanced microwave scanning radiometer-2 (AMSR2) brightness temperature. *IEEE Transactions on Geoscience and Remote Sensing*, *53*(8), 4568–4577. Retrieved from https://doi.org/10.1109/TGRS.2015.2402204

OSI-SAF, E. (2016). *Global sea ice concentration (SSMIS)* (Tech. Rep. No. OSI-401-b). Retrieved from https://doi.org/10.15770/EUM_SAF_OSI_NRT_2004

Peubey, C., & McNally, A. P. (2009). Characterization of the impact of geostationary clear-sky radiances on wind analyses in a 4D-Var context. *Quart. J. Roy. Meteorol. Soc.*, *135*, 1863 – 1876.

Picard, G., Sandells, M., & Löwe, H. (2018). SMRT: An active–passive microwave radiative transfer model for snow with multiple microstructure and scattering formulations (v1. 0). *Geoscientific Model Development*, *11*(7), 2763–2788.

Python Software Foundation. (2021). *Python* [code]. Retrieved from `https://github.com/python/cpython/releases/tag/v3.8.8rc1`

Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2017). Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations. *arXiv preprint arXiv:1711.10561*.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., et al. (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204.

Rodgers, C. D. (2000). *Inverse methods for atmospheric sounding: Theory and practice* (1st ed.). Singapore: World Scientific.

Sandells, M., Rutter, N., Wivell, K., Essery, R., Fox, S., Harlow, C., ... Toose, P. (2023). Simulation of arctic snow microwave emission in surface-sensitive atmosphere channels. *EGUsphere*, *2023*, 1–28. Retrieved from `https://egusphere.copernicus.org/preprints/2023/egusphere-2023-696/` doi: 10.5194/egusphere-2023-696

Sandven, S., Spreen, G., Heygster, G., Girard-Ardhuin, F., Farrell, S. L., Dierking, W., & Allard, R. A. (2023). Sea ice remote sensing—recent developments in methods and climate data sets. *Surveys in Geophysics*, 1–37. Retrieved from `https://doi.org/10.1007/s10712-023-09781-0`

Scarlat, R. C., Spreen, G., Heygster, G., Huntemann, M., Paţilea, C., Pedersen, L. T., & Saldo, R. (2020). Sea ice and atmospheric parameter retrieval from satellite microwave radiometers: Synergy of amsr2 and smos compared with the cimr candidate mission. *Journal of Geophysical Research: Oceans*, *125*(3), e2019JC015749. Retrieved from `https://doi.org/10.1029/2019JC015749`

Schneider, T., Lan, S., Stuart, A., & Teixeira, J. (2017). Earth system modeling 2.0: A blueprint for models that learn from observations and targeted high-resolution simulations. *Geophys. Res. Let.*, *44*(24), 12–396. Retrieved from `https://doi.org/10.1002/2017GL076101`

Smith, S. L., Kindermans, P.-J., Ying, C., & Le, Q. V. (2017). Don't decay the learning rate, increase the batch size. *arXiv preprint arXiv:1711.00489*.

Spreen, G., Kaleschke, L., & Heygster, G. (2008). Sea ice remote sensing using AMSR-E 89-GHz channels. *J. Geophys. Res.: Oceans*, *113*(C2), C02S03. Retrieved from `https://doi.org/10.1029/2005JC003384`

TensorFlow. (2021). *Tensorflow version 2.8.0* [code]. Retrieved from `https://doi.org/10.5281/zenodo.5799851`

Timmermann, R., Goosse, H., Madec, G., Fichefet, T., Ethe, C., & Duliere, V. (2005). On the representation of high latitude processes in the ORCA-LIM global coupled sea ice–ocean model. *Ocean Modelling*, *8*(1-2), 175–201. Retrieved from `https://doi.org/10.1016/j.ocemod.2003.12.009`

Zuo, H., Balmaseda, M. A., Tietsche, S., Mogensen, K., & Mayer, M. (2019). The ECMWF operational ensemble reanalysis–analysis system for ocean and sea ice: a description of the system and assessment. *Ocean science*, *15*(3), 779–808. Retrieved from `https://doi.org/10.5194/os-15-779-2019`