

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

Alan Jon Geer¹

¹European Centre for Medium Range Weather Forecasts

November 8, 2023

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.

1
2
3
4
5
6
7
8
9
10
11
12

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

Alan Geer¹

¹European 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

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

69 The aim of this work is to build tools and techniques for direct assimilation of satel-
70 lite microwave radiances for the cryosphere and in particular, the sea ice. This work is
71 ultimately intended for use in an operational weather forecasting system, to allow the
72 assimilation of microwave radiances over sea ice surfaces and to contribute to analyses
73 of the sea ice concentration. Microwave radiances have a strong sensitivity to sea ice, par-
74 ticularly its fractional coverage but also ice age, snow depth and snow microstructure.
75 Operational methods to retrieve the sea ice concentration use largely heuristic (‘tie point’)
76 approaches (e.g. Comiso et al., 2003; Spreen et al., 2008; Sandven et al., 2023) and an
77 experimental method using optimal estimation used a simplified forward model (e.g. Scar-
78 lat et al., 2020). This is because physical modelling of snow and ice radiative transfer
79 at microwave frequencies is difficult, with better results below 20 GHz than above (e.g.
80 Hirahara et al., 2020). A fundamental issue is the importance of centimetre to micron
81 scale details of the snow and ice, including air inclusions in ice or grain shapes and their
82 layout in snow, that are relevant to the radiative transfer. Physical models of snow ra-
83 diative transfer (e.g. Picard et al., 2018) are starting to provide convincing results up
84 to around 250 GHz over land, for example linking measurements from snow pit profiles
85 to aircraft radiance measurements with reasonable accuracy in late winter in the Cana-
86 dian Arctic (Sandells et al., 2023). Physical models for sea ice have also shown poten-
87 tial (Kang et al., 2023), though with the necessity of empirically adjusting the snow and
88 sea ice optical properties to fit each location. For global assimilation of snow and ice sur-
89 faces, observation models will need to be reliable across all seasons and hemispheres, and
90 will need to handle many different types of ice and snow, including wet and refrozen snow
91 in the warmer seasons. There is also the problem of defining the input state, since it is
92 impossible to provide snow pits and ice cores globally.

93 An alternative path to improving our modelling frameworks is to use the vast amount
94 of existing observational data to improve models in an empirical way (e.g. Schneider et
95 al., 2017; Geer, 2021; Gettelman et al., 2022). The idea of training models from obser-
96 vations has become more feasible in recent years following the advent of easy-to-use ma-
97 chine learning tools such as Keras (Chollet et al., 2015) and rapid progress has been made
98 in machine learning alternatives to weather forecasting models (e.g. Lam et al., 2022).
99 A partly empirical approach is taken in the current work, with the aim to train a new
100 observation model for sea ice using global data through the whole year, in order to han-
101 dle all seasons, hemispheres and snow and ice types. However, empirical geophysical mod-
102 els are normally built using ‘supervised learning’ strategies that require the inputs and
103 outputs of the empirical model to be known and available in large quantities. For ex-
104 ample, in the case of ML-based weather forecasting models, this relies on existing atmo-
105 spheric re-analysis datasets (e.g. Hersbach et al., 2020). But to train an empirical ob-
106 servation model to link the geophysical state to the satellite radiances for the land sur-
107 face or the cryosphere, the inputs, in other words the state of the soil, snow and ice, are
108 not well enough known on a global scale. In fact, we will likely only know the state pa-
109 rameters well enough on a global scale once an observation model is available to help de-
110 rive them from satellite observations, so training an empirical model for surface radia-
111 tive transfer can be seen as a chicken and egg problem.

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

117 scribe underlying statistical properties of, for example, the output snowflake shapes. A
 118 similar principle is used in the current work, which allows the empirical observation model
 119 to define its own latent space as its input, in order to represent the mostly unknown mi-
 120 crophysical structures of sea ice and snow that are needed to simulate microwave radi-
 121 ances. This will be referred to as an ‘empirical state’, which will be trained simultane-
 122 ously with the empirical model to go from the state to the observations. The meaning
 123 of the empirical state variables is learnt along with the empirical model.

124 However, a purely empirical framework is of no use if we want to infer geophysi-
 125 cal properties, not empirical variables. To solve this, we can impose known physical re-
 126 lationships, such as the radiative transfer of the atmosphere. Reichstein et al. (2019) de-
 127 scribed how empirical models could be used and trained in networks with physical mod-
 128 els. These networks can be represented in a Bayesian way that is mathematically equiv-
 129 alent to data assimilation (Geer, 2021). The Bayesian viewpoint brings the insight that
 130 physical models represent parts of the network where there is good prior physical knowl-
 131 edge, whereas machine learning models can represent areas where there is little or no prior
 132 knowledge. In the current work, the empirical modelling is limited to the surface emis-
 133 sivity of the sea ice, whereas physical modelling is used for the surface emissivity of the
 134 ocean and for the radiative transfer of the atmosphere. In particular, it is the inclusion
 135 of physical equations that allows the sea ice concentration to be derived within the sys-
 136 tem, but the equations would have been incomplete without the empirical contribution.
 137 The entire network can be referred to as a hybrid empirical-physical model.

138 A number of other aspects of geophysical modelling and observation are more well-
 139 developed in data assimilation frameworks than in typical machine learning methods (Geer,
 140 2021). One of these is mapping between geophysical fields on a regular time and space
 141 grid to the irregular arrangement of satellite observations in time and space, which is han-
 142 dled by an interpolation operator. Also used is background error modelling (to control
 143 the weight of prior physical knowledge against the new knowledge from observations) and
 144 observation errors (to account for the varying difficulty of forward modelling at differ-
 145 ent frequencies measured by the satellite). All these techniques are also employed in the
 146 current work, making a mix of data assimilation and machine learning, following much
 147 work merging the two fields (e.g. Hsieh & Tang, 1998; Bocquet et al., 2020; Brajard et
 148 al., 2020).

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

159 2 Method

160 2.1 Overview

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

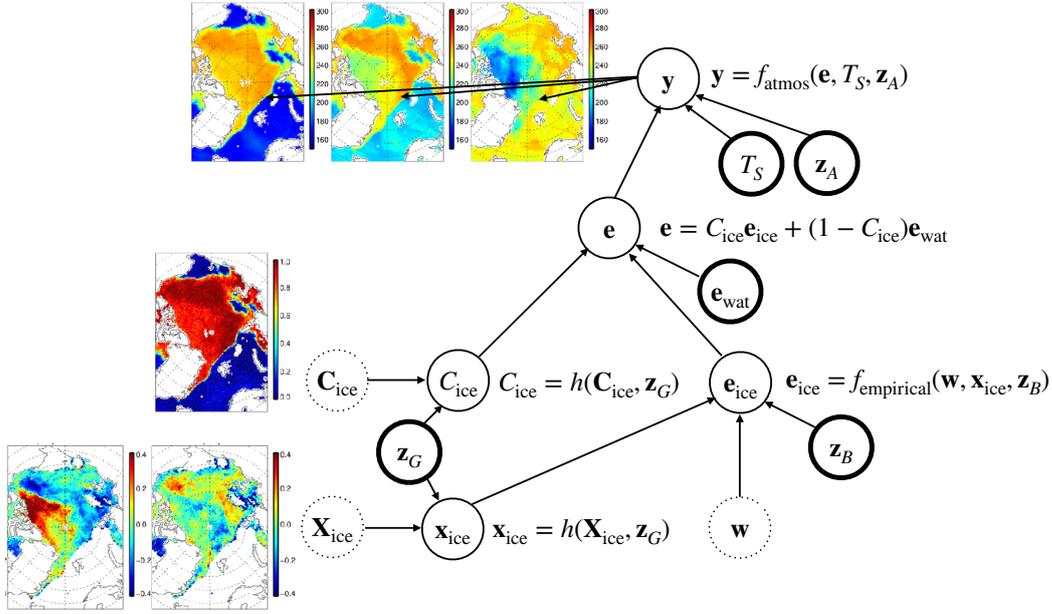


Figure 1. Simplified overview of the hybrid empirical-physical training network, for a single day and a single multi-channel observation. Circles represent variables that are trainable (dotted), dependent (thin solid), or fixed (thick solid). Arrows indicate dependencies and equations give the functional 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.

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

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

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

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

191 The hybrid network is trained by minimising a loss function. The aim is to be able
192 to replicate as closely as possible a year of Advanced Microwave Sounding Radiometer-
193 2 (AMSR2) observations over ocean and sea ice areas poleward of 45° latitude. The sim-
194 ulated observation \mathbf{y} (Fig. 1) represents an observation at one location comprising mea-
195 surements at different frequencies and polarisations, known as channels. The notation
196 here is that variables in lower-case bold represent vectors. The measured quantity is ra-
197 diance and this is given in terms of equivalent black body brightness temperature (TB).
198 Further detail on the microwave observations is in Sec. 22.3. The observed AMSR2 ob-
199 servations are \mathbf{y}_{obs} and the discrepancy with the simulations is measured by a mean squared
200 error loss function J_{obs} , which weights observations by an observation error \mathbf{r} following
201 standard DA practice:

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

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

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

222 The mixed surface emissivity \mathbf{e} must be estimated within the network to provide
223 the surface boundary conditions to the atmospheric radiative transfer model. The emis-
224 sivity is described as a linear combination of the sea ice and ocean water surface emis-
225 sivities \mathbf{e}_{ice} and \mathbf{e}_{wat} , weighted by the sea ice concentration in the satellite’s field of view,
226 C_{ice} . The contrast in the surface emissivity between ocean and ice is strongest and eas-
227 iest to detect at microwave wavelengths, and is fundamentally what allows the inference
228 of the sea ice concentration from satellite observations. This equation also defines
229 the meaning of sea ice concentration in this work, which is the fraction of the ocean sur-
230 face covered with ice or snow, which excludes melt ponds but can include ice bergs and
231 ice shelves. The ocean surface emissivity is described physically and is treated as a known
232 quantity. Not included in the simplified diagram is a windspeed error correction that is
233 applied to the ocean surface emissivity, which uses 10 trainable weights. This correction
234 is not a major aspect of the problem and is further described in Sec. 2.5.

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

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

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

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

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

275 2.2 Technical summary

276 Figure 1 represents the network for one single observation, but the training is done
 277 on approximately 0.6 billion observed radiances across one year and with approximately
 278 4.9 billion pieces of supporting information extracted from ECMWF short-range fore-
 279 casts relating to the atmospheric radiative transfer terms and the surface temperature.
 280 Table 1 summarises the technical details of the network and its training.

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

$$J = J_{\text{obs}} + J_{\text{seaice_bounds}} + J_{\text{seaice_tsfc}} + J_{\text{emis}} + J_{\text{bias}}. \quad (2)$$

283 The sea ice concentration is constrained by two loss functions representing physical con-
 284 straints. The first sea ice loss function, $J_{\text{seaice_bounds}}$, imposes the limits zero and 1 and
 285 the second, $J_{\text{seaice_tsfc}}$, imposes the climatological probability of sea ice as a function of
 286 surface temperature (Sec. 2.7). The bias correction requires a loss function J_{bias} (Sec. 2.3)
 287 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 2.7)	365 or 366
Trainable parameters	Sea ice map \mathbf{C}_{ice} (366 days, see Sec. 2.7)	22,874,634
	Empirical micro- and macro-physical properties \mathbf{X}_{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}_{emis}	10
	TOTAL	91,311,120
	Fixed parameters	Atmospheric radiative transfer \mathbf{z}_A
Surface (skin temperature (T_S, \mathbf{z}_B), surface wind-speed, ocean emissivity \mathbf{e}_{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	

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

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

302 Training of the hybrid model and geographical maps was done on the ECMWF super-
303 computer 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

304 maximum 48 h of wallclock time (no GPUs were available). This allowed up to 8 train-
305 ing epochs, which was sufficient in the current work. To train the model, the loss func-
306 tion (Eq. 2) was minimised using the Adam variant of the mini-batch stochastic gradi-
307 ent descent approach (Kingma & Ba, 2014) at its default settings, including its learn-
308 ing rate set to 0.001, with the exception of the use of a batch size of 1024 (Appendix A1).
309 From the point of view of Keras training, the fixed parameters were treated as input val-
310 ues (features), the observations were treated as output values (labels). Further techni-
311 cal aspects of the training and hyperparameters are explored in Appendix A1.

312 2.3 Observations

313 Observations are provided by AMSR2 (Okuyama & Imaoka, 2015), a conical scan-
314 ning microwave radiometer on the polar-orbiting GCOM-W2 satellite observing at fre-
315 quencies from 6.7 GHz to 89 GHz, each separately measuring vertically (v) and horizon-
316 tally (h) polarised radiances (Tab. 2). The original radiance observations were obtained
317 from the Japan Aerospace Exploitation Agency (JAXA, <https://gportal.jaxa.com>) and
318 averaged onto a 40 km reduced Gaussian grid (Hortal & Simmons, 1991). This averag-
319 ing standardises the measurement footprint, which varies with frequency (the footprint
320 varies both in size, Table 2, but also in central location by up to around 4 km). The spa-
321 tial locations of the resulting ‘superobservations’ or superobs are taken as those of the
322 centres of the Gaussian grid points. The times of the superobs are the mean of the times
323 of the original observations, which span only a few seconds, so the superob remains an
324 almost instantaneous observation of the earth. AMSR2 also has channels at 6 – 7 GHz
325 with excellent sea ice sensitivity but these have been left out due to their footprint be-
326 ing larger than the 40 km grid. Superobs are based on an average of 20 raw observations;
327 those based on less than 6 raw observations are discarded. Following current practice
328 in atmospheric data assimilation (e.g. Kazumori et al., 2016; Geer et al., 2018), the ob-
329 servations are used in all-sky conditions, i.e. clear, cloudy and precipitating.

330 Figure 2 illustrates these observations over the Arctic ocean. At 10 GHz, v-polarised,
331 (10v) the sea ice areas are relatively easily distinguished from ocean by the large con-
332 trast in brightness temperature, with sea ice showing a fairly uniform value around 250 K.
333 But at higher frequencies (19v – 89v) and in the horizontally-polarised channels (10h –
334 89h) the sea ice areas show more strongly variable brightness temperature, driven by the
335 micro and macrophysical characteristics of the sea ice and snow cover. The influence of
336 the atmosphere (e.g. clouds and water vapour) is also important, particularly at 89 GHz.

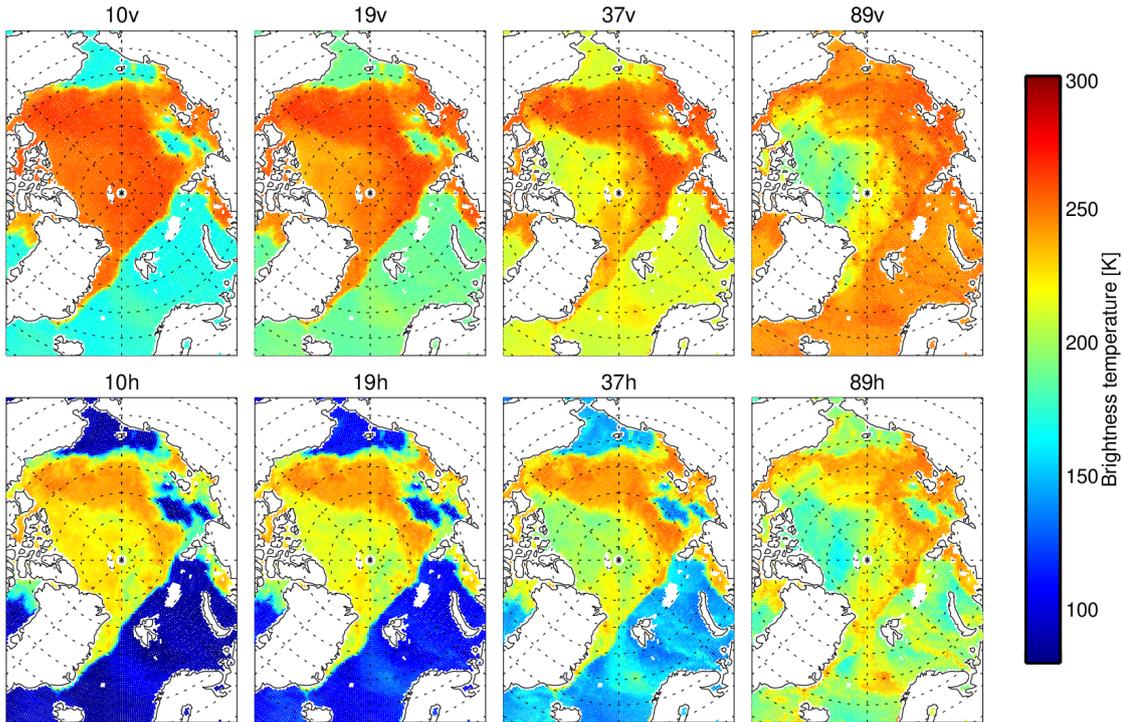


Figure 2. Observed AMSR2 brightness temperatures on the 40 km grid over the Arctic on 7th November 2020, showing only pure ocean scenes, and allowing observations from multiple orbits to overlay each other. Channels 24v and 24h are also used in this work, but are omitted from the figure to save space.

337 The ocean surface generally has much lower brightness temperatures than the sea ice and
 338 is more strongly polarised, so that h-polarisation measurements have much lower bright-
 339 ness temperatures than v-polarisation measurements, down to around 85 K at 10h. Ex-
 340 isting heuristic sea ice algorithms use higher frequencies to obtain better spatial reso-
 341 lution, and are based on empirically observed characteristics of the polarisation and fre-
 342 quency dependence of ocean and sea ice surfaces. These include methods using 19v, 37v,
 343 19h and 37h (e.g. Comiso et al., 2003; OSI-SAF, 2016) and those using 89v and 89h (e.g.
 344 Spreen et al., 2008) but in general no existing algorithm uses a physical description of
 345 the problem like the hybrid network used here (Fig. 1). This is mainly due to the lack
 346 of knowledge of the sea ice emissivity and its variations with underlying micro- and macro-
 347 physical ice and snow characteristics. The heuristic approaches can be vulnerable to at-
 348 mospheric effects such as cloud and water vapour along with variations in sea ice and
 349 ocean surface characteristics, which could incorrectly be interpreted as variations in sea
 350 ice concentration.

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

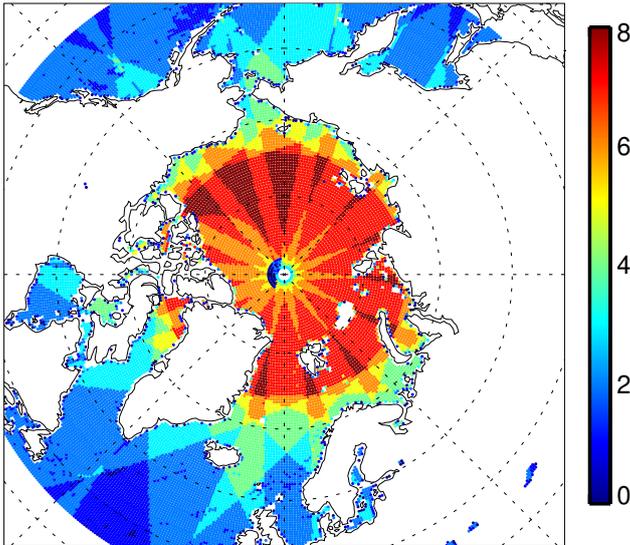


Figure 3. Number of AMSR2 pure ocean superobs available on the 40 km grid over the Arctic on 7th November 2020, in the 24 h period from 2100 UTC on 6th November to 2100 UTC on 7th November. Dashed lines indicate latitude every 10° and longitude every 45° .

360 the sea ice grid are on the same standardised locations. The result is that the daily es-
361 timates of sea ice concentration, and its the empirical properties, typically have to fit be-
362 tween 4 and 8 AMSR2 superobs and represent, in rough terms, a daily average.

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

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

368 The bias corrections are a function of the sea ice concentration, and are per channel, so
369 there is a vector of 10 bias corrections for for sea ice, \mathbf{b}_{ice} , and similarly for ocean wa-
370 ter, \mathbf{b}_{wat} . These are also trainable parameters, though in practice they are not allowed
371 to vary much. The bias correction is initialised with the background values $b_{\text{ice,bg}}$ and
372 $b_{\text{wat,bg}}$, respectively set to 2.5 K and 5.0 K (in all channels) using prior estimates of AMSR2
373 biases over ocean and land surfaces (Geer et al., 2022) and assuming that sea ice sur-
374 faces will have similar biases to those seen over land surfaces. Given that the bias cor-
375 rections are trainable values, they are constrained by a loss term

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

376 Here, the per-channel bias corrections are $b_{\text{ice},j}$ and $b_{\text{wat},j}$ with a background error of $b_{\text{bgerr}} =$
377 0.001 K, and the division by n , the number of training observations, standardises the loss
378 function to the Keras approach (Eq. 1). Given the very small chosen background error,
379 in practice the bias is forced to stay extremely close to prior values, rather than being
380 allowed to evolve to fit the observations. Hence the bias loss term exists mainly to ex-
381 plore the possibility of relaxing this constraint and evolving the bias correction in the
382 sensitivity tests described in A2. These show that if biases are allowed to evolve away

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

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

402 2.4 Atmospheric radiative transfer

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

$$y_{jk} = e_j T_S \Gamma_{jk} + (1 - e_j) T_{jk}^{\downarrow} \Gamma_{jk} + T_{jk}^{\uparrow}. \quad (5)$$

414 Here, T_{jk}^{\downarrow} is the downwelling radiation (TB) at the surface, T_{jk}^{\uparrow} is the component of up-
 415 welling radiation at the top of the atmosphere coming from the atmosphere itself, and
 416 Γ_{jk} is the atmospheric transmittance. The surface is represented using the approxima-
 417 tion of specular reflection, where the surface emits radiation according to the surface tem-
 418 perature T_S multiplied by the surface emissivity e_j , and reflects downwelling radiation
 419 modulated by a reflectivity $1 - e_j$. This is not a perfect assumption, because microwave
 420 radiation can penetrate centimetres to metres into snow and ice surfaces, and future work
 421 will use a more physical representation of the radiative transfer within the sea ice and
 422 snow (e.g. Picard et al., 2018). The final ‘all-sky’ brightness temperature is obtained by
 423 weighting the two sub-columns according to the effective cloud fraction C_{eff} (Geer et al.,
 424 2009):

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

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

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

441 2.5 Surface emissivity model

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

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

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

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

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

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

463 The trained values of \mathbf{w}_b and \mathbf{w}_0 to \mathbf{w}_3 are illustrated in Fig. 4 as well as being tab-
 464 ulated in Appendix B. Here, \mathbf{w}_b broadly represents a mean ice surface emissivity and
 465 other vectors describe variability around this as a function of the input values. Of these
 466 inputs, z_B is the only physical one: it is a scalar transformation of the surface temper-
 467 ature $z_B = \max(273.0 - T_S, 0.0)/30.0$. The factor 30.0 in the denominator is for nor-
 468 malisation and is chosen so that maximum values do not much exceed 1, as is common
 469 practice in machine learning. As mentioned in Sec. 2.4, the representation of the surface
 470 using an emissivity and a skin temperature is a big approximation. Quite often the ra-
 471 diation may be coming from within the sea ice or snow layer, where it is warmer, and
 472 the effective radiating temperature of the snow and sea ice (in Eq. 5) should be higher
 473 than the given skin temperature from the ECMWF model. To partly compensate this,
 474 the model is designed to allow an increase in the surface emissivity as the skin temper-
 475 ature gets colder. Figure 4a shows the modelled surface emissivity at $T_S = 273.0$ K ($z_B =$
 476 0) and $T_S = 261.0$ K ($z_B = 0.4$), assuming all other inputs are zero. Going to the lower

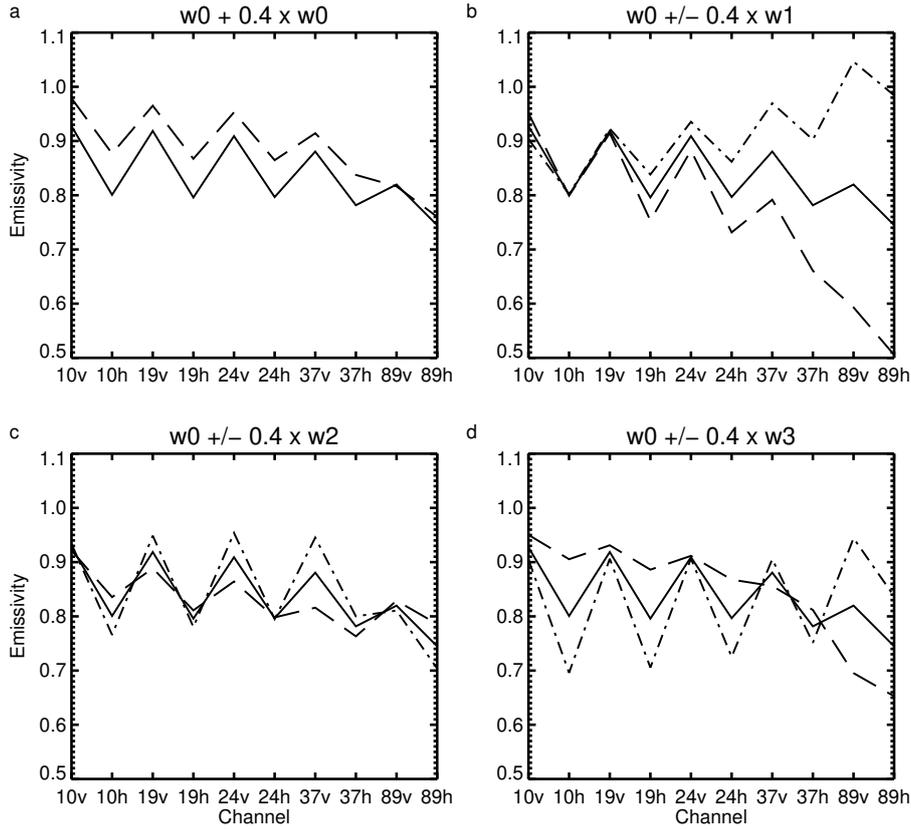


Figure 4. Illustrating the trained sea ice emissivity model. The ice emissivity with all inputs zero is w_b which is shown in all panels (solid line). If just one input parameter is perturbed to $+0.4$ (long dash) or -0.4 (dot dash) then the emissivity changes according to weights w_0 to w_3 .

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

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

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

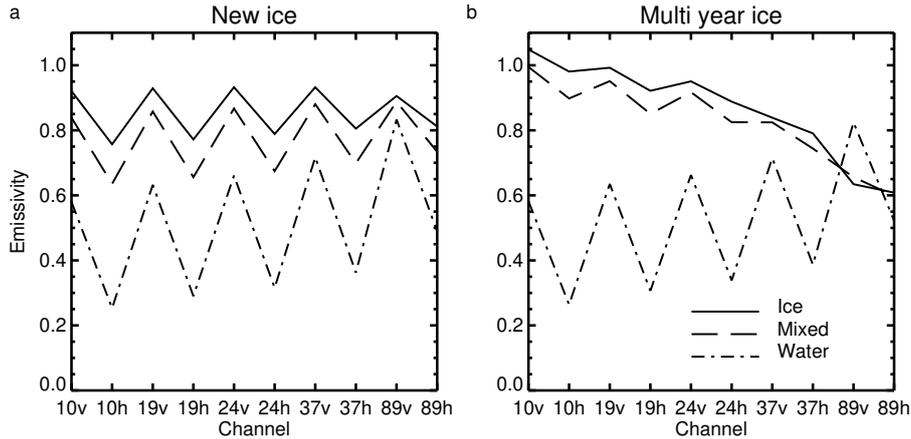


Figure 5. Emissivity as a function of channel at illustrative locations in the Arctic on 7th November (a) in the multi year ice and (b) in newly formed sea ice. The mixed emissivity (dashed) is generated from the sea ice concentration and the ice and water emissivities (solid and dot-dash) according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

496 vector \mathbf{w}_b , which gives the baseline surface emissivity for channel 10v:

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

497 In practice, the size of the background error here, $w_{\text{bgerr}} = 0.00001$, was chosen to keep
 498 $w_{b,1}$ extremely close to its background value. An emissivity of 0.93 is within the typical
 499 values for channel 10v, and further, the emissivity in this channel is thought to be
 500 relatively invariant between first year and multi year ice (Lee et al., 2017, their Table
 501 4). The posterior sensitivity tests in Appendix A2 show that this constraint on 10v emis-
 502 sivity could in practice have been relaxed or removed. In early testing before the addi-
 503 tion of sea ice concentration loss terms, the emissivity loss term was necessary to avoid
 504 completely non-physical sea ice concentrations being retrieved. However, it appears that
 505 the sea ice concentration loss terms are more effective and targeted, and would be suf-
 506 ficient on their own in future versions of this work. In any case, despite the constraint,
 507 the final trained model still has freedom to adjust the output emissivity at 10v down to
 508 at least 0.9 and up to at least 0.98 depending on the input parameters (Fig. 4).

509 Figure 5 illustrates sea ice surface emissivities generated by the empirical model
 510 (Eq. 8) at two locations in the Arctic, alongside ocean water emissivities (Eq. 10, to be
 511 described shortly) and the mixed-surface emissivity that is determined by the sea ice con-
 512 centration (Eq. 7). The location with new ice (panel a) generates relatively strong po-
 513 larisation but has little frequency dependence. The frequency dependence is in qualita-
 514 tive agreement with aircraft observations of new ice (e.g. Hewison & English, 1999). In
 515 contrast, a location in the multi-year ice is mostly unpolarised but its emissivity drops
 516 with frequency, as is generally observed (e.g. Baordo & Geer, 2015). Both sea ice emis-
 517 sivity spectra contrast strongly with the ocean water emissivity which is even more po-
 518 larised and increases with frequency. These are locations where the sea ice concentra-
 519 tion is close to 1, so the mixed-surface emissivity is relatively close to the ice emissiv-
 520 ity.

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

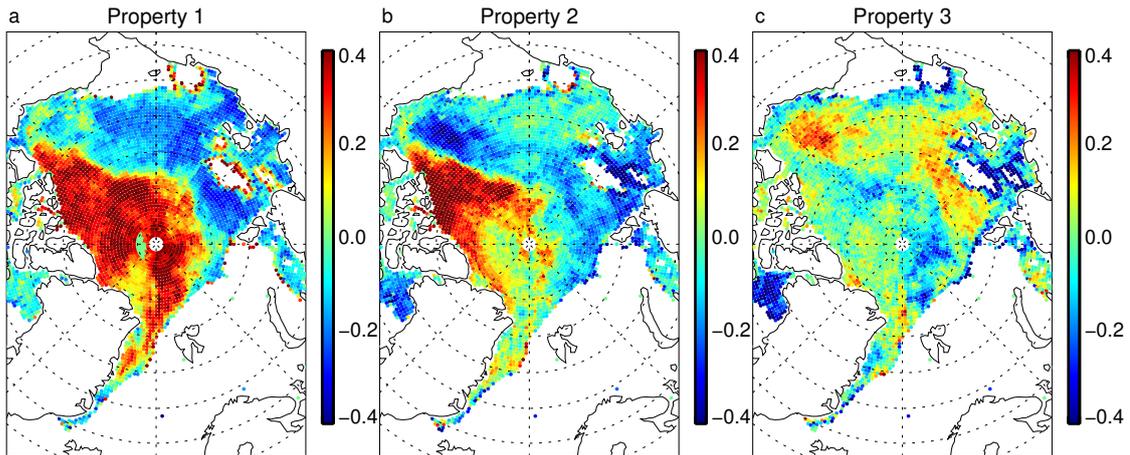


Figure 6. Estimated values of the three empirical sea ice properties in the Arctic on 7th November 2020. Values are only shown where the estimated sea ice concentration is greater than 0.2.

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

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

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

540 2.6 Empirical state representing micro- and macro-physical properties 541 of sea ice

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

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

556 2.7 Sea ice

557 In order to speed up the training, the sea ice maps \mathbf{C}_{ice} were initialised with a monthly
 558 mean sea ice concentration computed from the existing ECMWF sea ice concentration
 559 analysis at the AMSR2 locations. However, apart from the physical constraints described
 560 below, the sea ice concentration can be freely adjusted to best fit the observations, and
 561 is not constrained in any way to the ECMWF sea ice concentration. The existing ECMWF
 562 sea ice analysis is based on a combination of an ocean model, a sea ice model (LIM2, Tim-
 563 mermann et al., 2005) and the assimilation of highly processed observations, using the
 564 ocean data assimilation framework (OCEAN5, Zuo et al., 2019; de Rosnay et al., 2022)
 565 and will be referred to by this name subsequently. The observations are the OSTIA sea
 566 ice dataset (Good et al., 2020) which is ultimately based on sea ice retrievals (OSI-SAF,
 567 2016) obtained from a microwave sensor similar to AMSR2 using a traditional heuristic
 568 approach (e.g. Comiso et al., 2003). It is important that the new emissivity model
 569 should not try to fit these heuristic assumptions in any way. Further, due to the com-
 570 plex processing chain, the OCEAN5 sea ice can be at least 48 h behind reality once it
 571 has been mapped to observation locations (e.g. Baordo & Geer, 2015; Browne et al., 2019;
 572 de Rosnay et al., 2022, this work). As encountered when prototyping the current approach,
 573 if there is an incorrect sea ice concentration in Eq. 7, the sea ice emissivity model can
 574 take on characteristics of the ocean water surface emissivity, which would make the model
 575 essentially useless. For all these reasons, it was important not to allow the new sea ice
 576 analysis to be constrained by the OCEAN5 sea ice analysis in any way.

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

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

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

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

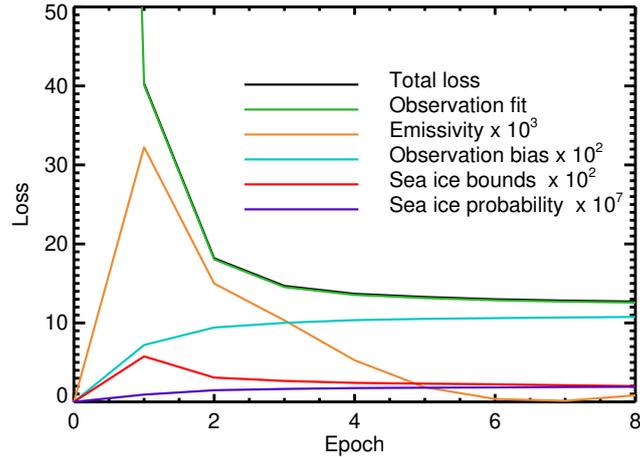


Figure 7. Loss terms during the training of the hybrid network (see Table 1) The total loss term is dominated by the observation fit, so it is mostly hidden behind the observation fit losses; the total loss at epoch zero is 340 and is off scale. The order of lines in the key follows the top-to-bottom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

601 sea ice values, the full loss term was:

$$J_{\text{seaice_tsfc}} = \frac{1}{n} \sum_d 4.0 \times \max(T_{S,d} - 273.2, 0.0) \quad (12)$$

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

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

617 3 Results

618 3.1 Training and fit to observations

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

625 the occurrence of non-physical sea ice concentrations. The loss terms have a very wide
 626 range of magnitudes and hence all but the observation term have required re-scaling to
 627 be visible on this figure. Despite this, the results show that most of these additional smaller
 628 loss terms are effective (see appendix). The hybrid model appears mainly converged after
 629 8 epochs, but Appendix A1 (using just a month of training data) investigates the use
 630 of larger numbers of epochs, and shows that the model can continue to evolve, but at
 631 the cost of some overfitting. Training on the yearly dataset is resource-limited to 8 epochs,
 632 but a beneficial, though involuntary, side effect has been that it is using ‘early stopping’
 633 to avoid over-fitting.

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

645 The RMS of the initial and final (or analysis) departures ($y_{\text{obs},ij} - y_{ij}$) are given
 646 in Table 2. Initial RMS departures are of order 10 K to 40 K. Panel g illustrates these
 647 departures in channel 37v, with large discrepancies coming both from the incorrect sur-
 648 face emissivity and the incorrect sea ice initial field. By contrast, the analysis departures
 649 (e.g. panel h) are mostly within ± 3 K and rarely larger than 10 K. The largest remain-
 650 ing departures are mainly found over open ocean and not sea ice, and come from errors
 651 in the location of cloud and precipitation in the ECMWF background fields, which can-
 652 not be adjusted in the current network. The RMS of the analysis departures ranges from
 653 2.7 K in channel 10v to 9.9 K in 89h (Table 2). These values are close to the prescribed
 654 observation errors, as intended.

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

663 At higher frequencies, the improvements between the initial model and the anal-
 664 ysis (Fig. 8 d and e) are driven not just by the improved sea ice concentration, but also
 665 by the development of the empirical terms of the surface emissivity model, and the em-
 666 pirical state inputs. Empirical properties 1 and 2 (Fig. 6) are the main driver in the anal-
 667 ysed brightness temperature at 37v, helping generate brightness temperatures down to
 668 210 K over the multi-year ice (empirical property 1 values of up to 0.4) and up to 260 K
 669 over the first year ice (empirical property 1 down to -0.4). Hence the hybrid model ap-
 670 pears to make a physically plausible set of adjustments in order to fit the observations.

671 3.2 Sea ice concentration

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

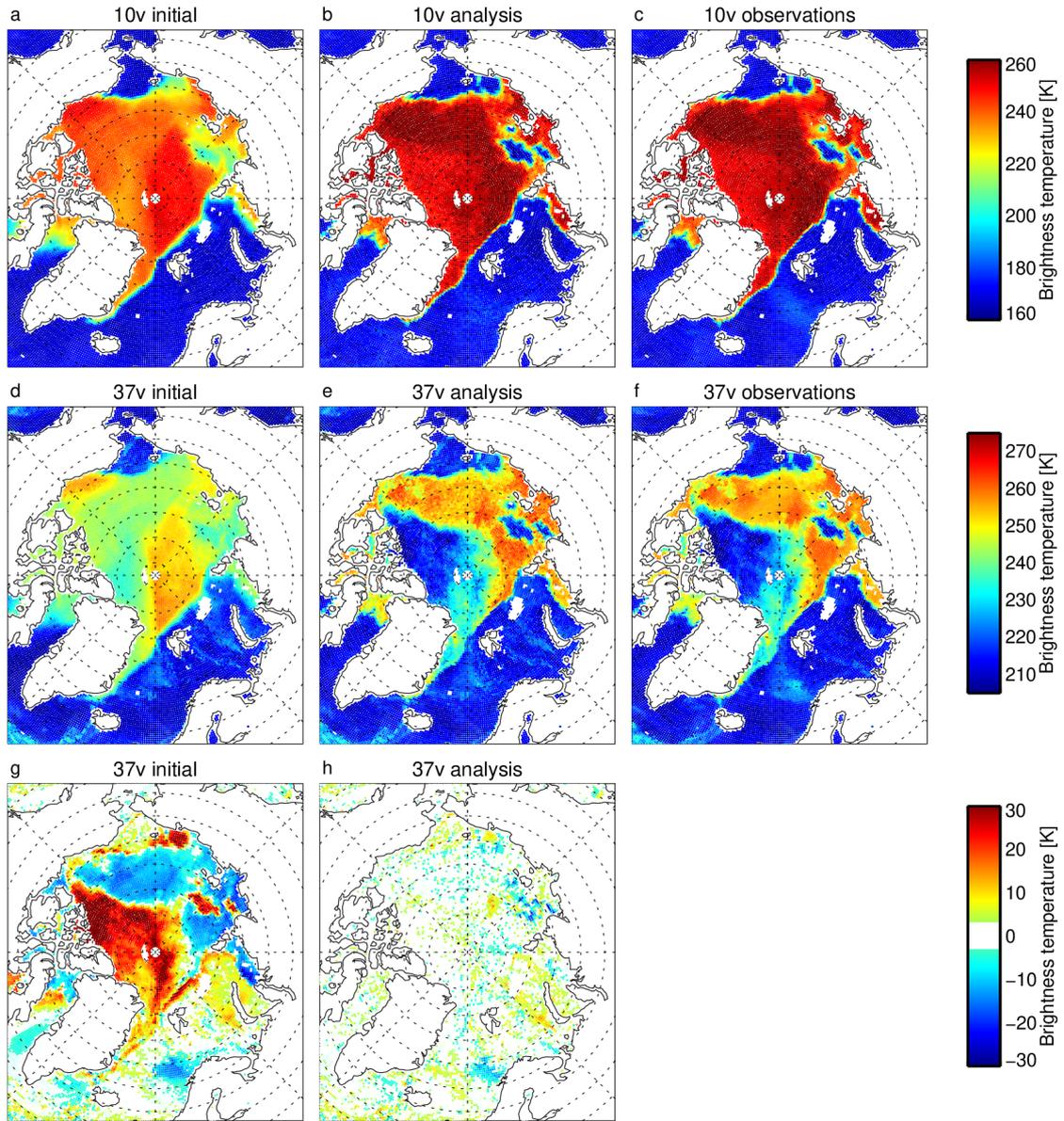


Figure 8. Comparison of simulated and observed brightness temperatures in channels 10v and 37v, over the Arctic on 7th November 2020: (a,d) initial guess; (b,e) after training, in other words the analysis; (c,f) observations; (g) the initial guess minus observation departure; (h) the analysis minus observation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

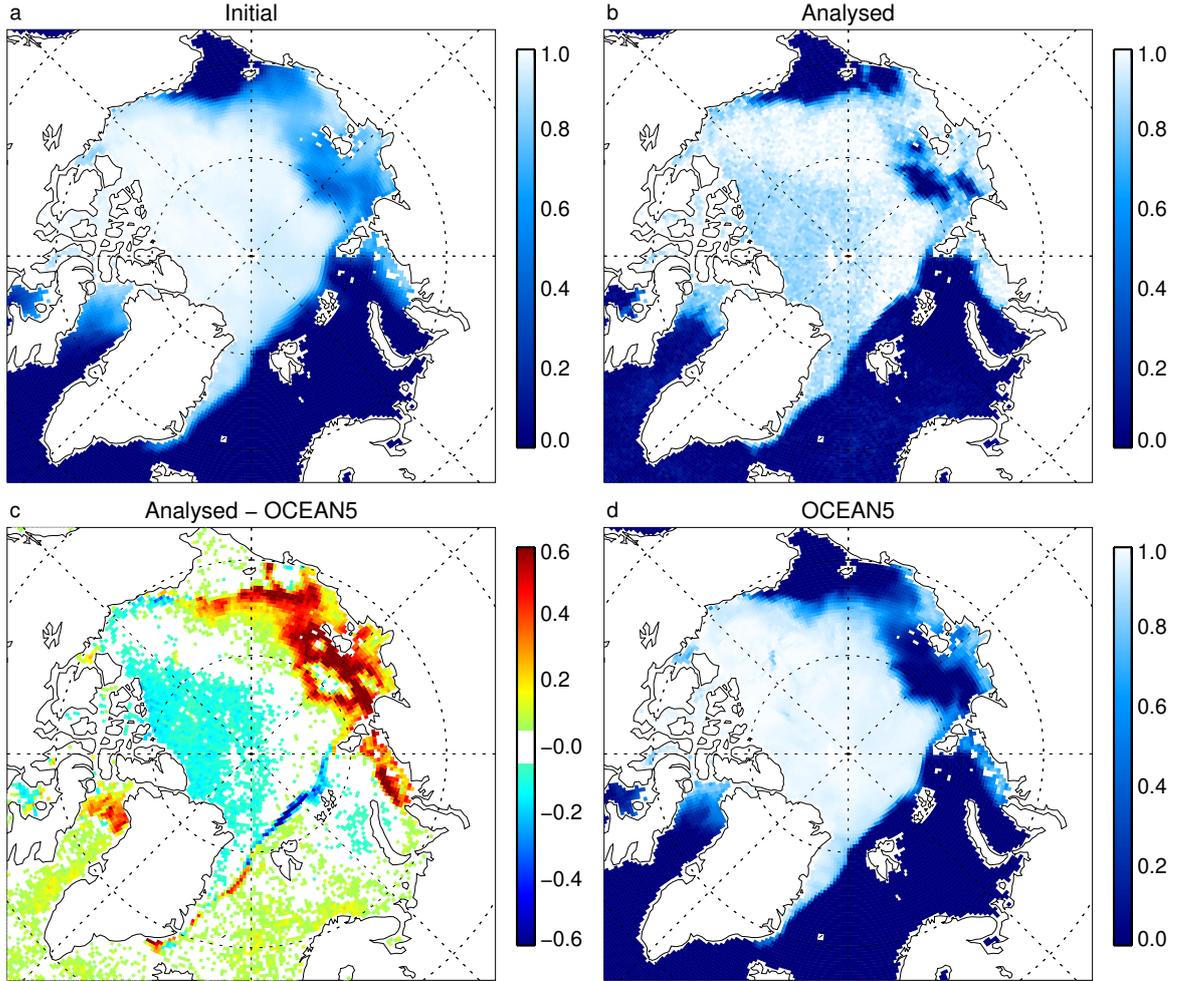


Figure 9. Comparison of sea ice concentration on 7th November 2020: (a) Initial monthly mean sea ice; (b) Analysed in this work; (c) Difference between analysed and OCEAN5; (d) OCEAN5 sea ice. Differences smaller than 0.05 are not shown. Latitude and longitude grid spacings are 10° and 45° respectively.

676 ysis (OCEAN5, panel d). There are also disagreements in the location of the ice edge
 677 by up to around 100 km in the region of Svalbard. The fundamental issue is the roughly
 678 48 h time delay in producing the OCEAN5 analysis; eventually the OCEAN5 sea ice catches
 679 up and provides a similar picture (not shown). Table 3 shows that the new analysed sea
 680 ice agrees best with OCEAN5 if it is artificially delayed by 2 or 3 days, consistent with
 681 previous expectations (Baordo & Geer, 2015; Browne et al., 2019; de Rosnay et al., 2022).
 682 There is also some low concentration (< 0.2) sea ice incorrectly identified in the cur-
 683 rent analysis, for example off the N coast of Scandinavia, visible mainly in the difference
 684 plot (panel c). This spurious sea ice is generated when clouds are present in the obser-
 685 vations but not in the ECMWF 12 h atmospheric forecast (z_A here). The network (Fig. 1)
 686 cannot adjust the atmosphere to add cloud and instead can only create sea ice to bet-
 687 ter fit the observations (an opposite effect would tend to create negative sea ice concen-
 688 trations where the ECMWF forecasts have too much cloud, but this is suppressed by the
 689 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

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

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

716 Figure 12 shows the annual cycle of ice area, using ‘cleaned up’ sea ice concentra-
 717 tions as described in the previous paragraph. Compared to OCEAN5, this work gives
 718 slightly higher ice area in the Antarctic winter, and slightly lower ice area in the Arc-
 719 tic winter. The better timeliness of the new sea ice analysis is also clear. In early Novem-
 720 ber in the Arctic, the current work identifies a rapid freezing event that is smoothed out
 721 and delayed in the OCEAN5 sea ice analysis, as also illustrated in Fig. 9. A similar pic-
 722 ture is seen in the Antarctic, from late February to mid-May. Apart from these issues,
 723 there is good agreement between the annual cycles in the two ice products.

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

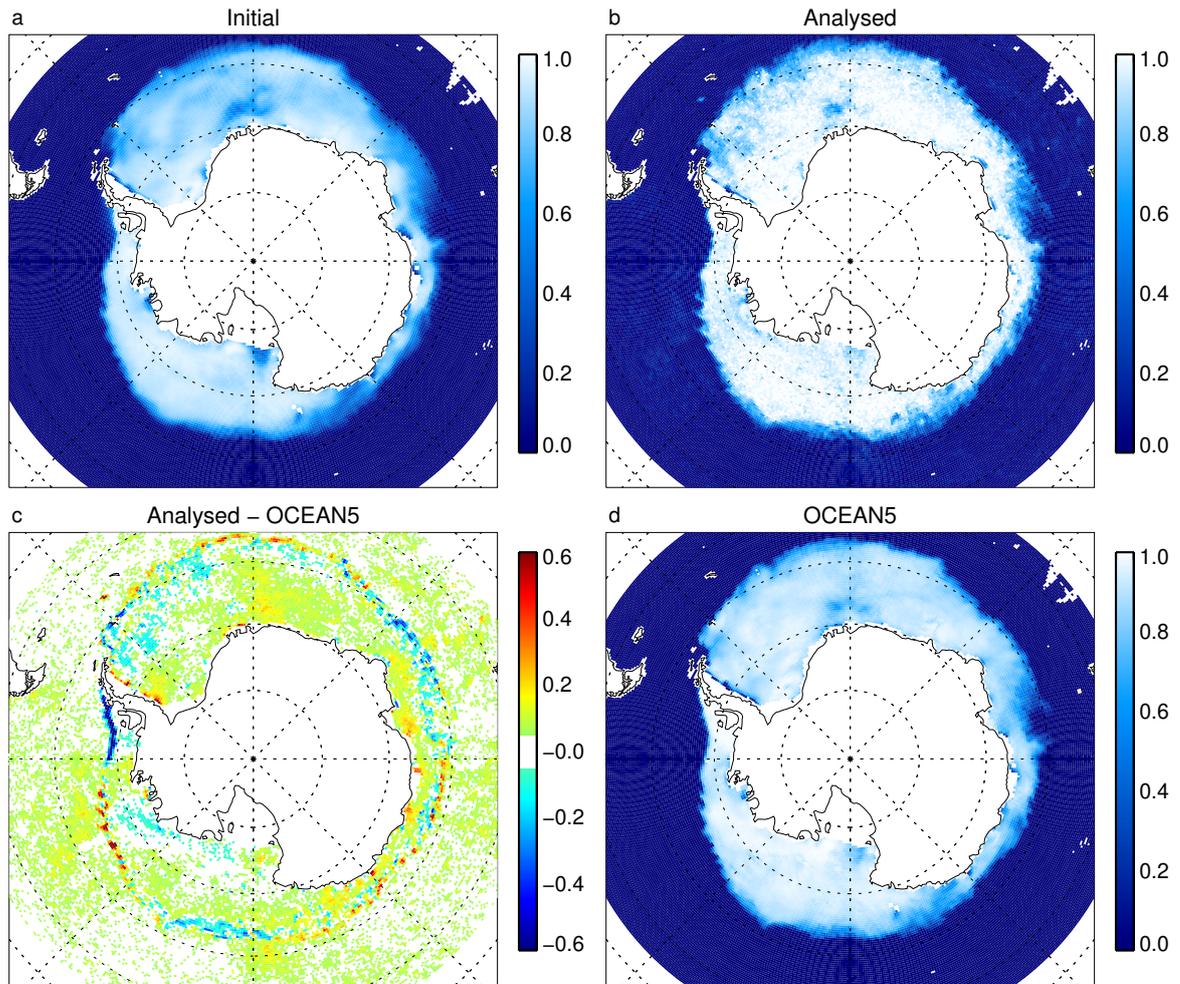


Figure 10. As Fig. 9 on 7th November 2020, but for the Antarctic. The Greenwich meridian is to the top.

727 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
 728 in the multi-year ice zone should be close to 1, like in OCEAN5, this appears to be
 729 a defect in the current approach and likely relates to the constraint of 10v ice surface
 730 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
 731 should have been allowed to decrease more at 10v. In the Antarctic (panel b)
 732 there is typically a small overestimation (around 0.02) compared to OCEAN5, and much
 733 larger positive differences in patches close to the Antarctic coast. These can be traced
 734 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
 735 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
 736 of the data this suggests there is a real defect in the OCEAN5 sea ice. There are small
 737 underestimations of the sea ice compared to OCEAN5 in the Weddel Sea which might
 738 also be due to the multi-year ice issue, since this is one of the few areas of the Antarctic
 739 where ice can persist from one year to the next. But apart from the areas of disagree-
 740
 741
 742
 743

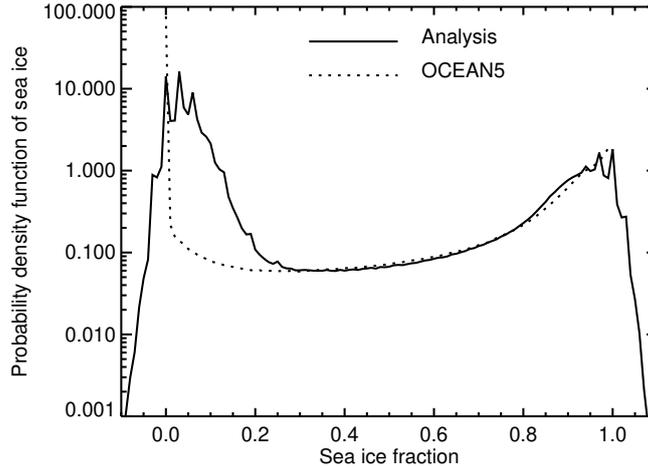


Figure 11. Probability density function for sea ice concentration for the new analysis and the existing OCEAN5, over the year and both hemispheres, using a log y-axis to better compare the full probability range.

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

746 4 Conclusion

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

761 There is no perfect truth against which to evaluate the resulting global daily maps
 762 of sea ice, since existing global satellite retrievals of sea ice are based on heuristic meth-
 763 ods which this work aims to replace, and in-situ measurements are very limited in cov-
 764 erage. The high quality of fit to observations suggests that the sea ice results are good,
 765 but comparisons are also made to the existing ECMWF sea ice analyses, generated by
 766 the OCEAN5 assimilation system. The new maps are 48 h to 72 h more timely than the
 767 sea ice concentration analysed in OCEAN5. The new maps have generally sharper res-
 768 olution including plausible mesoscale features like some apparently eddy-generated 300
 769 km wave features in the Antarctic sea ice edge. The new data also suggest there is a sub-
 770 stantial underestimate of sea ice cover in the OCEAN5 analyses in the Antarctic sum-
 771 mer. There are some limitations in the new data too, mainly an apparent underestimate

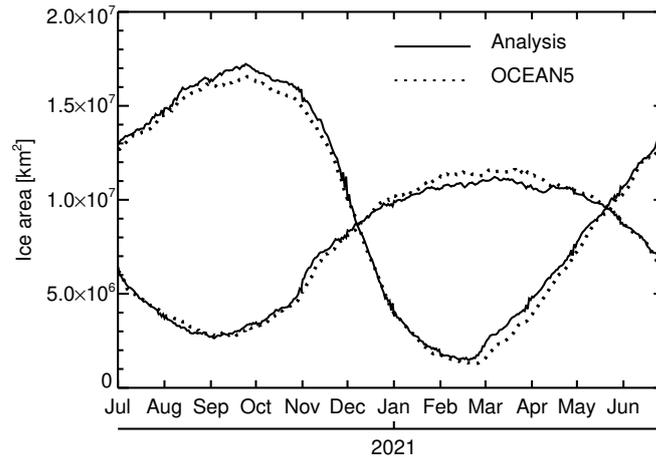


Figure 12. Estimated hemispheric sea ice area from the current work (solid) and OCEAN5 (dotted) from 1st July 2020 to 30th June 2021. The Antarctic ice extent is largest in September and the Arctic in March.

772 of the sea ice concentration in the Arctic winter multi-year ice, where it seems to be about
 773 0.1 too low, in compensation for remaining inaccuracies in the sea ice emissivity model.

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

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

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

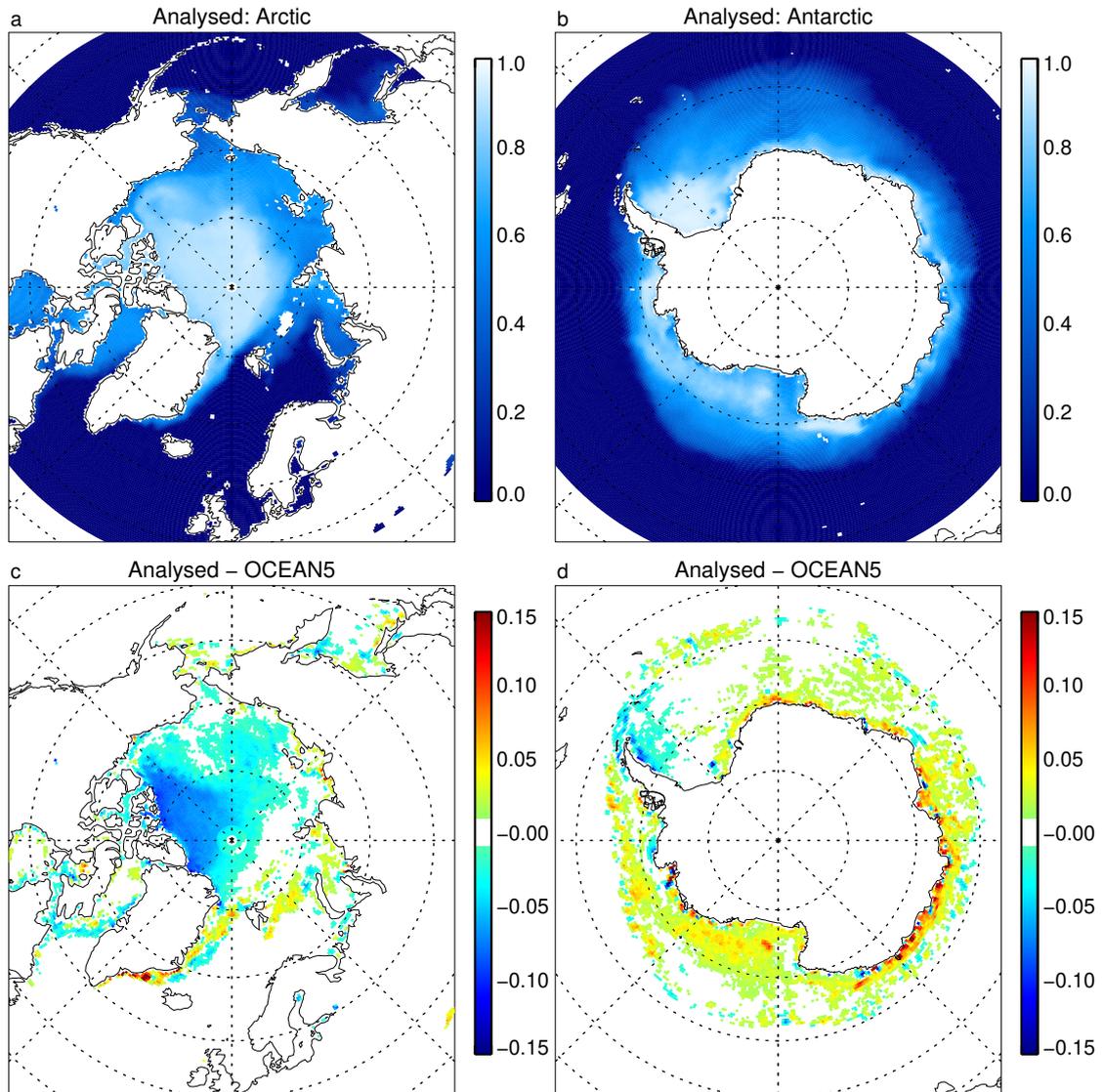


Figure 13. Annual mean sea ice concentration analysed in this work (a,b) and difference in annual mean between the current work and OCEAN5 (c,d). Differences smaller than 0.01 are not plotted.

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

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

824 The training of the hybrid physical-empirical model has demonstrated a number
 825 of innovations that have been facilitated by the availability of modern machine learn-
 826 ing and differential programming tools such as Keras and TensorFlow (Abadi et al., 2015).
 827 Considering the network in Fig. 1, if the sea ice emissivity was a known parameter, the
 828 lowermost empirical part of the network (devoted to the empirical state parameters and
 829 the empirical surface emissivity model) would be unnecessary, and this work would have
 830 been a standard data assimilation problem of estimating the sea ice concentration maps
 831 from the satellite observations. However, there would still be the novelty that it was im-
 832 plemented in a machine learning framework. If the sea ice concentration, and the micro-
 833 and macro-physical parameters of the sea ice and snow cover, were known globally through
 834 modelling or observations, those parameters could have been treated as known input vari-
 835 ables and the empirical sea ice emissivity model would be an ML component model trained
 836 inside an otherwise physical data assimilation network (e.g. Reichstein et al., 2019; Geer,
 837 2021) but not achieved practically as yet, to this author’s knowledge. But since the in-
 838 put variables are also unknown (the chicken and egg problem) the most novel aspect of
 839 this network is that it simultaneously trains an empirical model and works out what its
 840 inputs should be (the maps of sea ice parameters).

841 It is proposed to describe the simultaneous training of state and model as an ‘em-
 842 pirical state’ method, with the essential components being:

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

849 A similar approach could be extended to snow surfaces over land, soil moisture and veg-
 850 etation analyses from microwave sensors, and possibly to many other ‘chicken and egg’
 851 problems in the wider developments towards earth system assimilation. The approach
 852 can also be extended to a ‘hybrid state’, as in this work, where some of the inputs are
 853 physical and some empirical. By taking this mix of empirical and physical inputs, em-
 854 pirical methods can become progressively more physical, as models become able to sup-
 855 ply more and higher quality input parameters. For example, snow grain size, depth and
 856 temperature may eventually be available within the ECMWF model, and these could
 857 be included as inputs to the empirical surface emissivity model, and the number of em-
 858 pirical parameters representing unknown aspects of the snow and sea ice state could hope-
 859 fully be reduced. This means that empirical methods can be a quick way of getting started
 860 with a new assimilation domain, such as sea ice, but they can subsequently evolve to-
 861 wards more physical approaches as physical models become more capable within that
 862 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

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

881 Appendix A Sensitivities - overview

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

890 A1 Sensitivities - epochs, batch size, network complexity

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

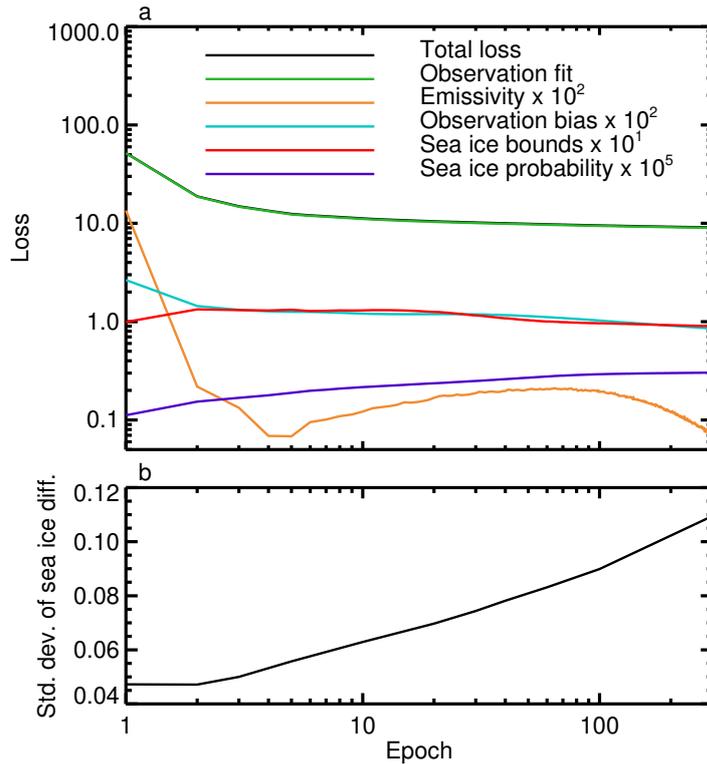


Figure A1. (a) Total loss and individual loss terms (scaled as noted in key) using an extended number of training epochs; (b) Standard deviation of the difference between analysed sea ice fraction with a 48 h delay and OCEAN5 sea ice fraction. Note the logarithmic x axis and, in panel a, also y axis,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

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

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

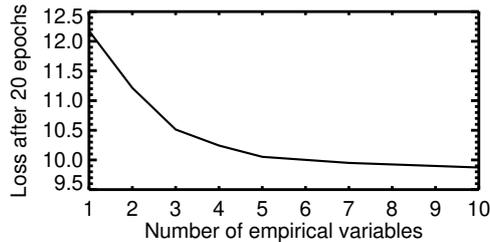


Figure A2. Total loss as a function of the number of empirical variables.

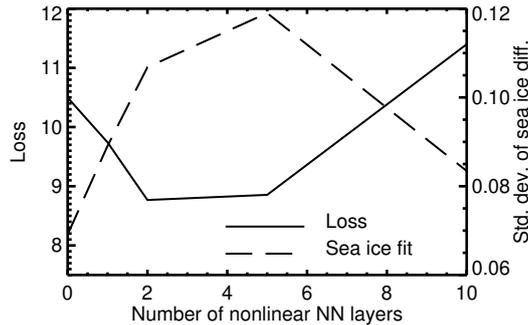


Figure A3. Total loss and sea ice fit to OCEAN5 with a 48 h offset, as a function of the number of neural network layers, based on 50 epochs of training and 20 neurons per layer. The data point for "0" layers is used to represent a single linear layer.

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

918 The chosen model for sea ice emissivity is linear, but the framework is perfectly ca-
 919 pable of training a nonlinear model, including deep networks. Sensitivity tests were per-
 920 formed using a sigmoid activation function and between 1 and 10 fully connected lay-
 921 ers using 20 neurons each. In these tests the number of epochs was set to 50 to ensure
 922 the deeper networks were converged. Fig. A3 shows the results in terms of the loss func-
 923 tion at 50 epochs and the fit of the sea ice field to ECMWF sea ice (with the 2 day off-
 924 set for best fit). The point marked 0 layers corresponds to the normal linear single layer
 925 model, but trained for 50 epochs. Going to a nonlinear activation function and adding
 926 up to 2 layers is capable of fitting the observations better, as indicated by the reductions
 927 in the loss function. However, this comes at the price of generating a poorer-quality sea
 928 ice field, as indicated by the increasing standard deviation of the difference with ECMWF
 929 / OCEAN5 sea ice becoming larger than 0.1. For 3 layers and greater, the picture re-
 930 verses, but this is likely because 50 epochs is insufficient to fully train the deeper net-
 931 works. In broad terms, adding multiple layers and nonlinearity to the surface emissiv-
 932 ity model seems to give greater possibilities for over-fitting the data, similar to increas-
 933 ing the number of epochs.

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

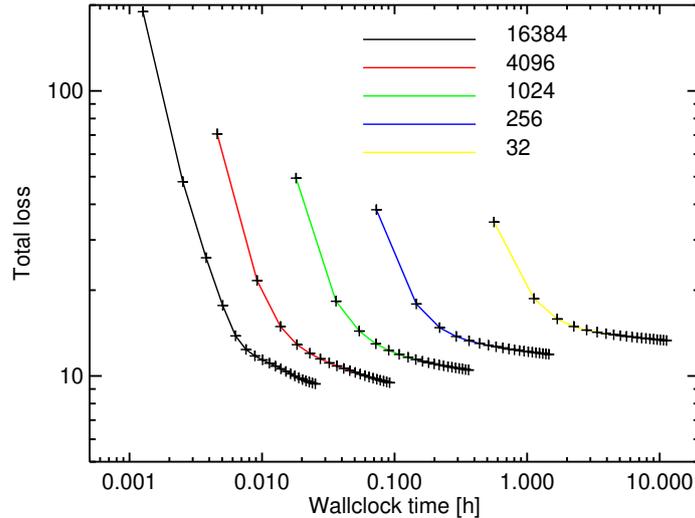


Figure A4. Total loss as a function of wallclock time for five different batch sizes, as indicated in the key. Crosses indicate the wallclock time and the corresponding loss after each of 20 epochs.

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

959 A2 Sensitivities - prior knowledge loss terms

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

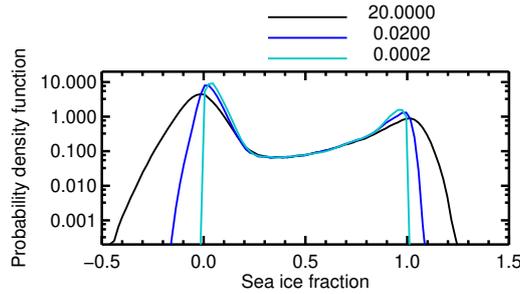


Figure A5. Probability density function of sea ice fraction for three settings of the sea ice background error, as given in the key.

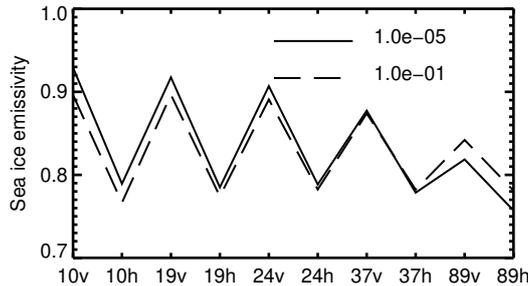


Figure A6. Sea ice surface emissivity baseline term w_b , as a function of AMSR2 channel identifier, for values of the surface emissivity background error (w_{bgerr} in Eq. 9) as indicated in the key.

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

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

981 Relaxing the bias background error b_{bgerr} in Eq. 4 to values larger than the chosen
 982 $b_{bgerr} = 0.001 K$ improves the fit to observations, and reduces the total loss from
 983 around 10.5 down to around 9.8 (no figure shown). However, this leads the network to
 984 generate unreasonably large bias corrections of up to -7 K over ocean and -16 K over sea
 985 ice (no figure shown). This likely shows that the problem is ill-posed when both TB bias
 986 and surface emissivity are allowed to vary. Hence the bias correction term is a very im-
 987 portant 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.

1020 Original AMSR2 data for this value added data product was provided by Japan
 1021 Aerospace Exploration Agency (JAXA, 2023) via the G-Portal. The user is entitled to
 1022 use JAXA G-Portal AMSR2 data free of charge without any restrictions (including com-
 1023 mercial use) except for the condition about acknowledgement of data credit as stipulated
 1024 in Article 7.(2) of the terms and conditions at [https://gportal.jaxa.jp/gpr/index/](https://gportal.jaxa.jp/gpr/index/eula?lang=en)
 1025 [eula?lang=en](https://gportal.jaxa.jp/gpr/index/eula?lang=en)

1026 The Python code was run on Python 3.8.8-01 (Python Software Foundation, 2021)
 1027 including Tensorflow and Keras 2.8.0 (Abadi et al., 2015; TensorFlow, 2021) on the ECMWF
 1028 ATOS supercomputer CPU nodes.

1029 Acknowledgments

1030 Niels Bormann, Fabrizio Baordo, Matthew Chantry and Tracy Scanlon are thanked for
 1031 reviews and suggestions.

1032 References

- 1033 Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., . . . Zheng, X.
 1034 (2015). *TensorFlow: Large-scale machine learning on heterogeneous systems*
 1035 [code]. Retrieved from <https://www.tensorflow.org/> (Software available
 1036 from tensorflow.org) doi: 10.5281/zenodo.4724125
- 1037 Baordo, F., & Geer, A. (2015). Microwave surface emissivity over sea-ice. *EUMET-*
 1038 *SAT NWP-SAF visiting scientist report NWPSAF_EC_VS_026*. Retrieved from
 1039 <https://nwpsaf.eu/publications/vs\reports/nwpsaf-ec-vs-026.pdf>
- 1040 Baordo, F., & Geer, A. J. (2016). Assimilation of SSMIS humidity-sounding chan-
 1041 nels in all-sky conditions over land using a dynamic emissivity retrieval. *Quart.*
 1042 *J. Roy. Meteorol. Soc.*, *142*, 2854–2866. Retrieved from [https://doi.org/10](https://doi.org/10.1002/qj.2873)
 1043 [.1002/qj.2873](https://doi.org/10.1002/qj.2873)
- 1044 Bauer, P., Moreau, E., Chevallier, F., & O’Keeffe, U. (2006). Multiple-scattering
 1045 microwave radiative transfer for data assimilation applications. *Quart. J. Roy.*
 1046 *Meteorol. Soc.*, *132*, 1259-1281. Retrieved from [https://doi.org/10.1256/qj](https://doi.org/10.1256/qj.05.153)
 1047 [.05.153](https://doi.org/10.1256/qj.05.153)
- 1048 Bauer, P., Thorpe, A., & Brunet, G. (2015). The quiet revolution of numerical
 1049 weather prediction. *Nature*, *525*(7567), 47–55.
- 1050 Berg, W., Bilanow, S., Chen, R., Datta, S., Draper, D., Ebrahimi, H., . . . Yang,
 1051 J. X. (2016). Intercalibration of the GPM microwave radiometer constellation.
 1052 *Journal of Atmospheric and Oceanic Technology*, *33*(12), 2639–2654. Retrieved
 1053 from <https://doi.org/10.1175/JTECH-D-16-0100.1>
- 1054 Bocquet, M., Brajard, J., Carrassi, A., & Bertino, L. (2019). Data assimilation
 1055 as a learning tool to infer ordinary differential equation representations of
 1056 dynamical models. *Nonlinear Processes in Geophysics*, *26*(3), 143–162.
- 1057 Bocquet, M., Brajard, J., Carrassi, A., & Bertino, L. (2020). Bayesian inference
 1058 of chaotic dynamics by merging data assimilation, machine learning and
 1059 expectation-maximization. *Foundations of Data Science*, *2*(1), 55–80. Re-
 1060 trieved from <https://doi.org/10.3934/fods.2020004>
- 1061 Brajard, J., Carrassi, A., Bocquet, M., & Bertino, L. (2020). Combining data as-
 1062 similation and machine learning to emulate a dynamical model from sparse
 1063 and noisy observations: A case study with the lorenz 96 model. *Journal of*
 1064 *Computational Science*, *44*, 101171.
- 1065 Browne, P. A., de Rosnay, P., Zuo, H., Bennett, A., & Dawson, A. (2019). Weakly
 1066 coupled ocean–atmosphere data assimilation in the ECMWF NWP system.
 1067 *Remote Sensing*, *11*(3), 234. Retrieved from [https://doi.org/10.3390/](https://doi.org/10.3390/rs11030234)
 1068 [rs11030234](https://doi.org/10.3390/rs11030234)
- 1069 Buehner, M., Caya, A., Pogson, L., Carrieres, T., & Pestieau, P. (2013). A new envi-
 1070 ronment canada regional ice analysis system. *Atmosphere-Ocean*, *51*(1), 18–34.

- 1071 Retrieved from <https://doi.org/10.1080/07055900.2012.747171>
- 1072 Chollet, F., et al. (2015). *Keras*. <https://keras.io>.
- 1073 Comiso, J. C., Cavalieri, D. J., & Markus, T. (2003). Sea ice concentration, ice tem-
1074 perature, and snow depth using AMSR-E data. *IEEE Trans. Geosci. Remote*
1075 *Sens.*, *41*(2), 243–252. Retrieved from <https://doi.org/10.1109/TGRS.2002>
1076 [.808317](https://doi.org/10.1109/TGRS.2002.808317)
- 1077 de Rosnay, P., Browne, P., de Boissésou, E., Fairbairn, D., Hirahara, Y., Ochi, K.,
1078 ... Rabier, F. (2022). Coupled assimilation at ECMWF: current status,
1079 challenges and future developments. *Quart. J. Roy. Meteorol. Soc.*, *148*,
1080 2672–2702. Retrieved from <https://doi.org/10.1002/qj.4330>
- 1081 Desroziers, G., Berre, L., Chapnik, B., & Poli, P. (2005). Diagnosis of observation,
1082 background and analysis-error statistics in observation space. *Quart. J. Roy.*
1083 *Meteorol. Soc.*, *131*, 3385–3396.
- 1084 Eyre, J. R., English, S. J., & Forsythe, M. (2020). Assimilation of satellite data in
1085 numerical weather prediction. Part I: The early years. *Quart. J. Roy. Meteorol.*
1086 *Soc.*, *146*(726), 49–68. Retrieved from <https://doi.org/10.1002/qj.3654>
- 1087 Geer, A. J. (2021). Learning earth system models from observations: Machine learn-
1088 ing or data assimilation? *Phil. Trans. R. Soc. A*, *379*, 20200089. Retrieved
1089 from <https://doi.org/10.1098/rsta.2020.0089>
- 1090 Geer, A. J. (2023a). *Code for simultaneous inference of sea ice state and surface*
1091 *emissivity model using machine learning and data assimilation* [code]. Re-
1092 trieved from <https://doi.org/10.5281/zenodo.10013542>
- 1093 Geer, A. J. (2023b). *Data for simultaneous inference of sea ice state and surface*
1094 *emissivity model using machine learning and data assimilation* [dataset]. Re-
1095 trieved from <https://doi.org/10.5281/zenodo.10033377>
- 1096 Geer, A. J. (2023c). Joint estimation of sea ice and atmospheric state from mi-
1097 crowave imagers in operational weather forecasting. , in preparation.
- 1098 Geer, A. J., Baordo, F., Bormann, N., English, S., Kazumori, M., Lawrence, H., ...
1099 Lupu, C. (2017). The growing impact of satellite observations sensitive to hu-
1100 midity, cloud and precipitation. *Quart. J. Roy. Meteorol. Soc.*, *143*, 3189–3206.
1101 Retrieved from <https://doi.org/10.1002/qj.3172>
- 1102 Geer, A. J., & Bauer, P. (2011). Observation errors in all-sky data assimilation.
1103 *Quart. J. Roy. Meteorol. Soc.*, *137*, 2024–2037. Retrieved from <https://doi>
1104 [.org/10.1002/qj.830](https://doi.org/10.1002/qj.830)
- 1105 Geer, A. J., Bauer, P., & O’Dell, C. W. (2009). A revised cloud overlap scheme for
1106 fast microwave radiative transfer. *J. App. Meteor. Clim.*, *48*, 2257–2270. Re-
1107 trieved from <https://doi.org/10.1175/2009JAMC2170.1>
- 1108 Geer, A. J., Lonitz, K., Duncan, D., & Bormann, N. (2022). *Improved surface treat-*
1109 *ment for all-sky microwave observations* (Tech. Memo. No. 894). Reading, UK:
1110 ECMWF. Retrieved from <https://doi.org/10.21957/zi7q6hau>
- 1111 Geer, A. J., Lonitz, K., Weston, P., Kazumori, M., Okamoto, K., Zhu, Y., ...
1112 Schraff, C. (2018). All-sky satellite data assimilation at operational weather
1113 forecasting centres. *Quart. J. Roy. Meteorol. Soc.*, *144*, 1191–1217. doi:
1114 [10.1002/qj.3202](https://doi.org/10.1002/qj.3202)
- 1115 Gettelman, A., Geer, A. J., Forbes, R. M., Carmichael, G. R., Feingold, G., Pos-
1116 selt, D. J., ... Zuidema, P. (2022). The future of earth system prediction:
1117 Advances in model-data fusion. *Science Advances*, *8*(14), eabn3488.
- 1118 Ghahramani, Z. (2015). Probabilistic machine learning and artificial intelligence.
1119 *Nature*, *521*(7553), 452–459.
- 1120 Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feed-
1121 forward neural networks. In *Proceedings of the thirteenth international confer-*
1122 *ence on artificial intelligence and statistics* (pp. 249–256).
- 1123 Golmant, N., Vemuri, N., Yao, Z., Feinberg, V., Gholami, A., Rothauge, K., ...
1124 Gonzalez, J. (2018). On the computational inefficiency of large batch sizes for
1125 stochastic gradient descent. *arXiv preprint arXiv:1811.12941*.

- 1126 Good, S., Fiedler, E., Mao, C., Martin, M. J., Maycock, A., Reid, R., ... others
 1127 (2020). The current configuration of the OSTIA system for operational pro-
 1128 duction of foundation sea surface temperature and ice concentration analyses.
 1129 *Remote Sensing*, 12(4), 720. Retrieved from [https://doi.org/10.3390/
 1130 rs12040720](https://doi.org/10.3390/rs12040720)
- 1131 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J.,
 1132 ... Thépaut, J. (2020). The ERA5 global reanalysis. *Quarterly Journal of
 1133 the Royal Meteorological Society*. Retrieved from [https://doi.org/10.1002/
 1134 qj.3803](https://doi.org/10.1002/qj.3803)
- 1135 Hewison, T. J., & English, S. J. (1999). Airborne retrievals of snow and ice surface
 1136 emissivity at millimeter wavelengths. *IEEE Trans. Geosci. Rem. Sens.*, 37(4),
 1137 1871–1879.
- 1138 Hirahara, Y., Rosnay, P. d., & Arduini, G. (2020). Evaluation of a microwave
 1139 emissivity module for snow covered area with CMEM in the ECMWF inte-
 1140 grated forecasting system. *Remote Sensing*, 12(18), 2946. Retrieved from
 1141 <https://doi.org/10.3390/rs12182946>
- 1142 Hortal, M., & Simmons, A. (1991). Use of reduced Gaussian grids in spectral mod-
 1143 els. *Monthly Weather Review*, 119(4), 1057–1074. Retrieved from [https://doi
 1144 .org/10.1175/1520-0493\(1991\)119<1057:UORGGI>2.0.CO;2](https://doi.org/10.1175/1520-0493(1991)119<1057:UORGGI>2.0.CO;2)
- 1145 Hsieh, W. W., & Tang, B. (1998). Applying neural network models to prediction
 1146 and data analysis in meteorology and oceanography. *Bulletin of the American
 1147 Meteorological Society*, 79(9), 1855–1870.
- 1148 JAXA. (2023). *Global Change Observation Mission – Water Advanced Microwave
 1149 Scanning Radiometer 2 level 1B* [dataset]. Retrieved from [https://doi.org/
 1150 10.57746/E0.01gs73ans548qghaknzdjyxd2h](https://doi.org/10.57746/E0.01gs73ans548qghaknzdjyxd2h)
- 1151 Kang, E.-J., Sohn, B.-J., Tonboe, R. T., Noh, Y.-C., Kwon, I.-H., Kim, S.-W., ...
 1152 Liu, C. (2023). Explicitly determined sea ice emissivity and emission temper-
 1153 ature over the arctic for surface-sensitive microwave channels. *Quart. J. Roy.
 1154 Meteorol. Soc.*. Retrieved from <https://doi.org/10.1002/qj.4492>
- 1155 Kazumori, M., & English, S. J. (2015). Use of the ocean surface wind direction sig-
 1156 nal in microwave radiance assimilation. *Quarterly Journal of the Royal Mete-
 1157 orological Society*, 141(689), 1354–1375. Retrieved from [https://doi.org/10
 1158 .1002/qj.2445](https://doi.org/10.1002/qj.2445)
- 1159 Kazumori, M., Geer, A. J., & English, S. J. (2016). Effects of all-sky assimilation
 1160 of GCOM-W/AMSR2 radiances in the ECMWF numerical weather predic-
 1161 tion system. *Quart. J. Roy. Meteorol. Soc.*, 142, 721–737. Retrieved from
 1162 <https://doi.org/10.1002/qj.2669>
- 1163 Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv
 1164 preprint arXiv:1412.6980*.
- 1165 Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M.,
 1166 Pritzel, A., ... Battaglia, P. (2022). GraphCast: Learning skillful medium-
 1167 range global weather forecasting. *arXiv preprint arXiv:2212.12794*.
- 1168 Lee, S.-M., Sohn, B.-J., & Kim, S.-J. (2017). Differentiating between first-year
 1169 and multiyear sea ice in the arctic using microwave-retrieved ice emissiv-
 1170 ities. *J. Geophys. Res.: Atmos.*, 122(10), 5097–5112. Retrieved from
 1171 <https://doi.org/10.1002/2016JD026275>
- 1172 Leinonen, J., Grazioli, J., & Berne, A. (2021). Reconstruction of the mass and
 1173 geometry of snowfall particles from multi-angle snowflake camera (MASC)
 1174 images. *Atmos. Meas. Tech.*, 14(10), 6851–6866.
- 1175 Okuyama, A., & Imaoka, K. (2015). Intercalibration of advanced microwave
 1176 scanning radiometer-2 (AMSR2) brightness temperature. *IEEE Transac-
 1177 tions on Geoscience and Remote Sensing*, 53(8), 4568–4577. Retrieved from
 1178 <https://doi.org/10.1109/TGRS.2015.2402204>
- 1179 OSI-SAF, E. (2016). *Global sea ice concentration (SSMIS)* (Tech. Rep. No. OSI-401-
 1180 b). Retrieved from https://doi.org/10.15770/EUM_SAF_OSI_NRT_2004

- 1181 Peubey, C., & McNally, A. P. (2009). Characterization of the impact of geostation-
 1182 ary clear-sky radiances on wind analyses in a 4D-Var context. *Quart. J. Roy.*
 1183 *Meteorol. Soc.*, *135*, 1863 – 1876.
- 1184 Picard, G., Sandells, M., & Löwe, H. (2018). SMRT: An active–passive microwave
 1185 radiative transfer model for snow with multiple microstructure and scattering
 1186 formulations (v1. 0). *Geoscientific Model Development*, *11*(7), 2763–2788.
- 1187 Python Software Foundation. (2021). *Python* [code]. Retrieved from [https://](https://github.com/python/cpython/releases/tag/v3.8.8rc1)
 1188 github.com/python/cpython/releases/tag/v3.8.8rc1
- 1189 Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2017). Physics informed deep learn-
 1190 ing (part i): Data-driven solutions of nonlinear partial differential equations.
 1191 *arXiv preprint arXiv:1711.10561*.
- 1192 Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N.,
 1193 et al. (2019). Deep learning and process understanding for data-driven Earth
 1194 system science. *Nature*, *566*(7743), 195–204.
- 1195 Rodgers, C. D. (2000). *Inverse methods for atmospheric sounding: Theory and prac-*
 1196 *tice* (1st ed.). Singapore: World Scientific.
- 1197 Sandells, M., Rutter, N., Wivell, K., Essery, R., Fox, S., Harlow, C., ... Toose, P.
 1198 (2023). Simulation of arctic snow microwave emission in surface-sensitive
 1199 atmosphere channels. *EGUsphere*, *2023*, 1–28. Retrieved from [https://](https://egusphere.copernicus.org/preprints/2023/egusphere-2023-696/)
 1200 egusphere.copernicus.org/preprints/2023/egusphere-2023-696/ doi:
 1201 [10.5194/egusphere-2023-696](https://doi.org/10.5194/egusphere-2023-696)
- 1202 Sandven, S., Spreen, G., Heygster, G., Girard-Ardhuin, F., Farrell, S. L., Dierking,
 1203 W., & Allard, R. A. (2023). Sea ice remote sensing—recent developments in
 1204 methods and climate data sets. *Surveys in Geophysics*, 1–37. Retrieved from
 1205 <https://doi.org/10.1007/s10712-023-09781-0>
- 1206 Scarlat, R. C., Spreen, G., Heygster, G., Huntemann, M., Pațilea, C., Pedersen,
 1207 L. T., & Saldo, R. (2020). Sea ice and atmospheric parameter retrieval from
 1208 satellite microwave radiometers: Synergy of amsr2 and smos compared with
 1209 the cimr candidate mission. *Journal of Geophysical Research: Oceans*, *125*(3),
 1210 e2019JC015749. Retrieved from <https://doi.org/10.1029/2019JC015749>
- 1211 Schneider, T., Lan, S., Stuart, A., & Teixeira, J. (2017). Earth system modeling
 1212 2.0: A blueprint for models that learn from observations and targeted high-
 1213 resolution simulations. *Geophys. Res. Lett.*, *44*(24), 12–396. Retrieved from
 1214 <https://doi.org/10.1002/2017GL076101>
- 1215 Smith, S. L., Kindermans, P.-J., Ying, C., & Le, Q. V. (2017). Don’t decay the
 1216 learning rate, increase the batch size. *arXiv preprint arXiv:1711.00489*.
- 1217 Spreen, G., Kaleschke, L., & Heygster, G. (2008). Sea ice remote sensing using
 1218 AMSR-E 89-GHz channels. *J. Geophys. Res.: Oceans*, *113*(C2), C02S03. Re-
 1219 trieved from <https://doi.org/10.1029/2005JC003384>
- 1220 TensorFlow. (2021). *Tensorflow version 2.8.0* [code]. Retrieved from [https://doi](https://doi.org/10.5281/zenodo.5799851)
 1221 [.org/10.5281/zenodo.5799851](https://doi.org/10.5281/zenodo.5799851)
- 1222 Timmermann, R., Goosse, H., Madec, G., Fichefet, T., Ethe, C., & Duliere, V.
 1223 (2005). On the representation of high latitude processes in the ORCA-LIM
 1224 global coupled sea ice–ocean model. *Ocean Modelling*, *8*(1-2), 175–201. Re-
 1225 trieved from <https://doi.org/10.1016/j.ocemod.2003.12.009>
- 1226 Zuo, H., Balmaseda, M. A., Tietsche, S., Mogensen, K., & Mayer, M. (2019). The
 1227 ECMWF operational ensemble reanalysis–analysis system for ocean and sea
 1228 ice: a description of the system and assessment. *Ocean science*, *15*(3), 779–
 1229 808. Retrieved from <https://doi.org/10.5194/os-15-779-2019>

1
2
3
4
5
6
7
8
9
10
11
12

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

Alan Geer¹

¹European 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

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

69 The aim of this work is to build tools and techniques for direct assimilation of satel-
70 lite microwave radiances for the cryosphere and in particular, the sea ice. This work is
71 ultimately intended for use in an operational weather forecasting system, to allow the
72 assimilation of microwave radiances over sea ice surfaces and to contribute to analyses
73 of the sea ice concentration. Microwave radiances have a strong sensitivity to sea ice, par-
74 ticularly its fractional coverage but also ice age, snow depth and snow microstructure.
75 Operational methods to retrieve the sea ice concentration use largely heuristic (‘tie point’)
76 approaches (e.g. Comiso et al., 2003; Spreen et al., 2008; Sandven et al., 2023) and an
77 experimental method using optimal estimation used a simplified forward model (e.g. Scar-
78 lat et al., 2020). This is because physical modelling of snow and ice radiative transfer
79 at microwave frequencies is difficult, with better results below 20 GHz than above (e.g.
80 Hirahara et al., 2020). A fundamental issue is the importance of centimetre to micron
81 scale details of the snow and ice, including air inclusions in ice or grain shapes and their
82 layout in snow, that are relevant to the radiative transfer. Physical models of snow ra-
83 diative transfer (e.g. Picard et al., 2018) are starting to provide convincing results up
84 to around 250 GHz over land, for example linking measurements from snow pit profiles
85 to aircraft radiance measurements with reasonable accuracy in late winter in the Cana-
86 dian Arctic (Sandells et al., 2023). Physical models for sea ice have also shown poten-
87 tial (Kang et al., 2023), though with the necessity of empirically adjusting the snow and
88 sea ice optical properties to fit each location. For global assimilation of snow and ice sur-
89 faces, observation models will need to be reliable across all seasons and hemispheres, and
90 will need to handle many different types of ice and snow, including wet and refrozen snow
91 in the warmer seasons. There is also the problem of defining the input state, since it is
92 impossible to provide snow pits and ice cores globally.

93 An alternative path to improving our modelling frameworks is to use the vast amount
94 of existing observational data to improve models in an empirical way (e.g. Schneider et
95 al., 2017; Geer, 2021; Gettelman et al., 2022). The idea of training models from obser-
96 vations has become more feasible in recent years following the advent of easy-to-use ma-
97 chine learning tools such as Keras (Chollet et al., 2015) and rapid progress has been made
98 in machine learning alternatives to weather forecasting models (e.g. Lam et al., 2022).
99 A partly empirical approach is taken in the current work, with the aim to train a new
100 observation model for sea ice using global data through the whole year, in order to han-
101 dle all seasons, hemispheres and snow and ice types. However, empirical geophysical mod-
102 els are normally built using ‘supervised learning’ strategies that require the inputs and
103 outputs of the empirical model to be known and available in large quantities. For ex-
104 ample, in the case of ML-based weather forecasting models, this relies on existing atmo-
105 spheric re-analysis datasets (e.g. Hersbach et al., 2020). But to train an empirical ob-
106 servation model to link the geophysical state to the satellite radiances for the land sur-
107 face or the cryosphere, the inputs, in other words the state of the soil, snow and ice, are
108 not well enough known on a global scale. In fact, we will likely only know the state pa-
109 rameters well enough on a global scale once an observation model is available to help de-
110 rive them from satellite observations, so training an empirical model for surface radia-
111 tive transfer can be seen as a chicken and egg problem.

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

117 scribe underlying statistical properties of, for example, the output snowflake shapes. A
 118 similar principle is used in the current work, which allows the empirical observation model
 119 to define its own latent space as its input, in order to represent the mostly unknown mi-
 120 crophysical structures of sea ice and snow that are needed to simulate microwave radi-
 121 ances. This will be referred to as an ‘empirical state’, which will be trained simultane-
 122 ously with the empirical model to go from the state to the observations. The meaning
 123 of the empirical state variables is learnt along with the empirical model.

124 However, a purely empirical framework is of no use if we want to infer geophysi-
 125 cal properties, not empirical variables. To solve this, we can impose known physical re-
 126 lationships, such as the radiative transfer of the atmosphere. Reichstein et al. (2019) de-
 127 scribed how empirical models could be used and trained in networks with physical mod-
 128 els. These networks can be represented in a Bayesian way that is mathematically equiv-
 129 alent to data assimilation (Geer, 2021). The Bayesian viewpoint brings the insight that
 130 physical models represent parts of the network where there is good prior physical knowl-
 131 edge, whereas machine learning models can represent areas where there is little or no prior
 132 knowledge. In the current work, the empirical modelling is limited to the surface emis-
 133 sivity of the sea ice, whereas physical modelling is used for the surface emissivity of the
 134 ocean and for the radiative transfer of the atmosphere. In particular, it is the inclusion
 135 of physical equations that allows the sea ice concentration to be derived within the sys-
 136 tem, but the equations would have been incomplete without the empirical contribution.
 137 The entire network can be referred to as a hybrid empirical-physical model.

138 A number of other aspects of geophysical modelling and observation are more well-
 139 developed in data assimilation frameworks than in typical machine learning methods (Geer,
 140 2021). One of these is mapping between geophysical fields on a regular time and space
 141 grid to the irregular arrangement of satellite observations in time and space, which is han-
 142 dled by an interpolation operator. Also used is background error modelling (to control
 143 the weight of prior physical knowledge against the new knowledge from observations) and
 144 observation errors (to account for the varying difficulty of forward modelling at differ-
 145 ent frequencies measured by the satellite). All these techniques are also employed in the
 146 current work, making a mix of data assimilation and machine learning, following much
 147 work merging the two fields (e.g. Hsieh & Tang, 1998; Bocquet et al., 2020; Brajard et
 148 al., 2020).

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

159 2 Method

160 2.1 Overview

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

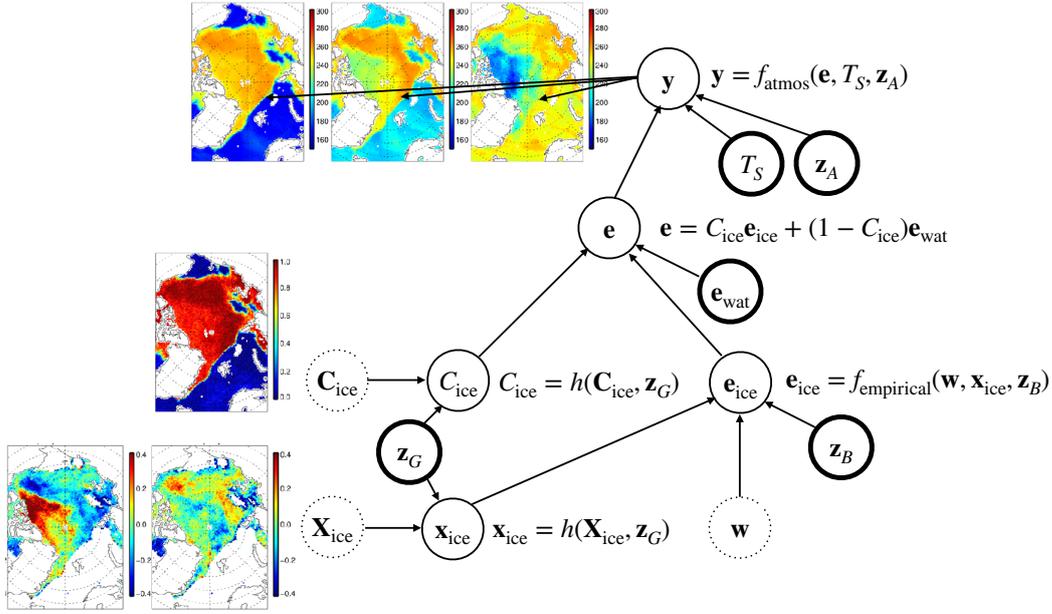


Figure 1. Simplified overview of the hybrid empirical-physical training network, for a single day and a single multi-channel observation. Circles represent variables that are trainable (dotted), dependent (thin solid), or fixed (thick solid). Arrows indicate dependencies and equations give the functional 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.

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

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

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

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

191 The hybrid network is trained by minimising a loss function. The aim is to be able
192 to replicate as closely as possible a year of Advanced Microwave Sounding Radiometer-
193 2 (AMSR2) observations over ocean and sea ice areas poleward of 45° latitude. The sim-
194 ulated observation \mathbf{y} (Fig. 1) represents an observation at one location comprising mea-
195 surements at different frequencies and polarisations, known as channels. The notation
196 here is that variables in lower-case bold represent vectors. The measured quantity is ra-
197 diance and this is given in terms of equivalent black body brightness temperature (TB).
198 Further detail on the microwave observations is in Sec. 22.3. The observed AMSR2 ob-
199 servations are \mathbf{y}_{obs} and the discrepancy with the simulations is measured by a mean squared
200 error loss function J_{obs} , which weights observations by an observation error \mathbf{r} following
201 standard DA practice:

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

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

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

222 The mixed surface emissivity \mathbf{e} must be estimated within the network to provide
223 the surface boundary conditions to the atmospheric radiative transfer model. The emis-
224 sivity is described as a linear combination of the sea ice and ocean water surface emis-
225 sivities \mathbf{e}_{ice} and \mathbf{e}_{wat} , weighted by the sea ice concentration in the satellite’s field of view,
226 C_{ice} . The contrast in the surface emissivity between ocean and ice is strongest and eas-
227 iest to detect at microwave wavelengths, and is fundamentally what allows the inference
228 of the sea ice concentration from from satellite observations. This equation also defines
229 the meaning of sea ice concentration in this work, which is the fraction of the ocean sur-
230 face covered with ice or snow, which excludes melt ponds but can include ice bergs and
231 ice shelves. The ocean surface emissivity is described physically and is treated as a known
232 quantity. Not included in the simplified diagram is a windspeed error correction that is
233 applied to the ocean surface emissivity, which uses 10 trainable weights. This correction
234 is not a major aspect of the problem and is further described in Sec. 2.5.

235 The sea ice concentration and its emissivity are the key unknowns in this work. To
236 describe the latter, the network includes a time-evolving geographical map of sea ice con-
237 centration \mathbf{C}_{ice} . Here the capitalised bold notation denotes a geographical map, possi-
238 bly 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}_{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}_{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}_{ice}) are interpolated from a geographical map \mathbf{X}_{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:

$$J = J_{\text{obs}} + J_{\text{seaice_bounds}} + J_{\text{seaice_tsfc}} + J_{\text{emis}} + J_{\text{bias}}. \quad (2)$$

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}_{ice} (366 days, see Sec. 2.7)	22,874,634
	Empirical micro- and macro-physical properties \mathbf{X}_{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}_{emis}	10
	TOTAL	91,311,120
	Fixed parameters	Atmospheric radiative transfer \mathbf{z}_A
Surface (skin temperature (T_S, \mathbf{z}_B), surface wind-speed, ocean emissivity \mathbf{e}_{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	

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

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

302 Training of the hybrid model and geographical maps was done on the ECMWF super-
303 computer 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

304 maximum 48 h of wallclock time (no GPUs were available). This allowed up to 8 train-
305 ing epochs, which was sufficient in the current work. To train the model, the loss func-
306 tion (Eq. 2) was minimised using the Adam variant of the mini-batch stochastic gradi-
307 ent descent approach (Kingma & Ba, 2014) at its default settings, including its learn-
308 ing rate set to 0.001, with the exception of the use of a batch size of 1024 (Appendix A1).
309 From the point of view of Keras training, the fixed parameters were treated as input val-
310 ues (features), the observations were treated as output values (labels). Further techni-
311 cal aspects of the training and hyperparameters are explored in Appendix A1.

312 2.3 Observations

313 Observations are provided by AMSR2 (Okuyama & Imaoka, 2015), a conical scan-
314 ning microwave radiometer on the polar-orbiting GCOM-W2 satellite observing at fre-
315 quencies from 6.7 GHz to 89 GHz, each separately measuring vertically (v) and horizon-
316 tally (h) polarised radiances (Tab. 2). The original radiance observations were obtained
317 from the Japan Aerospace Exploitation Agency (JAXA, <https://gportal.jaxa.com>) and
318 averaged onto a 40 km reduced Gaussian grid (Hortal & Simmons, 1991). This averag-
319 ing standardises the measurement footprint, which varies with frequency (the footprint
320 varies both in size, Table 2, but also in central location by up to around 4 km). The spa-
321 tial locations of the resulting ‘superobservations’ or superobs are taken as those of the
322 centres of the Gaussian grid points. The times of the superobs are the mean of the times
323 of the original observations, which span only a few seconds, so the superob remains an
324 almost instantaneous observation of the earth. AMSR2 also has channels at 6 – 7 GHz
325 with excellent sea ice sensitivity but these have been left out due to their footprint be-
326 ing larger than the 40 km grid. Superobs are based on an average of 20 raw observations;
327 those based on less than 6 raw observations are discarded. Following current practice
328 in atmospheric data assimilation (e.g. Kazumori et al., 2016; Geer et al., 2018), the ob-
329 servations are used in all-sky conditions, i.e. clear, cloudy and precipitating.

330 Figure 2 illustrates these observations over the Arctic ocean. At 10 GHz, v-polarised,
331 (10v) the sea ice areas are relatively easily distinguished from ocean by the large con-
332 trast in brightness temperature, with sea ice showing a fairly uniform value around 250 K.
333 But at higher frequencies (19v – 89v) and in the horizontally-polarised channels (10h –
334 89h) the sea ice areas show more strongly variable brightness temperature, driven by the
335 micro and macrophysical characteristics of the sea ice and snow cover. The influence of
336 the atmosphere (e.g. clouds and water vapour) is also important, particularly at 89 GHz.

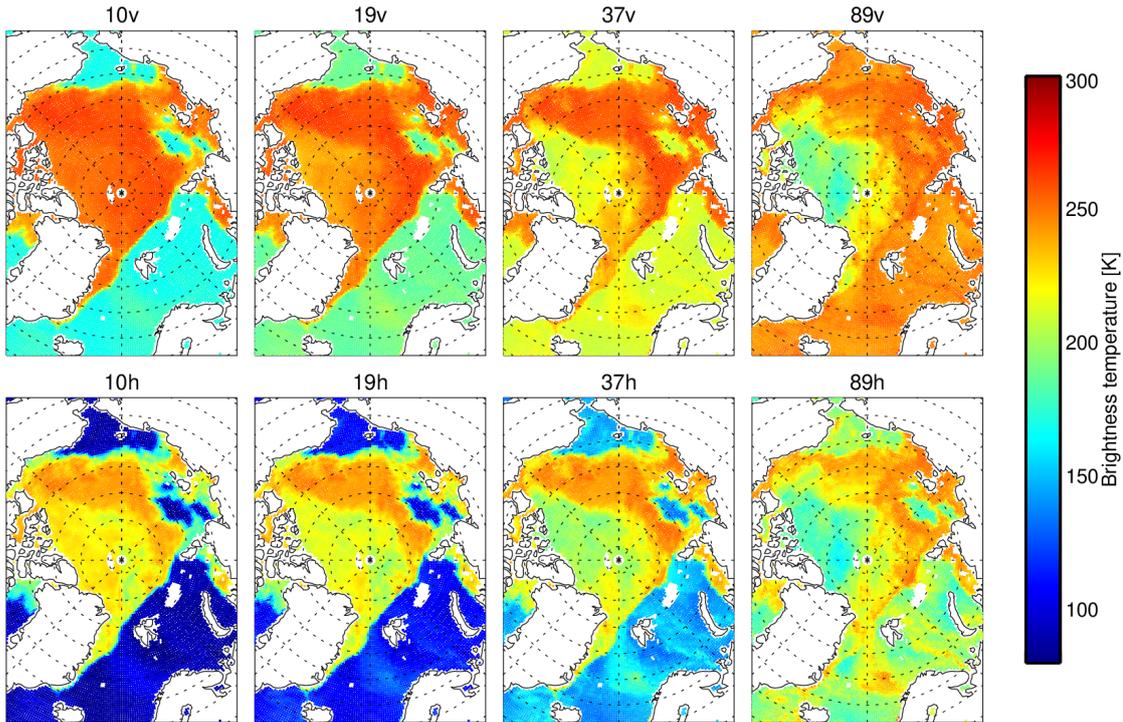


Figure 2. Observed AMSR2 brightness temperatures on the 40 km grid over the Arctic on 7th November 2020, showing only pure ocean scenes, and allowing observations from multiple orbits to overlay each other. Channels 24v and 24h are also used in this work, but are omitted from the figure to save space.

337 The ocean surface generally has much lower brightness temperatures than the sea ice and
 338 is more strongly polarised, so that h-polarisation measurements have much lower bright-
 339 ness temperatures than v-polarisation measurements, down to around 85 K at 10h. Ex-
 340 isting heuristic sea ice algorithms use higher frequencies to obtain better spatial reso-
 341 lution, and are based on empirically observed characteristics of the polarisation and fre-
 342 quency dependence of ocean and sea ice surfaces. These include methods using 19v, 37v,
 343 19h and 37h (e.g. Comiso et al., 2003; OSI-SAF, 2016) and those using 89v and 89h (e.g.
 344 Spreen et al., 2008) but in general no existing algorithm uses a physical description of
 345 the problem like the hybrid network used here (Fig. 1). This is mainly due to the lack
 346 of knowledge of the sea ice emissivity and its variations with underlying micro- and macro-
 347 physical ice and snow characteristics. The heuristic approaches can be vulnerable to at-
 348 mospheric effects such as cloud and water vapour along with variations in sea ice and
 349 ocean surface characteristics, which could incorrectly be interpreted as variations in sea
 350 ice concentration.

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

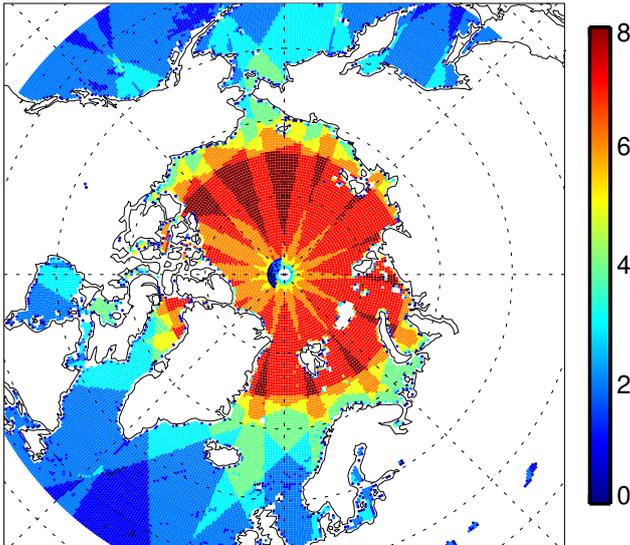


Figure 3. Number of AMSR2 pure ocean superobs available on the 40 km grid over the Arctic on 7th November 2020, in the 24 h period from 2100 UTC on 6th November to 2100 UTC on 7th November. Dashed lines indicate latitude every 10° and longitude every 45° .

360 the sea ice grid are on the same standardised locations. The result is that the daily es-
361 timates of sea ice concentration, and its the empirical properties, typically have to fit be-
362 tween 4 and 8 AMSR2 superobs and represent, in rough terms, a daily average.

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

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

368 The bias corrections are a function of the sea ice concentration, and are per channel, so
369 there is a vector of 10 bias corrections for for sea ice, \mathbf{b}_{ice} , and similarly for ocean wa-
370 ter, \mathbf{b}_{wat} . These are also trainable parameters, though in practice they are not allowed
371 to vary much. The bias correction is initialised with the background values $b_{\text{ice,bg}}$ and
372 $b_{\text{wat,bg}}$, respectively set to 2.5 K and 5.0 K (in all channels) using prior estimates of AMSR2
373 biases over ocean and land surfaces (Geer et al., 2022) and assuming that sea ice sur-
374 faces will have similar biases to those seen over land surfaces. Given that the bias cor-
375 rections are trainable values, they are constrained by a loss term

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

376 Here, the per-channel bias corrections are $b_{\text{ice},j}$ and $b_{\text{wat},j}$ with a background error of $b_{\text{bgerr}} =$
377 0.001 K, and the division by n , the number of training observations, standardises the loss
378 function to the Keras approach (Eq. 1). Given the very small chosen background error,
379 in practice the bias is forced to stay extremely close to prior values, rather than being
380 allowed to evolve to fit the observations. Hence the bias loss term exists mainly to ex-
381 plore the possibility of relaxing this constraint and evolving the bias correction in the
382 sensitivity tests described in A2. These show that if biases are allowed to evolve away

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

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

402 2.4 Atmospheric radiative transfer

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

$$y_{jk} = e_j T_S \Gamma_{jk} + (1 - e_j) T_{jk}^{\downarrow} \Gamma_{jk} + T_{jk}^{\uparrow}. \quad (5)$$

414 Here, T_{jk}^{\downarrow} is the downwelling radiation (TB) at the surface, T_{jk}^{\uparrow} is the component of up-
 415 welling radiation at the top of the atmosphere coming from the atmosphere itself, and
 416 Γ_{jk} is the atmospheric transmittance. The surface is represented using the approxima-
 417 tion of specular reflection, where the surface emits radiation according to the surface tem-
 418 perature T_S multiplied by the surface emissivity e_j , and reflects downwelling radiation
 419 modulated by a reflectivity $1 - e_j$. This is not a perfect assumption, because microwave
 420 radiation can penetrate centimetres to metres into snow and ice surfaces, and future work
 421 will use a more physical representation of the radiative transfer within the sea ice and
 422 snow (e.g. Picard et al., 2018). The final ‘all-sky’ brightness temperature is obtained by
 423 weighting the two sub-columns according to the effective cloud fraction C_{eff} (Geer et al.,
 424 2009):

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

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

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

441 2.5 Surface emissivity model

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

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

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

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

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

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

463 The trained values of \mathbf{w}_b and \mathbf{w}_0 to \mathbf{w}_3 are illustrated in Fig. 4 as well as being tab-
 464 ulated in Appendix B. Here, \mathbf{w}_b broadly represents a mean ice surface emissivity and
 465 other vectors describe variability around this as a function of the input values. Of these
 466 inputs, z_B is the only physical one: it is a scalar transformation of the surface temper-
 467 ature $z_B = \max(273.0 - T_S, 0.0)/30.0$. The factor 30.0 in the denominator is for nor-
 468 malisation and is chosen so that maximum values do not much exceed 1, as is common
 469 practice in machine learning. As mentioned in Sec. 2.4, the representation of the surface
 470 using an emissivity and a skin temperature is a big approximation. Quite often the ra-
 471 diation may be coming from within the sea ice or snow layer, where it is warmer, and
 472 the effective radiating temperature of the snow and sea ice (in Eq. 5) should be higher
 473 than the given skin temperature from the ECMWF model. To partly compensate this,
 474 the model is designed to allow an increase in the surface emissivity as the skin temper-
 475 ature gets colder. Figure 4a shows the modelled surface emissivity at $T_S = 273.0$ K ($z_B =$
 476 0) and $T_S = 261.0$ K ($z_B = 0.4$), assuming all other inputs are zero. Going to the lower

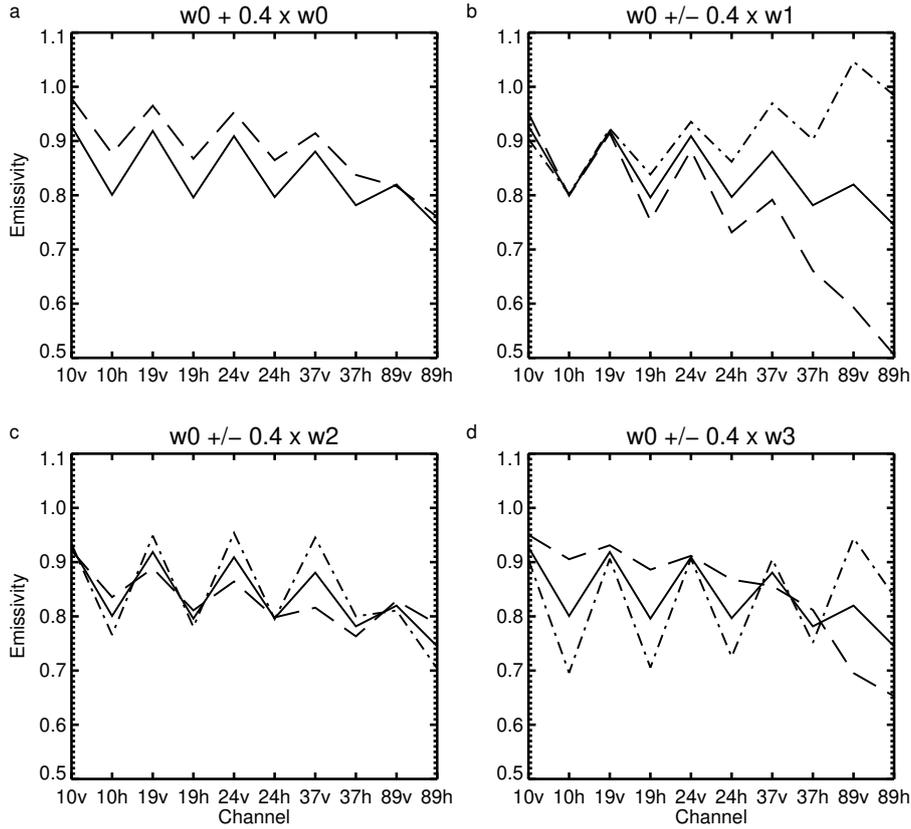


Figure 4. Illustrating the trained sea ice emissivity model. The ice emissivity with all inputs zero is w_b which is shown in all panels (solid line). If just one input parameter is perturbed to $+0.4$ (long dash) or -0.4 (dot dash) then the emissivity changes according to weights w_0 to w_3 .

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

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

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

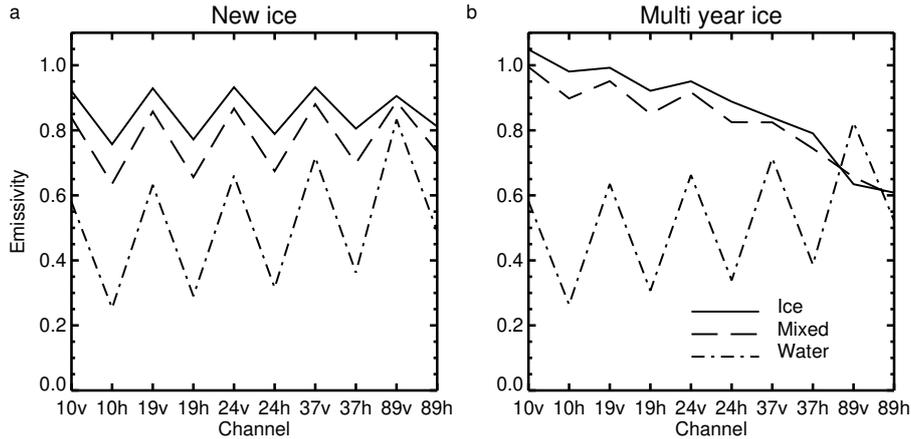


Figure 5. Emissivity as a function of channel at illustrative locations in the Arctic on 7th November (a) in the multi year ice and (b) in newly formed sea ice. The mixed emissivity (dashed) is generated from the sea ice concentration and the ice and water emissivities (solid and dot-dash) according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

496 vector \mathbf{w}_b , which gives the baseline surface emissivity for channel 10v:

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

497 In practice, the size of the background error here, $w_{\text{bgerr}} = 0.00001$, was chosen to keep
 498 $w_{b,1}$ extremely close to its background value. An emissivity of 0.93 is within the typical
 499 values for channel 10v, and further, the emissivity in this channel is thought to be
 500 relatively invariant between first year and multi year ice (Lee et al., 2017, their Table
 501 4). The posterior sensitivity tests in Appendix A2 show that this constraint on 10v emis-
 502 sivity could in practice have been relaxed or removed. In early testing before the addi-
 503 tion of sea ice concentration loss terms, the emissivity loss term was necessary to avoid
 504 completely non-physical sea ice concentrations being retrieved. However, it appears that
 505 the sea ice concentration loss terms are more effective and targeted, and would be suf-
 506 ficient on their own in future versions of this work. In any case, despite the constraint,
 507 the final trained model still has freedom to adjust the output emissivity at 10v down to
 508 at least 0.9 and up to at least 0.98 depending on the input parameters (Fig. 4).

509 Figure 5 illustrates sea ice surface emissivities generated by the empirical model
 510 (Eq. 8) at two locations in the Arctic, alongside ocean water emissivities (Eq. 10, to be
 511 described shortly) and the mixed-surface emissivity that is determined by the sea ice con-
 512 centration (Eq. 7). The location with new ice (panel a) generates relatively strong po-
 513 larisation but has little frequency dependence. The frequency dependence is in qualita-
 514 tive agreement with aircraft observations of new ice (e.g. Hewison & English, 1999). In
 515 contrast, a location in the multi-year ice is mostly unpolarised but its emissivity drops
 516 with frequency, as is generally observed (e.g. Baordo & Geer, 2015). Both sea ice emis-
 517 sivity spectra contrast strongly with the ocean water emissivity which is even more po-
 518 larised and increases with frequency. These are locations where the sea ice concentra-
 519 tion is close to 1, so the mixed-surface emissivity is relatively close to the ice emissiv-
 520 ity.

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

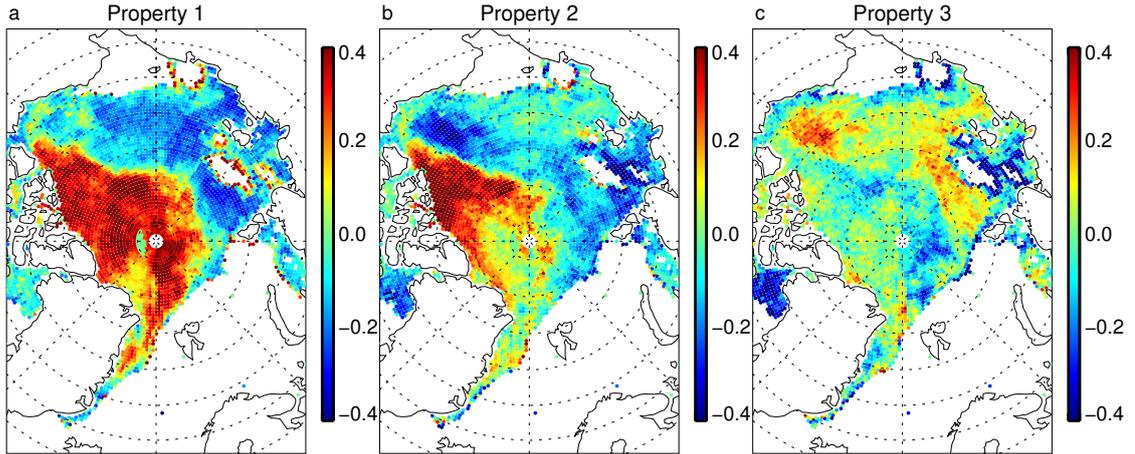


Figure 6. Estimated values of the three empirical sea ice properties in the Arctic on 7th November 2020. Values are only shown where the estimated sea ice concentration is greater than 0.2.

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

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

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

540 2.6 Empirical state representing micro- and macro-physical properties 541 of sea ice

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

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

556 2.7 Sea ice

557 In order to speed up the training, the sea ice maps \mathbf{C}_{ice} were initialised with a monthly
 558 mean sea ice concentration computed from the existing ECMWF sea ice concentration
 559 analysis at the AMSR2 locations. However, apart from the physical constraints described
 560 below, the sea ice concentration can be freely adjusted to best fit the observations, and
 561 is not constrained in any way to the ECMWF sea ice concentration. The existing ECMWF
 562 sea ice analysis is based on a combination of an ocean model, a sea ice model (LIM2, Tim-
 563 mermann et al., 2005) and the assimilation of highly processed observations, using the
 564 ocean data assimilation framework (OCEAN5, Zuo et al., 2019; de Rosnay et al., 2022)
 565 and will be referred to by this name subsequently. The observations are the OSTIA sea
 566 ice dataset (Good et al., 2020) which is ultimately based on sea ice retrievals (OSI-SAF,
 567 2016) obtained from a microwave sensor similar to AMSR2 using a traditional heuristic
 568 approach (e.g. Comiso et al., 2003). It is important that the new emissivity model
 569 should not try to fit these heuristic assumptions in any way. Further, due to the com-
 570 plex processing chain, the OCEAN5 sea ice can be at least 48 h behind reality once it
 571 has been mapped to observation locations (e.g. Baordo & Geer, 2015; Browne et al., 2019;
 572 de Rosnay et al., 2022, this work). As encountered when prototyping the current approach,
 573 if there is an incorrect sea ice concentration in Eq. 7, the sea ice emissivity model can
 574 take on characteristics of the ocean water surface emissivity, which would make the model
 575 essentially useless. For all these reasons, it was important not to allow the new sea ice
 576 analysis to be constrained by the OCEAN5 sea ice analysis in any way.

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

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

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

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

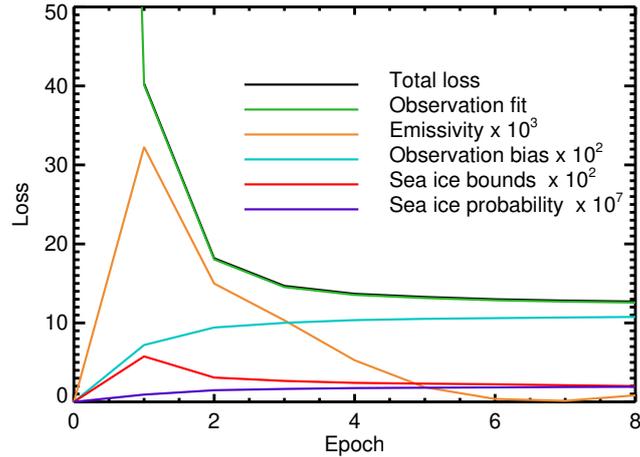


Figure 7. Loss terms during the training of the hybrid network (see Table 1) The total loss term is dominated by the observation fit, so it is mostly hidden behind the observation fit losses; the total loss at epoch zero is 340 and is off scale. The order of lines in the key follows the top-to-bottom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

601 sea ice values, the full loss term was:

$$J_{\text{seaice_tsfc}} = \frac{1}{n} \sum_d 4.0 \times \max(T_{S,d} - 273.2, 0.0) \quad (12)$$

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

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

617 3 Results

618 3.1 Training and fit to observations

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

625 the occurrence of non-physical sea ice concentrations. The loss terms have a very wide
 626 range of magnitudes and hence all but the observation term have required re-scaling to
 627 be visible on this figure. Despite this, the results show that most of these additional smaller
 628 loss terms are effective (see appendix). The hybrid model appears mainly converged after
 629 8 epochs, but Appendix A1 (using just a month of training data) investigates the use
 630 of larger numbers of epochs, and shows that the model can continue to evolve, but at
 631 the cost of some overfitting. Training on the yearly dataset is resource-limited to 8 epochs,
 632 but a beneficial, though involuntary, side effect has been that it is using ‘early stopping’
 633 to avoid over-fitting.

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

645 The RMS of the initial and final (or analysis) departures ($y_{\text{obs},ij} - y_{ij}$) are given
 646 in Table 2. Initial RMS departures are of order 10 K to 40 K. Panel g illustrates these
 647 departures in channel 37v, with large discrepancies coming both from the incorrect sur-
 648 face emissivity and the incorrect sea ice initial field. By contrast, the analysis departures
 649 (e.g. panel h) are mostly within ± 3 K and rarely larger than 10 K. The largest remain-
 650 ing departures are mainly found over open ocean and not sea ice, and come from errors
 651 in the location of cloud and precipitation in the ECMWF background fields, which can-
 652 not be adjusted in the current network. The RMS of the analysis departures ranges from
 653 2.7 K in channel 10v to 9.9 K in 89h (Table 2). These values are close to the prescribed
 654 observation errors, as intended.

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

663 At higher frequencies, the improvements between the initial model and the anal-
 664 ysis (Fig. 8 d and e) are driven not just by the improved sea ice concentration, but also
 665 by the development of the empirical terms of the surface emissivity model, and the em-
 666 pirical state inputs. Empirical properties 1 and 2 (Fig. 6) are the main driver in the anal-
 667 ysed brightness temperature at 37v, helping generate brightness temperatures down to
 668 210 K over the multi-year ice (empirical property 1 values of up to 0.4) and up to 260 K
 669 over the first year ice (empirical property 1 down to -0.4). Hence the hybrid model ap-
 670 pears to make a physically plausible set of adjustments in order to fit the observations.

671 3.2 Sea ice concentration

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

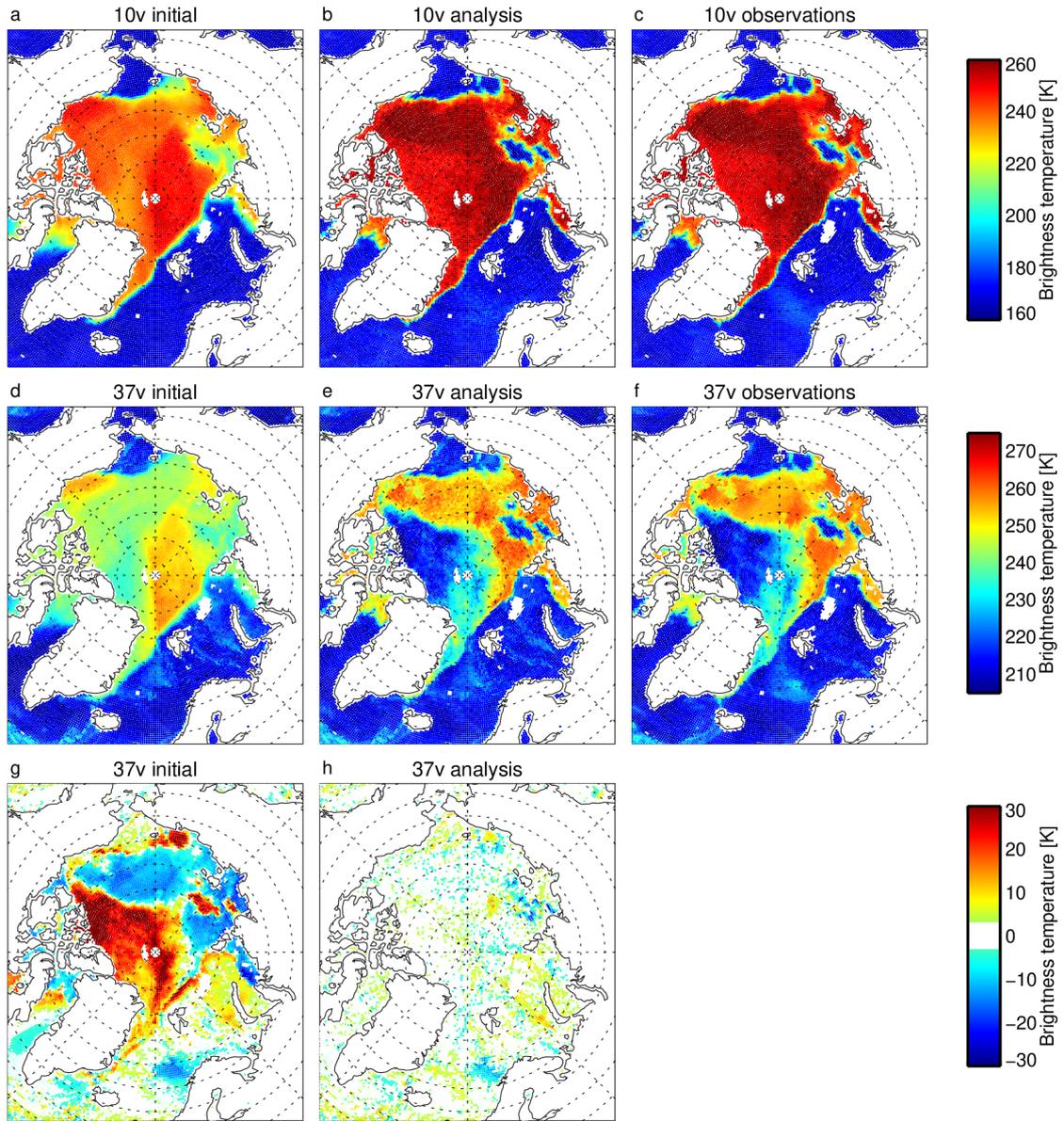


Figure 8. Comparison of simulated and observed brightness temperatures in channels 10v and 37v, over the Arctic on 7th November 2020: (a,d) initial guess; (b,e) after training, in other words the analysis; (c,f) observations; (g) the initial guess minus observation departure; (h) the analysis minus observation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

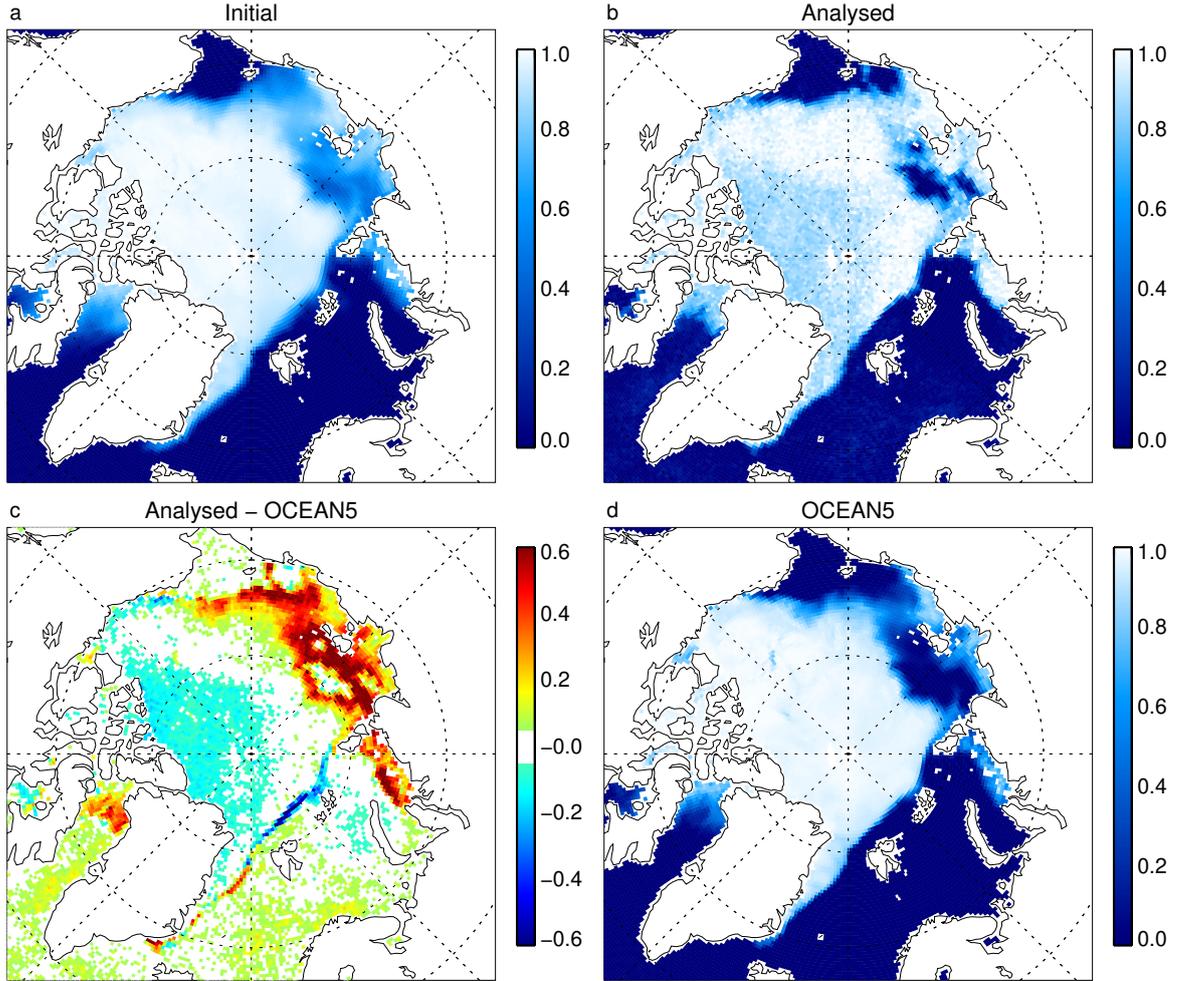


Figure 9. Comparison of sea ice concentration on 7th November 2020: (a) Initial monthly mean sea ice; (b) Analysed in this work; (c) Difference between analysed and OCEAN5; (d) OCEAN5 sea ice. Differences smaller than 0.05 are not shown. Latitude and longitude grid spacings are 10° and 45° respectively.

676 ysis (OCEAN5, panel d). There are also disagreements in the location of the ice edge
 677 by up to around 100 km in the region of Svalbard. The fundamental issue is the roughly
 678 48 h time delay in producing the OCEAN5 analysis; eventually the OCEAN5 sea ice catches
 679 up and provides a similar picture (not shown). Table 3 shows that the new analysed sea
 680 ice agrees best with OCEAN5 if it is artificially delayed by 2 or 3 days, consistent with
 681 previous expectations (Baordo & Geer, 2015; Browne et al., 2019; de Rosnay et al., 2022).
 682 There is also some low concentration (< 0.2) sea ice incorrectly identified in the cur-
 683 rent analysis, for example off the N coast of Scandinavia, visible mainly in the difference
 684 plot (panel c). This spurious sea ice is generated when clouds are present in the obser-
 685 vations but not in the ECMWF 12 h atmospheric forecast (z_A here). The network (Fig. 1)
 686 cannot adjust the atmosphere to add cloud and instead can only create sea ice to bet-
 687 ter fit the observations (an opposite effect would tend to create negative sea ice concen-
 688 trations where the ECMWF forecasts have too much cloud, but this is suppressed by the
 689 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

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

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

716 Figure 12 shows the annual cycle of ice area, using ‘cleaned up’ sea ice concentra-
 717 tions as described in the previous paragraph. Compared to OCEAN5, this work gives
 718 slightly higher ice area in the Antarctic winter, and slightly lower ice area in the Arc-
 719 tic winter. The better timeliness of the new sea ice analysis is also clear. In early Novem-
 720 ber in the Arctic, the current work identifies a rapid freezing event that is smoothed out
 721 and delayed in the OCEAN5 sea ice analysis, as also illustrated in Fig. 9. A similar pic-
 722 ture is seen in the Antarctic, from late February to mid-May. Apart from these issues,
 723 there is good agreement between the annual cycles in the two ice products.

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

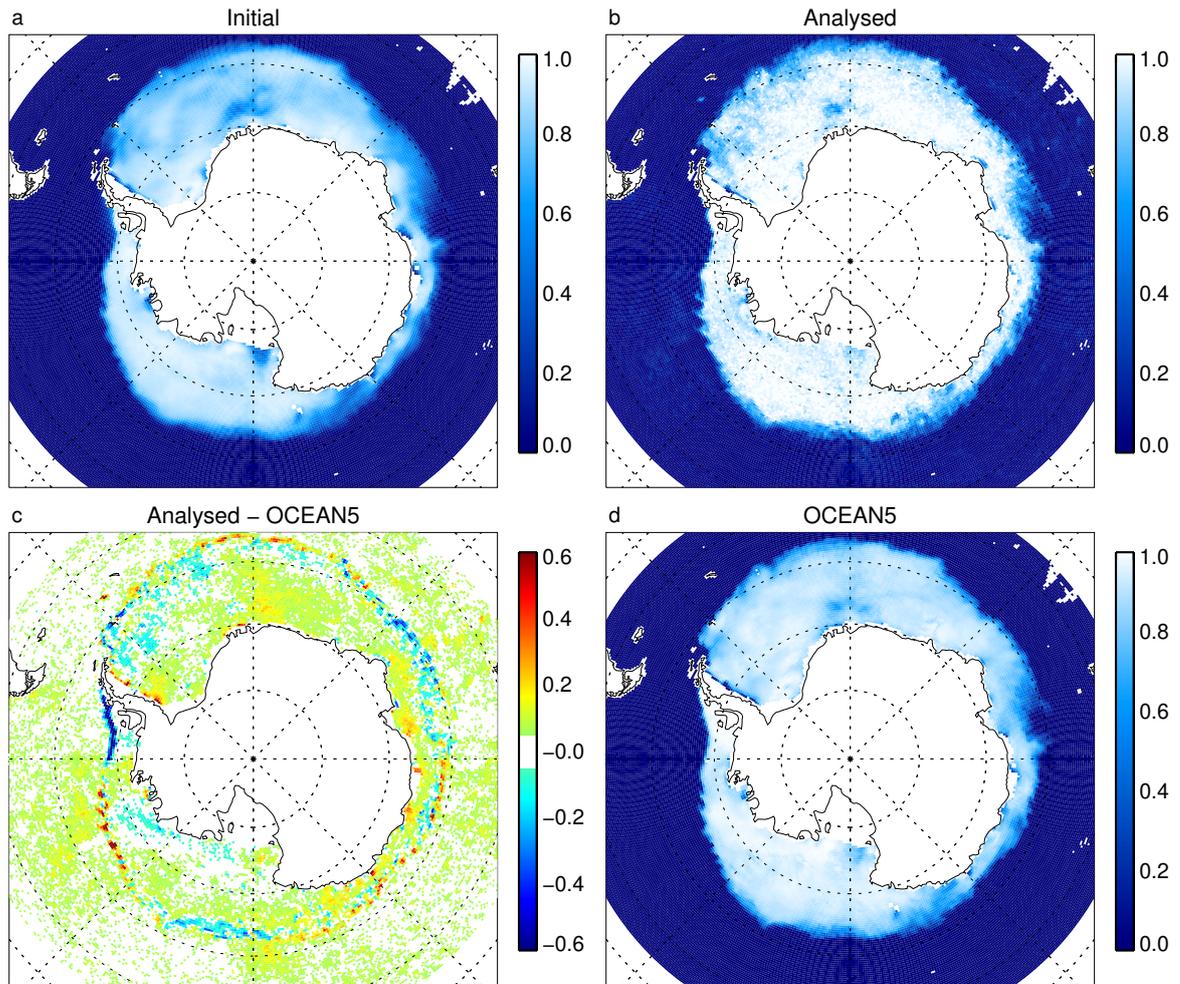


Figure 10. As Fig. 9 on 7th November 2020, but for the Antarctic. The Greenwich meridian is to the top.

727 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
 728 in the multi-year ice zone should be close to 1, like in OCEAN5, this appears to be
 729 a defect in the current approach and likely relates to the constraint of 10v ice surface
 730 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-
 731 year ice should have been allowed to decrease more at 10v. In the Antarctic (panel b)
 732 there is typically a small overestimation (around 0.02) compared to OCEAN5, and much
 733 larger positive differences in patches close to the Antarctic coast. These can be traced
 734 to the Antarctic coastline in the summer, January and February, where the new analyses
 735 show considerably more ice extent (much of it fractional ice, e.g. concentrations around
 736 0.5) along the coast than the OCEAN5 analyses. Given the careful treatment of the satellite
 737 field of view and land contamination in this work (Sec. 2.3) and the all-sky validity
 738 of the data this suggests there is a real defect in the OCEAN5 sea ice. There are small
 739 underestimations of the sea ice compared to OCEAN5 in the Weddel Sea which might
 740 also be due to the multi-year ice issue, since this is one of the few areas of the Antarctic
 741 where ice can persist from one year to the next. But apart from the areas of disagree-
 742
 743

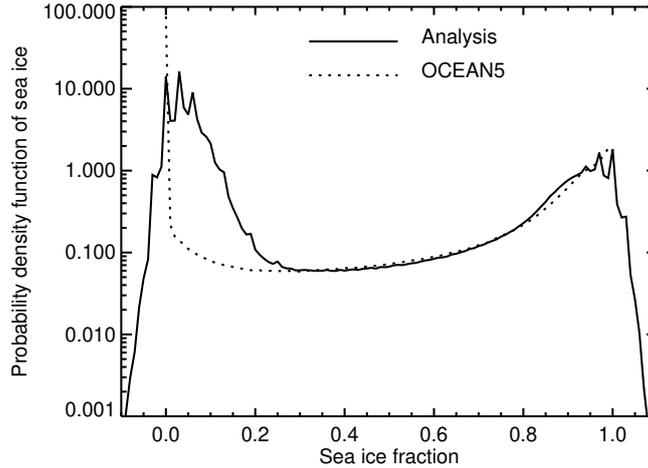


Figure 11. Probability density function for sea ice concentration for the new analysis and the existing OCEAN5, over the year and both hemispheres, using a log y-axis to better compare the full probability range.

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

746 4 Conclusion

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

761 There is no perfect truth against which to evaluate the resulting global daily maps
 762 of sea ice, since existing global satellite retrievals of sea ice are based on heuristic meth-
 763 ods which this work aims to replace, and in-situ measurements are very limited in cov-
 764 erage. The high quality of fit to observations suggests that the sea ice results are good,
 765 but comparisons are also made to the existing ECMWF sea ice analyses, generated by
 766 the OCEAN5 assimilation system. The new maps are 48 h to 72 h more timely than the
 767 sea ice concentration analysed in OCEAN5. The new maps have generally sharper res-
 768 olution including plausible mesoscale features like some apparently eddy-generated 300
 769 km wave features in the Antarctic sea ice edge. The new data also suggest there is a sub-
 770 stantial underestimate of sea ice cover in the OCEAN5 analyses in the Antarctic sum-
 771 mer. There are some limitations in the new data too, mainly an apparent underestimate

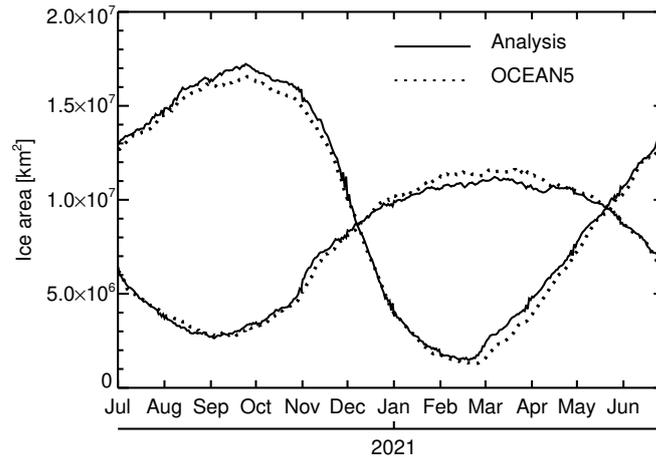


Figure 12. Estimated hemispheric sea ice area from the current work (solid) and OCEAN5 (dotted) from 1st July 2020 to 30th June 2021. The Antarctic ice extent is largest in September and the Arctic in March.

772 of the sea ice concentration in the Arctic winter multi-year ice, where it seems to be about
 773 0.1 too low, in compensation for remaining inaccuracies in the sea ice emissivity model.

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

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

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

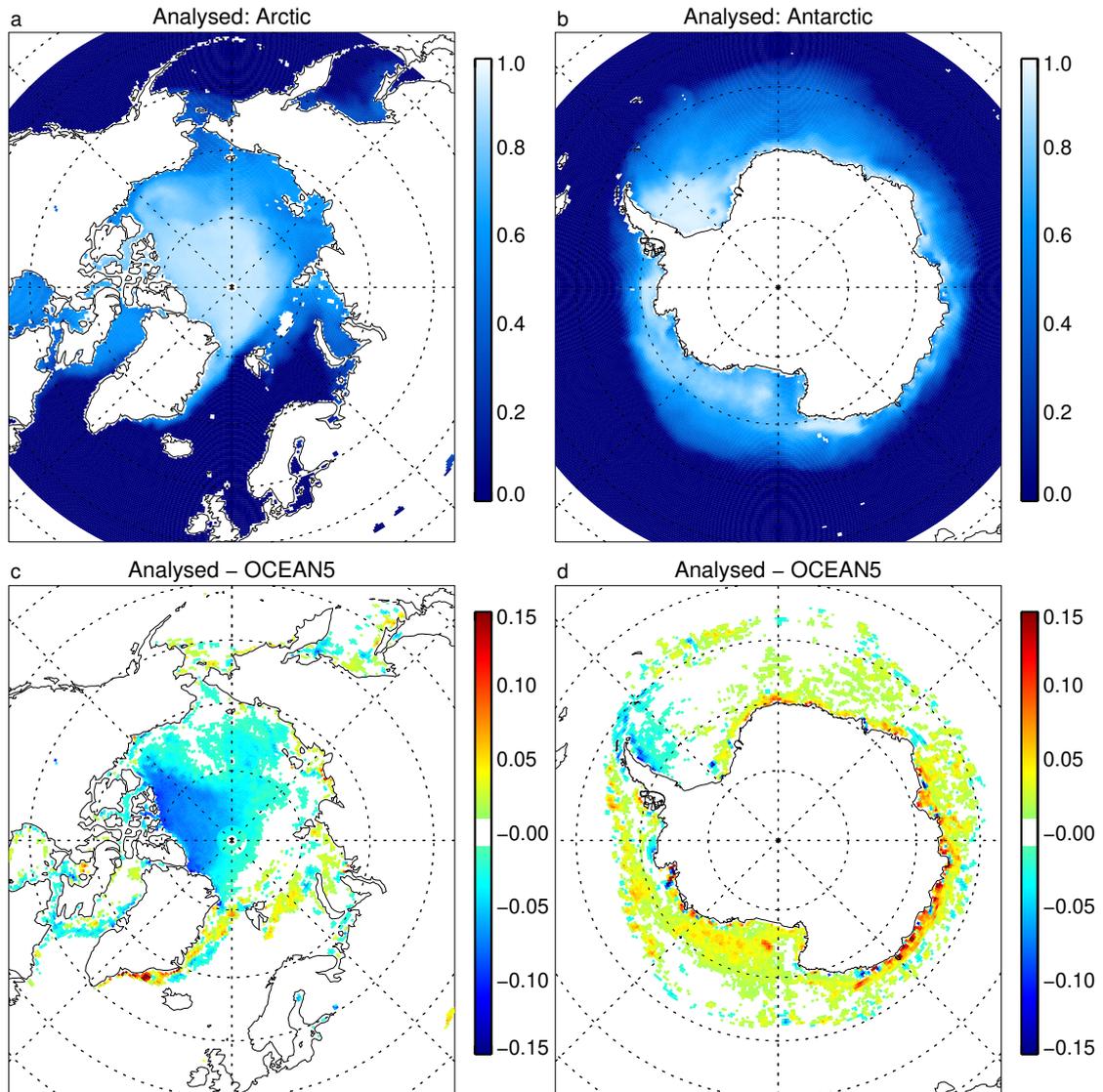


Figure 13. Annual mean sea ice concentration analysed in this work (a,b) and difference in annual mean between the current work and OCEAN5 (c,d). Differences smaller than 0.01 are not plotted.

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

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

824 The training of the hybrid physical-empirical model has demonstrated a number
 825 of innovations that have been facilitated by the availability of modern machine learn-
 826 ing and differential programming tools such as Keras and TensorFlow (Abadi et al., 2015).
 827 Considering the network in Fig. 1, if the sea ice emissivity was a known parameter, the
 828 lowermost empirical part of the network (devoted to the empirical state parameters and
 829 the empirical surface emissivity model) would be unnecessary, and this work would have
 830 been a standard data assimilation problem of estimating the sea ice concentration maps
 831 from the satellite observations. However, there would still be the novelty that it was im-
 832 plemented in a machine learning framework. If the sea ice concentration, and the micro-
 833 and macro-physical parameters of the sea ice and snow cover, were known globally through
 834 modelling or observations, those parameters could have been treated as known input vari-
 835 ables and the empirical sea ice emissivity model would be an ML component model trained
 836 inside an otherwise physical data assimilation network (e.g. Reichstein et al., 2019; Geer,
 837 2021) but not achieved practically as yet, to this author’s knowledge. But since the in-
 838 put variables are also unknown (the chicken and egg problem) the most novel aspect of
 839 this network is that it simultaneously trains an empirical model and works out what its
 840 inputs should be (the maps of sea ice parameters).

841 It is proposed to describe the simultaneous training of state and model as an ‘em-
 842 pirical state’ method, with the essential components being:

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

849 A similar approach could be extended to snow surfaces over land, soil moisture and veg-
 850 etation analyses from microwave sensors, and possibly to many other ‘chicken and egg’
 851 problems in the wider developments towards earth system assimilation. The approach
 852 can also be extended to a ‘hybrid state’, as in this work, where some of the inputs are
 853 physical and some empirical. By taking this mix of empirical and physical inputs, em-
 854 pirical methods can become progressively more physical, as models become able to sup-
 855 ply more and higher quality input parameters. For example, snow grain size, depth and
 856 temperature may eventually be available within the ECMWF model, and these could
 857 be included as inputs to the empirical surface emissivity model, and the number of em-
 858 pirical parameters representing unknown aspects of the snow and sea ice state could hope-
 859 fully be reduced. This means that empirical methods can be a quick way of getting started
 860 with a new assimilation domain, such as sea ice, but they can subsequently evolve to-
 861 wards more physical approaches as physical models become more capable within that
 862 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

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

881 **Appendix A Sensitivities - overview**

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

890 **A1 Sensitivities - epochs, batch size, network complexity**

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

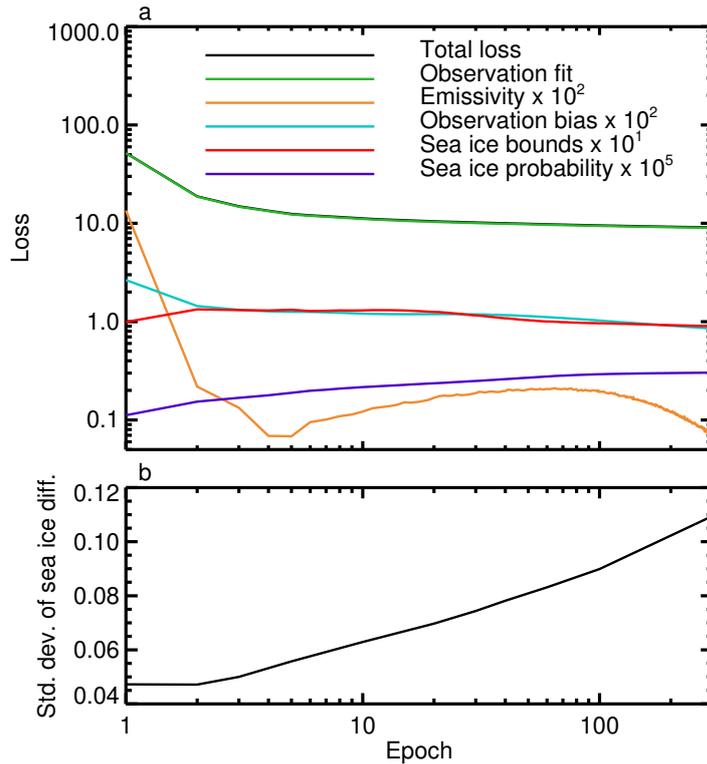


Figure A1. (a) Total loss and individual loss terms (scaled as noted in key) using an extended number of training epochs; (b) Standard deviation of the difference between analysed sea ice fraction with a 48 h delay and OCEAN5 sea ice fraction. Note the logarithmic x axis and, in panel a, also y axis,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

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

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

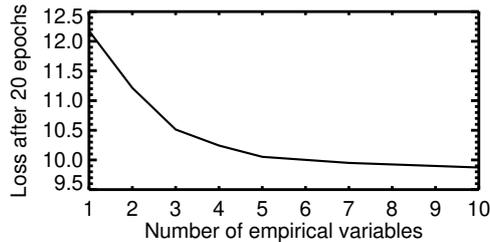


Figure A2. Total loss as a function of the number of empirical variables.

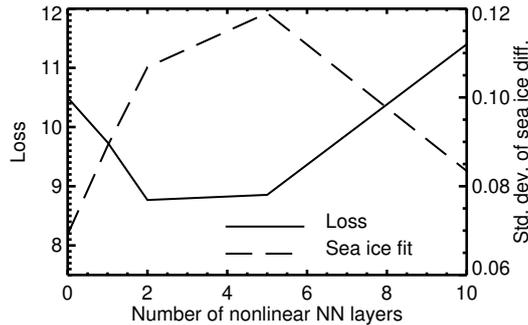


Figure A3. Total loss and sea ice fit to OCEAN5 with a 48 h offset, as a function of the number of neural network layers, based on 50 epochs of training and 20 neurons per layer. The data point for "0" layers is used to represent a single linear layer.

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

918 The chosen model for sea ice emissivity is linear, but the framework is perfectly ca-
 919 pable of training a nonlinear model, including deep networks. Sensitivity tests were per-
 920 formed using a sigmoid activation function and between 1 and 10 fully connected lay-
 921 ers using 20 neurons each. In these tests the number of epochs was set to 50 to ensure
 922 the deeper networks were converged. Fig. A3 shows the results in terms of the loss func-
 923 tion at 50 epochs and the fit of the sea ice field to ECMWF sea ice (with the 2 day off-
 924 set for best fit). The point marked 0 layers corresponds to the normal linear single layer
 925 model, but trained for 50 epochs. Going to a nonlinear activation function and adding
 926 up to 2 layers is capable of fitting the observations better, as indicated by the reductions
 927 in the loss function. However, this comes at the price of generating a poorer-quality sea
 928 ice field, as indicated by the increasing standard deviation of the difference with ECMWF
 929 / OCEAN5 sea ice becoming larger than 0.1. For 3 layers and greater, the picture re-
 930 verses, but this is likely because 50 epochs is insufficient to fully train the deeper net-
 931 works. In broad terms, adding multiple layers and nonlinearity to the surface emissiv-
 932 ity model seems to give greater possibilities for over-fitting the data, similar to increas-
 933 ing the number of epochs.

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

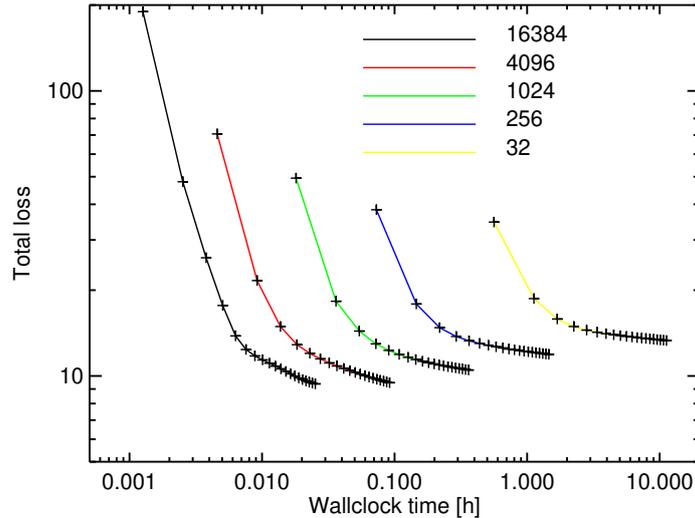


Figure A4. Total loss as a function of wallclock time for five different batch sizes, as indicated in the key. Crosses indicate the wallclock time and the corresponding loss after each of 20 epochs.

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

959 A2 Sensitivities - prior knowledge loss terms

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

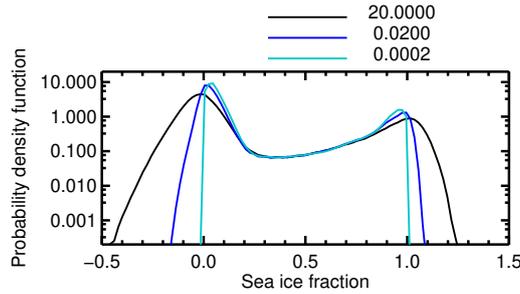


Figure A5. Probability density function of sea ice fraction for three settings of the sea ice background error, as given in the key.

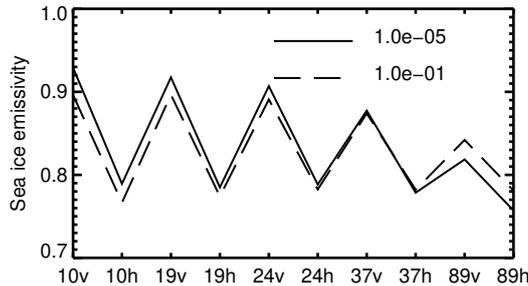


Figure A6. Sea ice surface emissivity baseline term w_b , as a function of AMSR2 channel identifier, for values of the surface emissivity background error (w_{bgerr} in Eq. 9) as indicated in the key.

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

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

981 Relaxing the bias background error b_{bgerr} in Eq. 4 to values larger than the chosen
 982 $b_{bgerr} = 0.001 K$ improves the fit to observations, and reduces the total loss from
 983 around 10.5 down to around 9.8 (no figure shown). However, this leads the network to
 984 generate unreasonably large bias corrections of up to -7 K over ocean and -16 K over sea
 985 ice (no figure shown). This likely shows that the problem is ill-posed when both TB bias
 986 and surface emissivity are allowed to vary. Hence the bias correction term is a very im-
 987 portant 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.

1020 Original AMSR2 data for this value added data product was provided by Japan
 1021 Aerospace Exploration Agency (JAXA, 2023) via the G-Portal. The user is entitled to
 1022 use JAXA G-Portal AMSR2 data free of charge without any restrictions (including com-
 1023 mercial use) except for the condition about acknowledgement of data credit as stipulated
 1024 in Article 7.(2) of the terms and conditions at [https://gportal.jaxa.jp/gpr/index/](https://gportal.jaxa.jp/gpr/index/eula?lang=en)
 1025 [eula?lang=en](https://gportal.jaxa.jp/gpr/index/eula?lang=en)

1026 The Python code was run on Python 3.8.8-01 (Python Software Foundation, 2021)
 1027 including Tensorflow and Keras 2.8.0 (Abadi et al., 2015; TensorFlow, 2021) on the ECMWF
 1028 ATOS supercomputer CPU nodes.

1029 Acknowledgments

1030 Niels Bormann, Fabrizio Baordo, Matthew Chantry and Tracy Scanlon are thanked for
 1031 reviews and suggestions.

1032 References

- 1033 Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., . . . Zheng, X.
 1034 (2015). *TensorFlow: Large-scale machine learning on heterogeneous systems*
 1035 [code]. Retrieved from <https://www.tensorflow.org/> (Software available
 1036 from tensorflow.org) doi: 10.5281/zenodo.4724125
- 1037 Baordo, F., & Geer, A. (2015). Microwave surface emissivity over sea-ice. *EUMET-*
 1038 *SAT NWP-SAF visiting scientist report NWPSAF_EC_VS_026*. Retrieved from
 1039 <https://nwpsaf.eu/publications/vs\reports/nwpsaf-ec-vs-026.pdf>
- 1040 Baordo, F., & Geer, A. J. (2016). Assimilation of SSMIS humidity-sounding chan-
 1041 nels in all-sky conditions over land using a dynamic emissivity retrieval. *Quart.*
 1042 *J. Roy. Meteorol. Soc.*, *142*, 2854–2866. Retrieved from [https://doi.org/10](https://doi.org/10.1002/qj.2873)
 1043 [.1002/qj.2873](https://doi.org/10.1002/qj.2873)
- 1044 Bauer, P., Moreau, E., Chevallier, F., & O’Keeffe, U. (2006). Multiple-scattering
 1045 microwave radiative transfer for data assimilation applications. *Quart. J. Roy.*
 1046 *Meteorol. Soc.*, *132*, 1259-1281. Retrieved from [https://doi.org/10.1256/qj](https://doi.org/10.1256/qj.05.153)
 1047 [.05.153](https://doi.org/10.1256/qj.05.153)
- 1048 Bauer, P., Thorpe, A., & Brunet, G. (2015). The quiet revolution of numerical
 1049 weather prediction. *Nature*, *525*(7567), 47–55.
- 1050 Berg, W., Bilanow, S., Chen, R., Datta, S., Draper, D., Ebrahimi, H., . . . Yang,
 1051 J. X. (2016). Intercalibration of the GPM microwave radiometer constellation.
 1052 *Journal of Atmospheric and Oceanic Technology*, *33*(12), 2639–2654. Retrieved
 1053 from <https://doi.org/10.1175/JTECH-D-16-0100.1>
- 1054 Bocquet, M., Brajard, J., Carrassi, A., & Bertino, L. (2019). Data assimilation
 1055 as a learning tool to infer ordinary differential equation representations of
 1056 dynamical models. *Nonlinear Processes in Geophysics*, *26*(3), 143–162.
- 1057 Bocquet, M., Brajard, J., Carrassi, A., & Bertino, L. (2020). Bayesian inference
 1058 of chaotic dynamics by merging data assimilation, machine learning and
 1059 expectation-maximization. *Foundations of Data Science*, *2*(1), 55–80. Re-
 1060 trieved from <https://doi.org/10.3934/fods.2020004>
- 1061 Brajard, J., Carrassi, A., Bocquet, M., & Bertino, L. (2020). Combining data as-
 1062 similation and machine learning to emulate a dynamical model from sparse
 1063 and noisy observations: A case study with the lorenz 96 model. *Journal of*
 1064 *Computational Science*, *44*, 101171.
- 1065 Browne, P. A., de Rosnay, P., Zuo, H., Bennett, A., & Dawson, A. (2019). Weakly
 1066 coupled ocean–atmosphere data assimilation in the ECMWF NWP system.
 1067 *Remote Sensing*, *11*(3), 234. Retrieved from [https://doi.org/10.3390/](https://doi.org/10.3390/rs11030234)
 1068 [rs11030234](https://doi.org/10.3390/rs11030234)
- 1069 Buehner, M., Caya, A., Pogson, L., Carrieres, T., & Pestieau, P. (2013). A new envi-
 1070 ronment canada regional ice analysis system. *Atmosphere-Ocean*, *51*(1), 18–34.

- 1071 Retrieved from <https://doi.org/10.1080/07055900.2012.747171>
- 1072 Chollet, F., et al. (2015). *Keras*. <https://keras.io>.
- 1073 Comiso, J. C., Cavalieri, D. J., & Markus, T. (2003). Sea ice concentration, ice tem-
1074 perature, and snow depth using AMSR-E data. *IEEE Trans. Geosci. Remote*
1075 *Sens.*, *41*(2), 243–252. Retrieved from <https://doi.org/10.1109/TGRS.2002>
1076 [.808317](https://doi.org/10.1109/TGRS.2002.808317)
- 1077 de Rosnay, P., Browne, P., de Boissésou, E., Fairbairn, D., Hirahara, Y., Ochi, K.,
1078 ... Rabier, F. (2022). Coupled assimilation at ECMWF: current status,
1079 challenges and future developments. *Quart. J. Roy. Meteorol. Soc.*, *148*,
1080 2672–2702. Retrieved from <https://doi.org/10.1002/qj.4330>
- 1081 Desroziers, G., Berre, L., Chapnik, B., & Poli, P. (2005). Diagnosis of observation,
1082 background and analysis-error statistics in observation space. *Quart. J. Roy.*
1083 *Meteorol. Soc.*, *131*, 3385–3396.
- 1084 Eyre, J. R., English, S. J., & Forsythe, M. (2020). Assimilation of satellite data in
1085 numerical weather prediction. Part I: The early years. *Quart. J. Roy. Meteorol.*
1086 *Soc.*, *146*(726), 49–68. Retrieved from <https://doi.org/10.1002/qj.3654>
- 1087 Geer, A. J. (2021). Learning earth system models from observations: Machine learn-
1088 ing or data assimilation? *Phil. Trans. R. Soc. A*, *379*, 20200089. Retrieved
1089 from <https://doi.org/10.1098/rsta.2020.0089>
- 1090 Geer, A. J. (2023a). *Code for simultaneous inference of sea ice state and surface*
1091 *emissivity model using machine learning and data assimilation* [code]. Re-
1092 trieved from <https://doi.org/10.5281/zenodo.10013542>
- 1093 Geer, A. J. (2023b). *Data for simultaneous inference of sea ice state and surface*
1094 *emissivity model using machine learning and data assimilation* [dataset]. Re-
1095 trieved from <https://doi.org/10.5281/zenodo.10033377>
- 1096 Geer, A. J. (2023c). Joint estimation of sea ice and atmospheric state from mi-
1097 crowave imagers in operational weather forecasting. , in preparation.
- 1098 Geer, A. J., Baordo, F., Bormann, N., English, S., Kazumori, M., Lawrence, H., ...
1099 Lupu, C. (2017). The growing impact of satellite observations sensitive to hu-
1100 midity, cloud and precipitation. *Quart. J. Roy. Meteorol. Soc.*, *143*, 3189–3206.
1101 Retrieved from <https://doi.org/10.1002/qj.3172>
- 1102 Geer, A. J., & Bauer, P. (2011). Observation errors in all-sky data assimilation.
1103 *Quart. J. Roy. Meteorol. Soc.*, *137*, 2024–2037. Retrieved from <https://doi>
1104 [.org/10.1002/qj.830](https://doi.org/10.1002/qj.830)
- 1105 Geer, A. J., Bauer, P., & O’Dell, C. W. (2009). A revised cloud overlap scheme for
1106 fast microwave radiative transfer. *J. App. Meteor. Clim.*, *48*, 2257–2270. Re-
1107 trieved from <https://doi.org/10.1175/2009JAMC2170.1>
- 1108 Geer, A. J., Lonitz, K., Duncan, D., & Bormann, N. (2022). *Improved surface treat-*
1109 *ment for all-sky microwave observations* (Tech. Memo. No. 894). Reading, UK:
1110 ECMWF. Retrieved from <https://doi.org/10.21957/zi7q6hau>
- 1111 Geer, A. J., Lonitz, K., Weston, P., Kazumori, M., Okamoto, K., Zhu, Y., ...
1112 Schraff, C. (2018). All-sky satellite data assimilation at operational weather
1113 forecasting centres. *Quart. J. Roy. Meteorol. Soc.*, *144*, 1191–1217. doi:
1114 [10.1002/qj.3202](https://doi.org/10.1002/qj.3202)
- 1115 Gettelman, A., Geer, A. J., Forbes, R. M., Carmichael, G. R., Feingold, G., Pos-
1116 selt, D. J., ... Zuidema, P. (2022). The future of earth system prediction:
1117 Advances in model-data fusion. *Science Advances*, *8*(14), eabn3488.
- 1118 Ghahramani, Z. (2015). Probabilistic machine learning and artificial intelligence.
1119 *Nature*, *521*(7553), 452–459.
- 1120 Glorot, X., & Bengio, Y. (2010). Understanding the difficulty of training deep feed-
1121 forward neural networks. In *Proceedings of the thirteenth international confer-*
1122 *ence on artificial intelligence and statistics* (pp. 249–256).
- 1123 Golmant, N., Vemuri, N., Yao, Z., Feinberg, V., Gholami, A., Rothauge, K., ...
1124 Gonzalez, J. (2018). On the computational inefficiency of large batch sizes for
1125 stochastic gradient descent. *arXiv preprint arXiv:1811.12941*.

- 1126 Good, S., Fiedler, E., Mao, C., Martin, M. J., Maycock, A., Reid, R., ... others
 1127 (2020). The current configuration of the OSTIA system for operational pro-
 1128 duction of foundation sea surface temperature and ice concentration analyses.
 1129 *Remote Sensing*, 12(4), 720. Retrieved from [https://doi.org/10.3390/
 1130 rs12040720](https://doi.org/10.3390/rs12040720)
- 1131 Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J.,
 1132 ... Thépaut, J. (2020). The ERA5 global reanalysis. *Quarterly Journal of
 1133 the Royal Meteorological Society*. Retrieved from [https://doi.org/10.1002/
 1134 qj.3803](https://doi.org/10.1002/qj.3803)
- 1135 Hewison, T. J., & English, S. J. (1999). Airborne retrievals of snow and ice surface
 1136 emissivity at millimeter wavelengths. *IEEE Trans. Geosci. Rem. Sens.*, 37(4),
 1137 1871–1879.
- 1138 Hirahara, Y., Rosnay, P. d., & Arduini, G. (2020). Evaluation of a microwave
 1139 emissivity module for snow covered area with CMEM in the ECMWF inte-
 1140 grated forecasting system. *Remote Sensing*, 12(18), 2946. Retrieved from
 1141 <https://doi.org/10.3390/rs12182946>
- 1142 Hortal, M., & Simmons, A. (1991). Use of reduced Gaussian grids in spectral mod-
 1143 els. *Monthly Weather Review*, 119(4), 1057–1074. Retrieved from [https://doi
 1144 .org/10.1175/1520-0493\(1991\)119<1057:UORGGI>2.0.CO;2](https://doi.org/10.1175/1520-0493(1991)119<1057:UORGGI>2.0.CO;2)
- 1145 Hsieh, W. W., & Tang, B. (1998). Applying neural network models to prediction
 1146 and data analysis in meteorology and oceanography. *Bulletin of the American
 1147 Meteorological Society*, 79(9), 1855–1870.
- 1148 JAXA. (2023). *Global Change Observation Mission – Water Advanced Microwave
 1149 Scanning Radiometer 2 level 1B* [dataset]. Retrieved from [https://doi.org/
 1150 10.57746/E0.01gs73ans548qghaknzdjyxd2h](https://doi.org/10.57746/E0.01gs73ans548qghaknzdjyxd2h)
- 1151 Kang, E.-J., Sohn, B.-J., Tonboe, R. T., Noh, Y.-C., Kwon, I.-H., Kim, S.-W., ...
 1152 Liu, C. (2023). Explicitly determined sea ice emissivity and emission temper-
 1153 ature over the arctic for surface-sensitive microwave channels. *Quart. J. Roy.
 1154 Meteorol. Soc.*. Retrieved from <https://doi.org/10.1002/qj.4492>
- 1155 Kazumori, M., & English, S. J. (2015). Use of the ocean surface wind direction sig-
 1156 nal in microwave radiance assimilation. *Quarterly Journal of the Royal Mete-
 1157 orological Society*, 141(689), 1354–1375. Retrieved from [https://doi.org/10
 1158 .1002/qj.2445](https://doi.org/10.1002/qj.2445)
- 1159 Kazumori, M., Geer, A. J., & English, S. J. (2016). Effects of all-sky assimilation
 1160 of GCOM-W/AMSR2 radiances in the ECMWF numerical weather predic-
 1161 tion system. *Quart. J. Roy. Meteorol. Soc.*, 142, 721–737. Retrieved from
 1162 <https://doi.org/10.1002/qj.2669>
- 1163 Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv
 1164 preprint arXiv:1412.6980*.
- 1165 Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M.,
 1166 Pritzel, A., ... Battaglia, P. (2022). GraphCast: Learning skillful medium-
 1167 range global weather forecasting. *arXiv preprint arXiv:2212.12794*.
- 1168 Lee, S.-M., Sohn, B.-J., & Kim, S.-J. (2017). Differentiating between first-year
 1169 and multiyear sea ice in the arctic using microwave-retrieved ice emissiv-
 1170 ities. *J. Geophys. Res.: Atmos.*, 122(10), 5097–5112. Retrieved from
 1171 <https://doi.org/10.1002/2016JD026275>
- 1172 Leinonen, J., Grazioli, J., & Berne, A. (2021). Reconstruction of the mass and
 1173 geometry of snowfall particles from multi-angle snowflake camera (MASC)
 1174 images. *Atmos. Meas. Tech.*, 14(10), 6851–6866.
- 1175 Okuyama, A., & Imaoka, K. (2015). Intercalibration of advanced microwave
 1176 scanning radiometer-2 (AMSR2) brightness temperature. *IEEE Transac-
 1177 tions on Geoscience and Remote Sensing*, 53(8), 4568–4577. Retrieved from
 1178 <https://doi.org/10.1109/TGRS.2015.2402204>
- 1179 OSI-SAF, E. (2016). *Global sea ice concentration (SSMIS)* (Tech. Rep. No. OSI-401-
 1180 b). Retrieved from https://doi.org/10.15770/EUM_SAF_OSI_NRT_2004

- 1181 Peubey, C., & McNally, A. P. (2009). Characterization of the impact of geostation-
 1182 ary clear-sky radiances on wind analyses in a 4D-Var context. *Quart. J. Roy.*
 1183 *Meteorol. Soc.*, *135*, 1863 – 1876.
- 1184 Picard, G., Sandells, M., & Löwe, H. (2018). SMRT: An active–passive microwave
 1185 radiative transfer model for snow with multiple microstructure and scattering
 1186 formulations (v1. 0). *Geoscientific Model Development*, *11*(7), 2763–2788.
- 1187 Python Software Foundation. (2021). *Python* [code]. Retrieved from [https://](https://github.com/python/cpython/releases/tag/v3.8.8rc1)
 1188 github.com/python/cpython/releases/tag/v3.8.8rc1
- 1189 Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2017). Physics informed deep learn-
 1190 ing (part i): Data-driven solutions of nonlinear partial differential equations.
 1191 *arXiv preprint arXiv:1711.10561*.
- 1192 Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N.,
 1193 et al. (2019). Deep learning and process understanding for data-driven Earth
 1194 system science. *Nature*, *566*(7743), 195–204.
- 1195 Rodgers, C. D. (2000). *Inverse methods for atmospheric sounding: Theory and prac-*
 1196 *tice* (1st ed.). Singapore: World Scientific.
- 1197 Sandells, M., Rutter, N., Wivell, K., Essery, R., Fox, S., Harlow, C., ... Toose, P.
 1198 (2023). Simulation of arctic snow microwave emission in surface-sensitive
 1199 atmosphere channels. *EGUsphere*, *2023*, 1–28. Retrieved from [https://](https://egusphere.copernicus.org/preprints/2023/egusphere-2023-696/)
 1200 egusphere.copernicus.org/preprints/2023/egusphere-2023-696/ doi:
 1201 [10.5194/egusphere-2023-696](https://doi.org/10.5194/egusphere-2023-696)
- 1202 Sandven, S., Spreen, G., Heygster, G., Girard-Ardhuin, F., Farrell, S. L., Dierking,
 1203 W., & Allard, R. A. (2023). Sea ice remote sensing—recent developments in
 1204 methods and climate data sets. *Surveys in Geophysics*, 1–37. Retrieved from
 1205 <https://doi.org/10.1007/s10712-023-09781-0>
- 1206 Scarlat, R. C., Spreen, G., Heygster, G., Huntemann, M., Pațilea, C., Pedersen,
 1207 L. T., & Saldo, R. (2020). Sea ice and atmospheric parameter retrieval from
 1208 satellite microwave radiometers: Synergy of amsr2 and smos compared with
 1209 the cimr candidate mission. *Journal of Geophysical Research: Oceans*, *125*(3),
 1210 e2019JC015749. Retrieved from <https://doi.org/10.1029/2019JC015749>
- 1211 Schneider, T., Lan, S., Stuart, A., & Teixeira, J. (2017). Earth system modeling
 1212 2.0: A blueprint for models that learn from observations and targeted high-
 1213 resolution simulations. *Geophys. Res. Lett.*, *44*(24), 12–396. Retrieved from
 1214 <https://doi.org/10.1002/2017GL076101>
- 1215 Smith, S. L., Kindermans, P.-J., Ying, C., & Le, Q. V. (2017). Don't decay the
 1216 learning rate, increase the batch size. *arXiv preprint arXiv:1711.00489*.
- 1217 Spreen, G., Kaleschke, L., & Heygster, G. (2008). Sea ice remote sensing using
 1218 AMSR-E 89-GHz channels. *J. Geophys. Res.: Oceans*, *113*(C2), C02S03. Re-
 1219 trieved from <https://doi.org/10.1029/2005JC003384>
- 1220 TensorFlow. (2021). *Tensorflow version 2.8.0* [code]. Retrieved from [https://doi](https://doi.org/10.5281/zenodo.5799851)
 1221 [.org/10.5281/zenodo.5799851](https://doi.org/10.5281/zenodo.5799851)
- 1222 Timmermann, R., Goosse, H., Madec, G., Fichefet, T., Ethe, C., & Duliere, V.
 1223 (2005). On the representation of high latitude processes in the ORCA-LIM
 1224 global coupled sea ice–ocean model. *Ocean Modelling*, *8*(1-2), 175–201. Re-
 1225 trieved from <https://doi.org/10.1016/j.ocemod.2003.12.009>
- 1226 Zuo, H., Balmaseda, M. A., Tietsche, S., Mogensen, K., & Mayer, M. (2019). The
 1227 ECMWF operational ensemble reanalysis–analysis system for ocean and sea
 1228 ice: a description of the system and assessment. *Ocean science*, *15*(3), 779–
 1229 808. Retrieved from <https://doi.org/10.5194/os-15-779-2019>