Parameter sensitivity analysis of a sea ice melt pond parametrisation and its emulation using neural networks

Simon Driscoll [a],[*], Alberto Carrassi [b],[a], Julien Brajard [c], Laurent Bertino [c], Marc Bocquet [d], Einar Örn Ólason [c]

[a] *University of Reading, Department of Meteorology, Earley Gate, PO Box 243, Reading, RG6 6BB, United Kingdom*
[b] *Department of Physics and Astronomy "Augusto Righi", University of Bologna, Viale Carlo Berti Pichat, 6/2, Bologna, 40127, Italy*
[c] *Nansen Environmental and Remote Sensing Center, Jahnebakken 3, Bergen, N-5007, Norway*
[d] *CEREA, École des Ponts and EDF R&D, Île-de-France, France*

A B S T R A C T

Accurate simulation of sea ice is critical for predictions of future Arctic sea ice loss, looming climate change impacts, and more. A key feature in Arctic sea ice is the formation of melt ponds. Each year melt ponds develop on the surface of the ice and primarily via affecting the albedo, they have an enormous effect on the energy budget and climate of the Arctic. As melt ponds are subgrid scale and their evolution occurs due to a number of competing, poorly understood factors, their representation in models is parametrised.

Sobol sensitivity analysis, a form of variance based global sensitivity analysis is performed on an advanced melt pond parametrisation (MPP), in Icepack, a state-of-the-art thermodynamic column sea ice model. Results show that the model is very sensitive to changing its uncertain MPP parameter values, and that these have varying influences over model predictions both spatially and temporally. Such extreme sensitivity to parameters makes MPPs a potential source of prediction error in sea-ice models, given that the (often many) parameters in MPPs are usually poorly known.

Machine learning (ML) techniques have shown great potential in learning and replacing subgrid scale processes in models. Given the complexity of melt pond physics and the need for accurate parameter values in MPPs, we propose an alternative data-driven MPPs that would prioritise the accuracy of albedo predictions. In particular, we constructed MPPs based either on linear regression or on nonlinear neural networks, and investigate if they could substitute the original physics-based MPP in Icepack.

Our results shown that linear regression are insufficient as emulators, whilst neural networks can learn and emulate the MPP in Icepack very reliably. Icepack with the MPPs based on neural networks only slightly deviates from the original Icepack and overall offers the same long term model behaviour.

We then searched for the smallest possible emulator that achieves good performance by performing features selection based on mutual information. Results indicates that a smaller model, based only on a portion of the full set of input variables needed by the physical MPP, is also sufficient to approximate and replace the physical MPP. This smaller emulators are not only computationally faster but also easier to interpret on a physical ground.

Several and diverse challenges still exist, yet this study is an encouraging first step, prior to using real data, towards the adoption of data-driven MPPs in sea-ice models.

1. Introduction

Arctic sea ice plays an essential role in global ocean circulation and in regulating Earth's climate and weather [e.g., 1–3]. Growing from a minimum in September, until the Arctic ice pack reaches its maximum extent in March [4], sea ice is a good thermal insulator, and also becomes a platform for an even greater insulator, snow. It acts as a lid on the surface of the polar oceans, and governs transfer of heat and water vapour between a comparatively warmer ocean and cold atmosphere [5].

Of fundamental importance in its own right, sea ice is also one of the Earth system components most sensitive to climate change [6], and reductions in Arctic ice cover have been considerable over the satellite record since 1979 [7,8]. Climate models project that Arctic sea ice loss

---

will continue with the possibility of ice-free summers occurring before or by the end of the 21st century [e.g., 9,10]. Losing high-albedo sea ice causes exposure of the dark ocean surface, more sunlight is absorbed, which in turn enhances surface warming [11]. This feedback plays a major role in the Arctic amplification of global warming [12].

A key question for the climate system, with important biological [13], economic [14] and geopolitical [15] implications, is understanding when sea ice may entirely vanish from Arctic summers. Climate models have traditionally significantly underestimated the observed trend in Arctic sea ice decline [9], and have also failed to capture physical changes in other warming periods, such as the elevated Arctic temperatures during the last interglacial (LIG) period [16].

1.1. Sea-ice melt ponds and their role in Arctic climate

Whilst reductions in sea ice and leads opening between the ice alter the albedo and thus the climate of the Arctic by exposing the ocean surface, another profound impact to the Arctic's climate occurs on the ice itself, through the formation of *melt ponds*. Melt ponds develop over the sea ice during the melting season from the accumulation of melt water from ice and snow, and may cover up to 50% of the Arctic sea ice surface itself [17]. Whilst the albedo of bare sea ice can be up to 0.8, melt ponds may have an albedo as low as 0.15 [17], and their evolution on the Arctic sea ice in summer is one of the main factors affecting sea ice albedo and hence the polar climate system [18].

The seasonal evolution of Arctic sea ice albedo undergoes five distinct phases: dry snow, melting snow, pond formation, pond evolution, and freeze-up during Autumn [19]. In April (with much of dry snow) the albedo tends to be high, and generally uniform, whilst as summer progresses it lowers and becomes spatially inhomogeneous. Accounting for these phases in numerical models of the sea ice requires a parametrisation of melt pond effects. Melt ponds shape, size, and coverage are determined by a number of factors that include ice surface topography and the total water amount available [20]. Melt ponds exist at a higher percentage over thinner younger ice (at the end of the Artic summer) and the impact of melt ponds will increase, as the area of young ice is expected to increase due to climate change. The role and the impact of melt ponds on the physics and dynamics of the sea ice is also evidenced by numerical experiments with sea ice models. The latter exhibit a large sensitivity, particularly in terms of ice thickness to melt pond parametrisations [e.g., 21]. Flocco et al. [22,23] have shown that models lacking a melt pond parametrisation (MPP) can overestimate the summertime sea ice thickness by up to 40%.

Other results indicate the importance of melt pond processes and point to the need for their accurate simulation. With improved model physics and a sophisticated melt pond parametrisation, Guarino et al. [16] showed that the Hadley Centre HadGEM3 climate model hindcasted a complete loss of Arctic summer sea ice in LIG period simulations. They concluded that loss of Arctic sea ice in their simulations is largely due to increased net short-wave radiation, and not to a slowdown in the Atlantic Meridional Overturning Circulation (AMOC) as previously hypothesised. Their results suggested that local thermodynamic and melt pond formation processes play a key role in capturing the Arctic sea ice loss. Therefore, skilful simulations of future Arctic sea ice extension calls for accurate simulations of MPPs.

1.2. Using machine learning for parametrisation of physical processes in climate models

Machine learning (ML) can accurately and efficiently recognise complex patterns and emulate nonlinear dynamics [24]. Interest in using ML to emulate parametrised processes in climate simulations has grown considerably in the last decade [e.g., 25–27], largely (but not only) since emulators may offer improvements in computation speed as compared to traditional physics-based parametrisations [28]. When

trained against high resolution models that explicitly resolves the processes intended to be parametrised, ML-based parametrisations showed potential to improve model skills [e.g.,29–31, for the case of moist convection in global climate models]. This "perfect model scenario" whereby training is based on high-resolution models, is functional to the chronic lack of sufficient spatio-temporal dense accurate dataset required by proper training. This issue can be combated by combining data assimilation methods with ML, in such a way that the former handles efficiently the sparse and noise data and provides the analysed fields to train the ML [32,33].

1.3. Sensitivity of melt ponds parametrisations and their discovery using machine learning

This work has two connected goals. In the first part (Section 3) we study in detail the parametric sensitivity of the Icepack model, a state-of-the-art model of the thermodynamics processes in the sea ice [34], to its physics-based MPP. We adopt a rigorous approach to quantify this sensitivity, using the Sobol sensitivity analysis method [35–37]. As mentioned earlier inappropriate MPPs in sea ice model can cause overestimate of the summertime sea ice thickness by up to 40% [22,23]. Understanding the sources of uncertainty in MPPs and reducing them are both key to accurate climate simulations.

We shall see in our study that Icepack shows an intricate, generally high, dependency on the parameters of its MPPs, leading to large impacts in the predictions of crucial sea ice quantities. Understanding the source of uncertainty, narrowing it, and fine tuning the parameters is paramount. However, this is usually difficult to achieve as the true physics and geometry of melt ponds are very complex, and thus the MPP are usually over-parametrised.

To bypass the undesirable sensitivity of the physical MPP, and the limited accuracy against the costs of tuning the parameters, in the second part of this study (Section 4), we investigate the capability of modern ML to emulate the MPPs with the fewest possible inputs. We shall show that neural networks (NN) can successfully learn the functional relationship between inputs and outputs in the MPP, and that this ML-based MPP can substitute Icepack into the full sea-ice model providing accurate and stable simulations. Finally, in Section 4.4 we search for the smallest possible ML-based MPP by performing feature selection based on mutual information tools.

We work under the "perfect model" scenario, whereby the training relies on the output of the MPP intended to be emulated, thus a first step towards using real data that will be considered in the follow-up study. we will allude to this point in our conclusions and discussion on future directions in Section 5.

2. The sea-ice model *Icepack* and its melt pond parametrisations

The Los Alamos Sea Ice Model (CICE) was developed to create an efficient sea ice component for a fully coupled atmosphere-ice-ocean-land global climate model [38]. CICE has different interactive components: an ice dynamics model, a transport model, and the sub module *Icepack* that simulates all vertical processes, mainly of thermodynamics nature, in CICE [34]. Icepack module can be used as a stand-alone model and this is the way we shall implement it in this work. A schematic of Icepack together with the illustration of the important vertical processes are given in Fig. 1.

2.1. Melt pond parametrisations in Icepack

Formation of melt ponds on ice is simulated in Icepack considering pond area and depth. When the top of the pond refreezes and snowfall on top of the refrozen lid blocks solar radiation, the "effective pond area" that is used for the radiation calculations can decrease while the pond volume remains constant. The effective pond area is therefore what influences the sea ice albedo. Three different schemes are
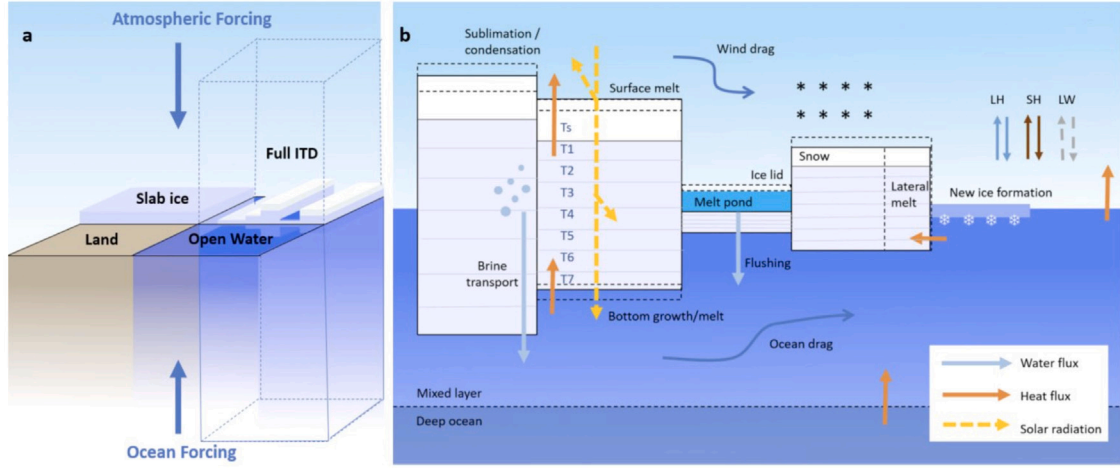
Fig. 1. Schematic overview of the Icepack model. Panel a: the columnar model set-up, with 4 tiles representing land, open water, slab ice (without snow) and a full ice-thickness distribution (ITD). Panel b: the most important vertical processes represented in Icepack when using the ITD. Courtesy of Hofsteenge and Ganzeveld [39]. For input and output to the MPP model, see Table 3.

available in Icepack to explicitly model melt ponds: (1) *CESM* [20], *topo* [22], and *level-ice* [40].

The first MPP is made specifically for the Community Earth System Model (CESM). In the CESM scheme, ponds can grow when rain or snow and ice melt water is added and shrink through refreezing. The melt pond processes are described empirically in this scheme and pond depth and area are linearly related.

The *topo* scheme is said topographic in that it simulates the concept that melt water collects on the lowest parts of the ice [41]. Since Icepack does not explicitly model ice topography, the ice thickness distribution is split into a surface height and basal depth distribution relative to sea level. Melt water is collected on the ice of the lowest surface height. In this scheme pond water can refreeze (affecting the effective pond area) and drain vertically when the sea ice becomes permeable.

The last pond scheme, the *level-ice* [40], builds upon *topo* but in additions it also accounts for gravity effects in a way where ponds can only form on the level (undeformed) ice areas per ice category. In the level-ice scheme, melt pond water can also refreeze and drain to the ocean, depending on the permeability of the ice.

Being the most advanced formulation in Icepack [34], in this study we focus on the level-ice MPP only, and shall hereafter refer to it as the MPP, without the need for further specification.

2.2. Icepack's melt pond parameters

Icepack has a total of seven melt pond parameters [34], although not all of them are used in the level-ice MPP. We know that one parameter is used only in the "topo" scheme, whilst the parameter *dpscale*, the "alter e-folding time scale for flushing pond refreezing parametrisation", is not used when the mushy thermodynamics option [that treats the sea ice as a mushy layer in which a salinity profile is allowed to evolve, as in 42] is used. See Hunke et al. [34] for more details.

Table 1 lists all melt pond parameters in the level-ice MPP together with uncertainty ranges as given by Urrego-Blanco et al. [43]. As can be seen, given these uncertainties, understanding the impact of these on the model is of crucial importance. It motivates us to employ a Sobol sensitivity analysis (SSA) on these parameters and explore the sensitivity of the sea ice variables in the Icepack model to these parameter values (see Fig. 2).

3. Sobol sensitivity analysis of melt pond parameters

3.1. An outline of sobol sensitivity analysis

Sensitivity analysis may be defined as "the study of how uncertainty in the output of a model (numerical or otherwise) can be apportioned to different sources of uncertainty in the model input" [37]. Here we choose to perform a Sobol Sensitivity Analysis [SSA, 35,36] on the Icepack MPP parameters. A global, variance-based method, SSA attributes variance in model output to individual or multiple parameters simultaneously, by taking into account the interactions between them [36,44]. In studies of hydrological models, for instance, SSA provided sensitivity indices with best accuracy and robustness, especially with models with strong parameter interactions [e.g., 45].

Another possible option could be using Shapley values that quantify the importance of each input variable to a predictive model by distributing the prediction value among features, offering a measure of each feature's impact on the model prediction [46]. Shapley values play an important role in sensitivity studies especially when the inputs are correlated. Since the melt pond parameters are almost completely independent, and in order to provide consistency between our study and that of Urrego-Blanco et al. [43] using CICE, we choose SSA to investigate the sensitivity of Icepack to its melt pond parameters.

For a detailed description and explanation of the SSA, see e.g., Sobol [35,36], Saltelli et al. [37]. We briefly outline it in the following. Let us consider the model:

$$\mathbf{Y} = f(\lambda) = f(\lambda_1, \lambda_2, \dots, \lambda_d), \tag{1}$$

with $\mathbf{Y} \in \mathbb{R}^m$ the model state vector, $\lambda_i \in \mathbb{R}$, for $i = 1, \dots, d$, the model parameters, thought to be independent random variables, described by known distributions that reflect uncertain knowledge on the system. Let $\mathscr{D} = \{1, \dots, d\}$. With $\Psi_i \subseteq \mathscr{D}$ being the set containing the ith parameter, we define $\tilde{\Psi}_i := \mathscr{D} \setminus i$, as its complement set in $\mathscr{D}$, such that:

$$\Psi_i \cup \tilde{\Psi}_i = \mathscr{D}, \qquad \Psi_i \cap \tilde{\Psi}_i = \emptyset. \tag{2}$$

Assume that we know the true parameter values, where $\lambda_i = \lambda_i^*$, for $i = 1, \dots, d$, and we would like to understand the impact of each parameter λ_i upon the variance of the output $\mathbf{Y}$. We obtain the first order sensitivity index for parameter λ_i:

$$S_i = \frac{\mathbb{V}_{\Psi_i}(\mathbb{E}_{\tilde{\Psi}_i}(\mathbf{Y}|\lambda_i))}{\mathbb{V}(\mathbf{Y})}, \tag{3}$$

where $\mathbb{E}$ and $\mathbb{V}$ are the expectation and variance operator, respectively.

Table 1

Parameter symbols, descriptions and statistical distributions as given by Urrego-Blanco et al. [43]. The distributions we use for SSA are visualised in Fig. 2.

Parameters perturbed for SSA of the *level-ice* MPP in Icepack					
Parameter symbol	Parameter description	Prior probability distribution	Min value	Max value	Default value
hs0	Snow depth of transition to bare ice (m)	Triangular	0	0.1	0.03
hs1	Snow depth of transition to pond ice (m)	Triangular	0	0.1	0.03
rfracmin	Minimum retained fraction of meltwater	Triangular	0	1	0.15
rfracmax	Maximum retained fraction of meltwater	Triangular	0	1	0.85
pndaspect	Ratio of pond depth to pond fraction	Uniform	0.4	1.2	0.8

Fig. 2. Probability distribution functions we sample from using the SALib library for sensitivity analysis, for the 5 parameters invoked in the level-ice melt pond parametrisation with the mushy thermodynamics scheme (triangular and uniform). Red lines indicate default parameter values in the model. So that the minimum retained fraction of meltwater does not exceed the maximum, we bound the minimum to $[0, 0.5]$ and the maximum to $[0.5, 1]$.

First order sensitivity indices characterise the fraction of the variance due to the parameter λ_i only, i.e. without interaction with others, and from the normalised law of total variance, we see that $S_i \leq 1$.

Total sensitivity indices have been introduced in Homma and Saltelli [47]: they account for all the contributions to the variation in output due to parameter λ_i plus all its interactions with the other parameters. Total indices can be given as

$$ST_i = \frac{\mathbb{E}_{\tilde{\Psi}_i}(\mathbb{V}_{\lambda_i}(\mathbf{Y}|\tilde{\Psi}_i))}{\mathbb{V}(\mathbf{Y})}, \qquad (4)$$

where $\tilde{\Psi}_i$ denotes all uncertain parameters except Ψ_i, and we have therefore $0 \leq S_i \leq ST_i \leq 1$, implying by definition that the total index will always bound the normal index from above.

3.2. SSA of Icepack parameters

To conduct Sobol sensitivity analysis on the Icepack model we use the Sensitivity Analysis Library in Python [SALib, 48]. Parameter uncertainty estimates come from Urrego-Blanco et al. [43], adapted to be suitable for SALib. Whilst the logit-normal probability distribution function as given by Urrego-Blanco et al. [43] for certain parameters is not available in SALib, our aim is to investigate sensitivity to parameter values, rather than reducing parameter uncertainty, thus we approximate this with a triangular distribution. These input parameters (and their uncertainty range) for Icepack's melt pond parameters are given in Table 1. Using SALib, a global sample of this parameter space is taken using Sobol–Saltelli sampling to achieve a uniform coverage of the space [36,49].

Icepack is a 1D thermodynamic column model that runs at a single geographic location and it requires forcing data to provide outputs of sea ice area fraction, average ice thickness, and so on. The data we choose to force Icepack with comes from the NCEP Climate Forecast System Version 2 (CFSv2) [50]. CFSv2 is a fully coupled model representing the interaction between the Earth's atmosphere, oceans, land and sea ice. Initialised four times per day (00:00, 06:00, 12:00, and 18:00 UTC), it provides us with more than a decade (2011–2022 period) of hourly data over many Arctic locations with which to force Icepack.

The Saltelli sampler of a provided parameter space in the SALib library generates $N \times (2d + 2)$ samples, where N is a chosen integer, required to be a power of 2 for equal subsampling and d is the number of parameter inputs. A low value of N would imply insufficient sampling of the parameter space. In this study to guarantee a sufficient sampling

Table 2

The locations used for our SSA of the Icepack model represent 12 physically distinct areas of the Arctic.

Locations used for Sobol sensitivity analysis			
Location	Region	Latitude	Longitude
1	Beaufort Sea	72° N	137° W
2	Chukchi Sea	69° N	172° W
3	East Siberian Sea	72° N	163° E
4	Laptev Sea	76.3° N	125.6° E
5	Kara Sea	77° N	77° E
6	Barents Sea	75° N	40° E
7	Greenland Sea	76° N	8° W
8	Fram Strait	78° N	0° E
9	Lincoln Sea	82° N	58° W
10	Baffin Bay	74° N	68° W
11	Hudson Bay	60° N	86° W
12	Arctic Ocean	90° N	0° E

of our parameter space, we use $N = 128$, with $d = 5$ parameters (cf Table 1), yielding 1536 simulations at each location. Icepack is then run with these combinations of perturbed parameter inputs, and an SSA is performed on the subsequent model output to compute sensitivity indices.

We choose to focus on 12 physically distinct locations in the Arctic (Table 2) so as to characterise the response in various Arctic regions. These 12 locations with 1536 simulations over the period of 2011–2022 represent approximately 200,000 years of Icepack model data with which we perform our SSA of the Icepack model to its melt pond parameters.

Arguably the two most fundamental variables to consider when modelling sea ice are the sea ice thickness and area fraction, effectively representing the total volume of sea ice in a given region. Fig. 3 shows the results of SSA: the first order and total order Sobol indices for these key sea ice variables. Therefore in Fig. 3 we are analysing the sensitivity of the Icepack model's variables related to sea ice volume with respect to its melt pond parameters.

Fig. 3 displays the monthly averaged first and total order sensitivity values averaged over the whole simulation period (2011–2022). We show the sensitivity to four over the five parameters in the melt pond parametrisation of Icepack (cf Table 1): the parameter "snow depth of transition to bare ice" ($hs0$ in Table 1) has no effect in the current scheme, and is thus excluded from the analysis. The Sobol indices are averaged across the 12 locations to produce a single average
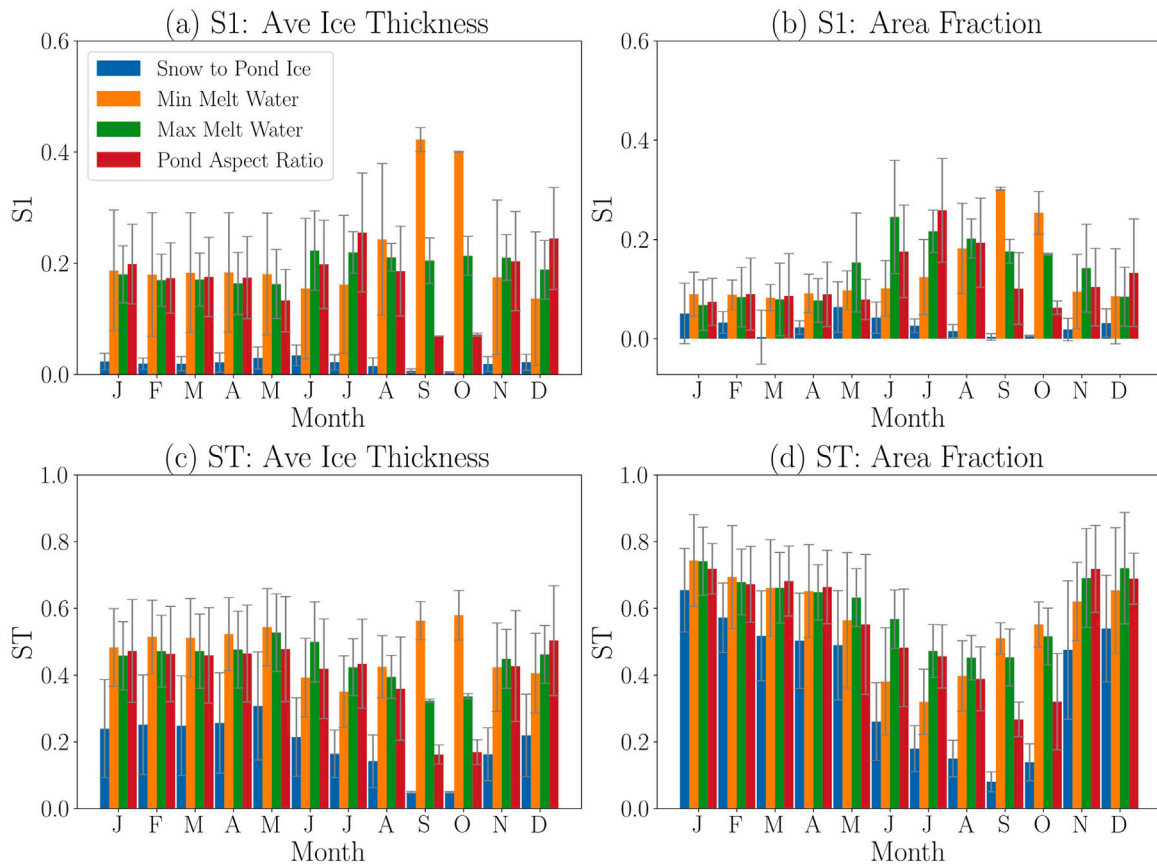
Fig. 3. Monthly first and total order Sobol indices relative to the 5 parameters in Table 1, averaged over the period 2012–2022, and averaged over all the 12 locations in Table 2. Whisker plots on the monthly averages show the standard deviation of total order Sobol indices over 12 locations.

first and total Sobol indices, and the vertical grey lines indicate the standard deviation over the 12 geographical locations. Thus Fig. 3 quantifies the sensitivity of the simulated ice thickness and are fraction to four parameters, expressing also its seasonality and the geographical variability.

Fig. 3 shows that there exists a difference in the sensitivity of the model to each parameter - i.e. not all parameters are equal in their influence. The parameter *hs1* (marking the snow depth of transition to pond ice, in blue) has the least influence across all Sobol indices and variables. The other parameters are associated with the minimum and maximum amount of available melt water to be added to melt ponds, and the last governs the pond aspect ratio (i.e. the area/height ratio of a pond of given volume). For a given volume of pond, altering the aspect ratio to make a smaller depth but bigger surface area pond would cause more absorption of incoming solar radiation, and therefore further melting. By contrast, a deep, narrow pond of the same volume would absorb incoming solar radiation, but would leave larger areas for exposed bare ice to reflect away the incoming solar radiation. Our SSA supports the fact that the amount of melt water added to a melt pond and the surface area of the pond are fundamental and obviously influential features when modelling melt ponds.

The qualitative differences in first order and total order indices between variables (e.g. Fig. 3a-c, and Fig. 3b-d) indicate that the Icepack model's sensitivity to all of its melt pond parameters (as measured by ST) is more complicated than the effects of varying parameters individually (measured by S1). By construction the total order indices will always be larger than first order indices, as the total contains the sensitivity to that parameter alone, and higher order interactions. For instance, the parameter with the smallest S1 in Fig. 3a and (b) (*hs1*) has a much larger influence on the models output when considering interactions with other parameters (compare S1 and ST of this parameter in

Fig. 3a and Fig. 3c). The same holding true for all parameters suggests that individually fine-tuning parameters in a sea ice model may not necessarily be the best approach, given this would neglect and ignore higher order effects. Whilst the first order effects are important, they neglect higher order interactions and the "total contribution" of each parameter towards the model output. Thus we will focus the remainder of this analysis predominantly on the total order Sobol indices.

The sensitivity of Icepack outputs to all of these parameters varies according to the time of year. We observe a decreasing ST of the Icepack model's area fraction to its melt pond parameters (Fig. 3d) as it enters the summer. This may be due to large (and increasing) areas in the Arctic having ice free periods as the summer approaches, thus limiting the influence of the melt pond parameters: there will be less ice to melt. Nonetheless, not all areas in the Arctic will fully melt to create an ice free area, and thus we can see that even throughout the summer the average values for any index and variable are non-zero. This rationale is further supported by the large standard deviation over the locations (grey vertical lines), suggesting that while locations at lower-latitude might be free of ice, higher-latitudes ones still present ice.

To further unveil the geographical dependence of the model sensitivity to parameters, a manifestation of diverse and differently dominant physical process, we look at the model runs at two exemplar locations. The first is the Southernmost location: the Hudson Bay (#11 in Table 2): it freeze over completely in winter but thaw in the summer. The second is the Northernmost location, the Arctic Ocean (#12 in Table 2). Fig. 4 shows the time series of the monthly average sea ice thickness for each perturbed parameter member in the ensemble used for SSA. The difference between the two locations is huge. We see that not only does altering melt pond parameter values have substantially different effects depending on the location, but also that, for the Arctic Ocean location (panel a) the differences in predicted sea ice thickness can differ by ≈ 2 − 3 metres.
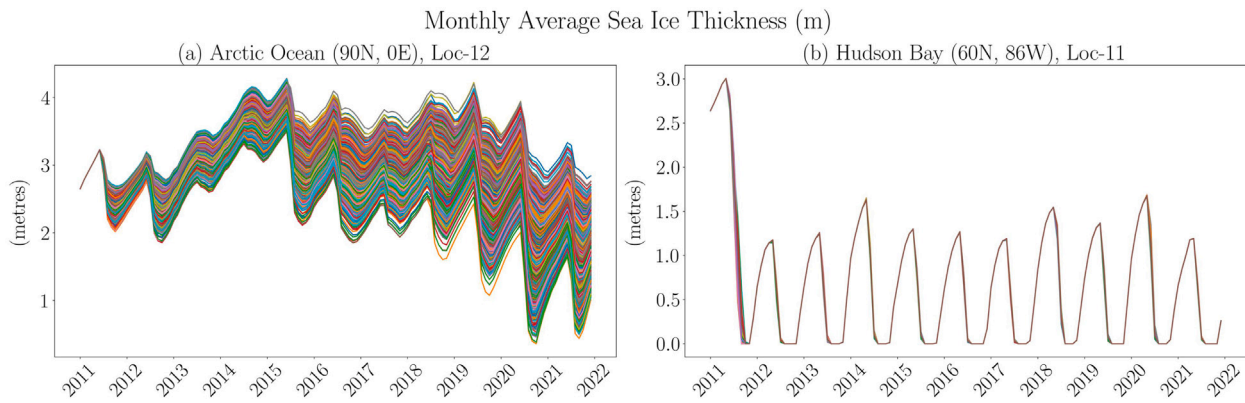
Monthly Average Sea Ice Thickness (m)

Fig. 4. Time series of monthly averaged sea ice thickness (metres) over the southernmost and northernmost locations chosen for SSA across all sampled parameter values. Each line represents an individual simulation in the 1536 member ensemble at each location.

Overall our results reveal the Icepack model is highly sensitive to its melt pond parameters and that this sensitivity is both temporally and spatially varying. This implies that parameters tested on various locations, and fine tuned to these locations, will generally not be optimal elsewhere. Note that the physical processes governing melt pond evolution in the marginal ice zone are thought to be considerably different to processes for example in the Central Arctic. A parametrisation that could be tested and fine tuned for stability at one location (where the model output may be much less sensitive to these parameter values). At the same time, these small fine tunings for stability may have a large influence on model output at another location (i.e. a location with differing ice properties and where the parameter values could play a larger role on the model output). However, we can see here that small changes in parameter values that appear innocuous at one location could have enormous implications for sea ice thickness predictions at another location. That a small perturbation in these parameters can produce up to a few metres difference predicted sea ice thickness in the Arctic Ocean over approximately a decade is especially important for long term and climate change predictions.

Climate model parameterisations, as the MPP in Icepack, are in general geographically and temporally invariant, "one-size-fits-all" approach. Ideally parametrisation should incorporate a spatio-temporal dependency and the complexity of processes ongoing in the Arctic that lead to the formation of melt ponds. What may seem relatively inconsequential uncertainty in one area, can lead to dramatically different sea ice predictions in another area. Can data-driven parametrisations emulate accurately physical parametrisation of the melt ponds in different geographical locations and seasons? Can they offer an alternative, computationally cheaper solution? Are they less sensitive to parameters' specifications? Those are the questions addressed in the remaining of this paper.

4. Emulation of a melt pond parametrisation

We seek to understand if an emulator can learn and replace a physics-based MPP. We shall work under a "perfect data" assumption, in which our emulator is trained against exact (and complete) output of the MPP that it intends to replicate. While this assumption will be relaxed in follow-up studies where we will train our emulator against real data (see Section 5 on our ongoing and future directions), this initial step is powerful as it allows for a detailed understanding of the appropriate NN architecture, its capability to learn the relevant physical processes, and its computational stability.

We will follow a two-stage experimental protocol. First, in the so called "offline emulation" (Section 4.2) we aim at learning the physical MPP input-to-output mapping. Later, in the "online emulation" (Section 4.3) the ML-based MPP is plugged into the full sea-ice model Icepack, substituting the original physical MPP. This latter stage is substantially more challenging as it also looks at the computational stability of the ML-based parametrisation in the long run.

Fig. 5. Locations of Icepack simulations, chosen at random from CFSv2 over its $1.875° \times \approx 1.904°$ resolution grid at latitudes between 60°N and 90°N. Yellow locations are those used as training data locations, a total of 80, whilst black locations represent those withheld for the testing stage (20 locations).

4.1. Dataset and neural network setup

We create many simulations of the Icepack model, thus building a training dataset for NNs to learn the parametrisation from. Like with SSA, the forcing data we choose for Icepack comes from the NCEP Climate Forecast System Version 2 (CFSv2). As opposed to our SSA study, there is no need to limit the database samples to an extraction from 12 locations. Thus we choose 100 randomly selected oceanic grid points over the Arctic circle and nearby latitudes in the CFSv2 dataset for our ML experiments (see Fig. 5). The lowest occurrence of sea ice in the Northern Hemisphere occurs along the coast of Hokkaido, Japan [$\approx$ 43°N, 51]. Accordingly, our 100 locations are randomly chosen from the CFSv2 dataset (i.e. over the $1.875° \times \approx 1.904°$ resolution grid at latitudes between 60°N and 90°N). This yields mostly locations within the Arctic Circle whilst permitting training and testing locations in, for example, Hudson Bay, Canada, known for substantial sea ice coverage.

Input to the MPP comprises many model variables, such as surface temperature, level-ice area fraction, water fraction retained for melt ponds, rainfall rate, snow melt, top ice melt, depth difference for snow

Table 3

All variable features (inputs) and targets (outputs) of the melt pond parametrisation in the level-ice scheme with the "mushy thermodynamics" option. Constants are also supplied to the MPP, such as *nilyr* (the number of ice layers), *dt* (time step in seconds), *ktherm* (thermodynamics scheme option), and not all variables supplied to the MPP may be used for all schemes. For example, as the mushy thermodynamics scheme handles flushing, *dpscale*, the alter e-folding time scale for flushing, and *qicen* (the ice layer enthalpy) which would used for explicit calculations for the brine permeability of the ice, are not required. Our table comprises all input that is relevant for the level-ice melt pond parametrisation using the mushy thermodynamics scheme, using all default parameter options in the Icepack model - see Hunke et al. [34] for more details. For each timestep Icepack makes 5 calls to the MPP, one for each ice category.

Inputs/Features and Outputs/Targets in the MPP of Icepack	
Features	**Description**
rfrac	Water fraction retained for melt ponds
meltt	Top melt rate (m/s)
melts	Snow melt rate (m/s)
frain	Rainfall rate (kg/m^2/s)
Tair	Air temperature (K)
fsurfn	Atmosphere-ice surface heat flux (W/m^2)
dhs	Depth difference for snow on sea ice and pond ice (m)
aicen	Ice area fraction
vicen	Ice volume (m)
vsnon	Snow volume (m)
Tsfcn	Surface temperature (° Celsius)
alvl	Level-ice area fraction
apnd[a]	Melt pond area fraction
hpnd[a]	Melt pond height/depth
ipnd[a]	Melt pond refrozen lid thickness
Targets	**Description**
apnd	Melt pond area fraction
hpnd	Melt pond height/depth
ipnd	Melt pond refrozen lid thickness
ffrac	Fraction of fsurfn to melt ipnd

[a] Implies the value from the previous timestep.

on sea ice and pond ice, and so on. The four output variables (e.g., the targets) calculated in the MPP are pond area, pond height, and frozen lid height (their calculated/outputted values at the previous timestep are also used as inputs to the subsequent timestep) and fraction of atmosphere/ice heat flux used to melt frozen lids which is calculated purely from the input variables anew at each timestep, therefore three of the four target variables are calculated explicitly using also their values at the previous timestep. The full list of inputs and targets is given in Table 3.

We train four NNs, one per target, rather than a single multi-target NN. When choosing the number of NNs, one may choose single-task learning (when there is one target per NN) or multi-task learning (when there are multiple targets per NN). See, for example, Caruana [52] for a discussion on single- and multi-task learning.

Our targets are not linked in the way that, e.g., an RGB channel might be where a multi-output NN might be necessary: the processes that lead to different MPP targets can be quite different. We have nonetheless tested both single output NNs and compared this with a single multi output NN (both approaches tuned using the hyperband hyperparameter optimisation). We achieve very similar results, with slightly superior performance scores when using the single output NNs, and thus we choose multiple single target NNs.

As common practice the full dataset is split into a training, validation and test data so as to understand the performance of the model on data that was not used in the training: some data is held back and we ask the model to make predictions for that period. For time series data like we have in this study, data are usually split time-wise [e.g., walk-forward 53]. As we deal with many time series over

multiple locations, we have instead split the data by location, i.e. NNs are trained on full time series data from Icepack simulations in the period 2012–2021 (inclusive) and over 80 locations across the Arctic (yellow dots in Fig. 5). The 20 test locations (black dots in Fig. 5) are used to assess if our emulator has learned the underlying physics and can emulate sea ice melting at other regions in the Arctic. Results using this spatial partition of training and testing data yields close agreement (not shown) to a more standard time partition in which the data 2011–2019 were used for training and 2020–2021 used as test data.

Given the importance of hyperparameters in NNs performance, we initially performed a hyperparameter optimisation [54] using Hyperband [55], a novel bandit-based approach: a form of random search that is sped up through early stopping and adaptive resource allocation. We searched over a hyperparameter space consisting of the number of nodes, layers, and types of activation functions. Hyperband selected four deep NNs that returned slightly better scores for offline training. Performance was assessed relative to a validation set where a training, validation and test set were split by 60, 20 and 20 locations. These deep NNs caused Icepack to become slightly unstable, thus we reduced the number of hidden layers and trained shallower NNs on a merged training and validation set (all 80 yellow locations in Fig. 5). We choose these shallow NNs subjectively, as a trade-off between accuracy and complexity. Hereafter, all results are relative to these latter shallow NNs that are illustrated in Fig. 6. These NNs are trained on the 80 locations over 50 epochs, and we report their performance finally relative to the test locations, which are held back from the model selection process.

4.2. ML-based melt ponds parametrisation - Offline emulation

Fig. 7 shows the scatter plots of the predicted values against the targets (the values predicted by the physical MPP), together with the corresponding R2 score and mean square error (MSE). The figure shows that overall the NNs have been able to learn the feature-to-target relationship in the physical MPP. The performance is very good for three targets, pond area, pond height, and the "ipond height", or height of the ice lid on a frozen pond, achieving extremely high R2 scores ($> \approx$ 0.97 for all three targets). For the fourth target, "Frac A/I Heat Flx", the performance is of medium skill (R2 = 0.568), but this is indeed a much more complex variable representing the fraction of atmosphere/ice heat flux calculated to melt the frozen lids.

The *ffrac* target is almost always zero or one, where samples with $ffrac = 1$ making up $\approx$ 0.25% of the dataset, samples where $0 < ffrac < 1 \approx$ 0.04% of the dataset, and all other samples being where $ffrac = 0$. Thus it represents an extreme imbalanced learning problem, and as *ffrac* is almost always either zero or one, it is very close to an imbalanced binary classification problem. In contrast, most of the ML algorithms, including those we use here, assume that the number of samples in each class is approximately equal [e.g., 56] thus the results will be biased towards the majority group. As learning of *ffrac* represents learning a nonlinear and extremely imbalanced problem, where only rarely does all the ice-atmosphere heat flux contributes to melting an ice lid, it differs substantially to the other three targets. Nonetheless, we shall see in Section 4.3 that the skill of the NNs is sufficient for online emulation of our parametrisation.

From the point of the view the NNs' goal of learning the input–output relation in the physical MPP, the difficulty in the variable *ffrac* relies also in its strong temporal variability. By contrast, the other three variables directly depend on their values at the previous timestep (an input to the MPP, see Table 3), making it possible to predict their future value with sufficient accuracy, even with a mere linear regression; see Table 4. Notwithstanding, Table 4 shows also that, whilst linear regressions do an excellent job in capturing two of the four target variables (pond area and height), NNs perform better than linear regressions for all target variables. In particular, for the thermodynamic-flavoured target variables (*ffrac*), NNs show substantial improvements where linear regressions show almost no predictive ability.
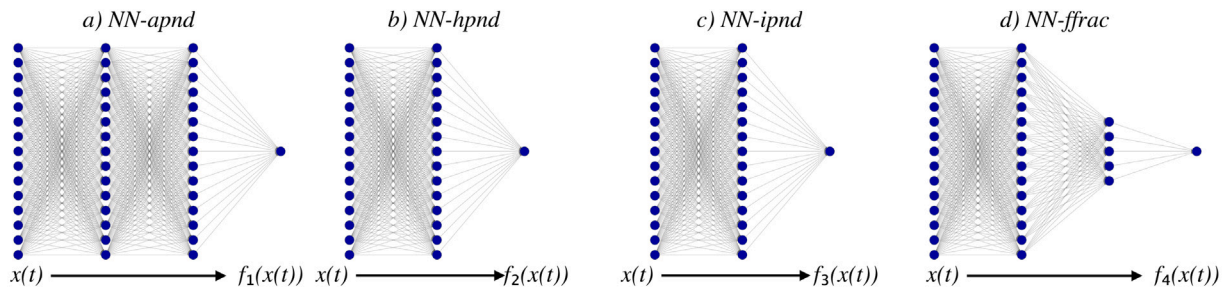
Fig. 6. Representation of the architecture of the neural networks for pond area, pond height, frozen lid ("ipond") height and fraction of atmosphere-ice heat flux used to met frozen lids. All networks use the sigmoid activation function as the output layer, which also comprises the activation functions in the hidden layers for all but the final neural network, which uses the swish activation function for the hidden layers. 50 epochs are used to train the NNs.
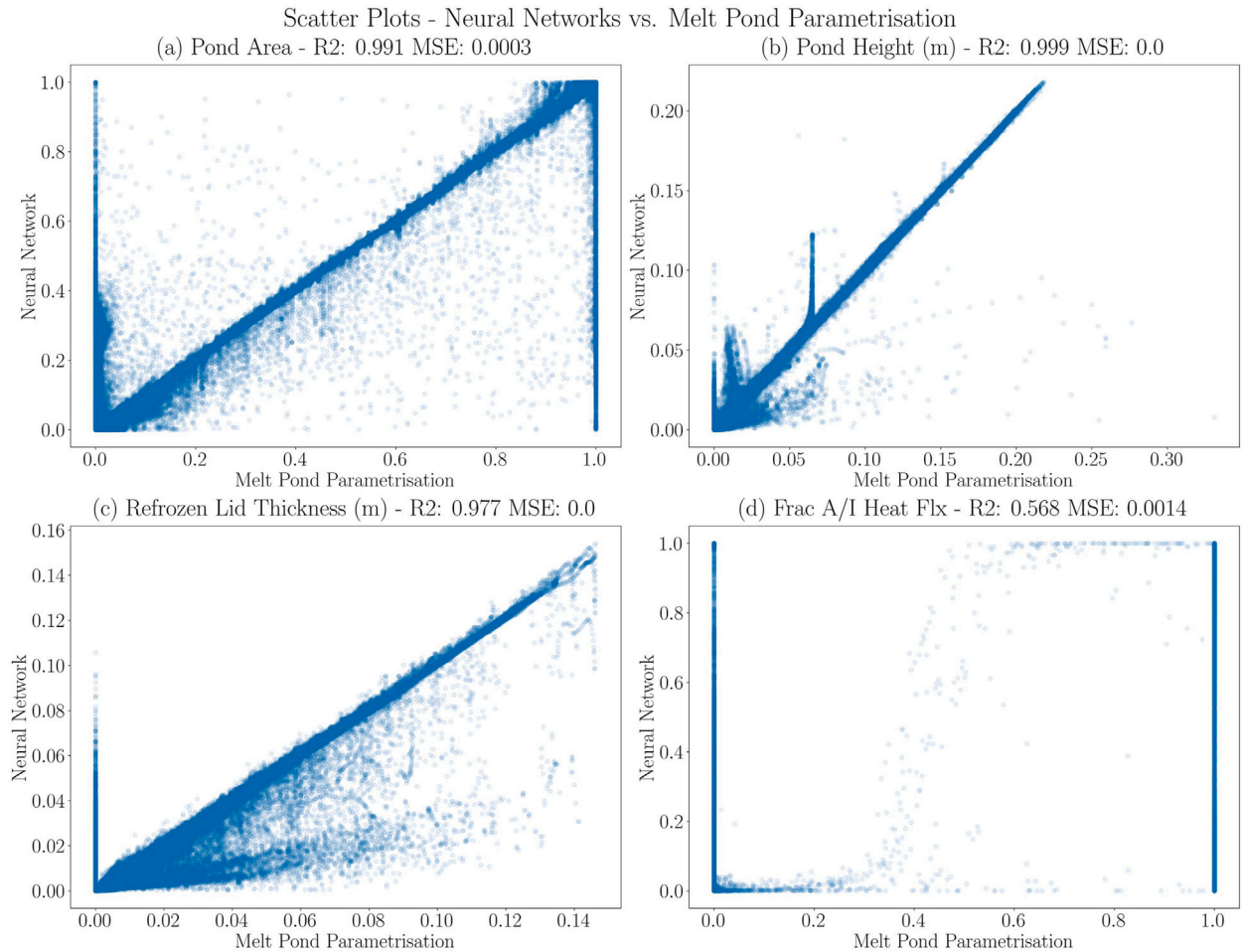

Fig. 7. Scatter plots of the four target variables comparing their true values and the predicted values as given by the neural networks.

4.2.1. Incremental scores

Our aim is to find the simplest possible stable online emulation of a sea ice MPP using NNs, thus offering an alternate route to the existing 'parametric' parametrisations. A key to understanding why it is stable in the online emulation is by exploring the incremental scores in the offline setting. When judging linear regressions by their incremental R2 score instead of their full R2 scores, we see that their predictive skill drops to almost zero. This suggests that the skill the linear regressions show in predicting the full state is essentially due to memory of the previous state of the sea ice. This will be useful in understanding differences in online performance. In the next section, we will show that the moderately good skill of the NNs in learning the incremental changes (R2≥ 0.60 - see Table 5) will reflect in an highly skilled and stable online emulation.

4.3. ML-based melt ponds parametrisation – Online emulation

The "online emulation" challenge consists in substituting the original physics-based MPP in the full sea-ice model Icepack with the ML-based MPP and in studying the new ML-based Icepack accuracy and stability. For completeness and for the sake of understanding the nature and complexity of the processes leading to melt ponds, we compare the original Icepack, hereafter referred to as *Icepack-Phys*, with our nonlinear NN-based MPP, *Icepack-NN*, and with *Icepack-LR* whereby the MPP is based on linear regression. As for the offline emulation, experiments run for the period of 2011–2021, over the 20 test locations displayed in Fig. 5; performance is assessed with regards to locations that both LR- and NN-based emulators have not seen during the training.

Fig. 8. Error time-series in the four MPP output variables: (a) Pond area fraction (apnd), (b) Melt pond height (hpnd), (c) Melt pond refrozen lid thickness (ipnd) and (d) Fraction of atmosphere-ice net heat flux at the surface used to melt (ffrac). Values are computed against Icepack-phys and refer to Icepack-LR (blue) and Icepack-NN (red). Solid lines represent the daily averaged values across all simulations (over all geographic locations), whilst blue (Icepack-LR) and yellow (Icepack-NN) shading above and below represents one standard deviation either side of the mean, respectively. Pearson correlations of the daily mean values of the four target variables between Icepack-phys and Icepack-LR are given in panel e and for Icepack-NN in panel f.

Table 4
Scores of linear regressions and neural networks for each of the four target variables on the test data. Errors are calculated relative to the target variables as given by the MPP itself. *apnd* and *ffrac* are unitless variables, *hpnd* and *ipnd* are measured in units of metres.

Linear Regression and Neural Network Test Data Scores			
Mean Square Error			
	apnd	hpnd	ipnd	ffrac
Lin Reg	6.03×10^{-4}	1.45×10^{-6}	5.56×10^{-6}	3.07×10^{-3}
NN	**3.23×10^{-4}**	**6.24×10^{-7}**	**1.65×10^{-6}**	**1.35×10^{-3}**
R2 Score			
Lin Reg	0.983	0.997	0.923	0.017
NN	**0.991**	**0.999**	**0.977**	**0.568**

Bold numbers indicate better performance.

Table 5
R2 scores as computed when using an increment over small (hourly) timesteps.

Incremental R2 scores					
Linear regressions - R2 scores			Neural networks - R2 scores		
apnd	hpnd	ipnd	apnd	hpnd	ipnd
0.02	0.08	0.08	0.60	0.73	0.71

The linear regression optimised for the offline learning, and presented in Table 4, could not be directly used in the online setting

because, being unbounded, it often predicts statistically sound but non-physical values, such as area fractions of melt ponds greater than one, or less than zero. While this already points to the inherent limitation of standard LR without constraint, for the sake of the comparison, in the experiments that follow we cap the LR output, to only yield physically plausible values. This entails coercing all four target variables with a minimum predicted value of zero, and both *apnd* and *ffrac* with a maximum value of one. For instance, if LRs predict a negative pond height, we set the pond height to zero. It is worth noting that NNs are bounded by construction thanks to the use of the sigmoid activation function for all outputs. Very similar results (not shown here) occur with alternative activation functions, such as ReLU and tanh.

Fig. 8 shows the error time-series of Icepack-LR and Icepack-NN; values are computed relative to Icepack-Phys. In contrast to the very accurate prediction of Pond area fraction (panel a) of Icepack-NN, Icepack-LR drifts rapidly away from Icepack-Phys, simulating unrealistic very large pond area fractions, and near year round substantial pond coverage. Pond heights (panel b) in Icepack-LR are substantially different from Icepack-Phys, and the results suggest overall that Icepack-LR represents a state of very large areas of ponds that are covered with a frozen lid. Correlations between Icepack-Phys and Icepack-LR (panel e) are not significantly correlated. This highlights the poor representation of melt ponds that can be achieved using linear regressions which in turn manifests further the nonlinear characters of the processes.

Fig. 9 has the same format and content as Fig. 8, except that rather than focusing on MPP outputs, it displays error timeseries of key sea-ice variables simulated in Icepack: sea-ice area fraction (*SIAF*; panel a),

Fig. 9. Time series of error for linear regressions and neural networks, both compared to the physical MPP for key output variables. Shading represents one standard deviation width above and below the mean solid lines. (NB: SIAF = "Sea ice area fraction"; AIT = "Average Ice Thickness"; PA = "Pond Albedo"; EPA = "Effective Pond Albedo").

average ice thickness (*AIT*; panel b), pond albedo (*PA*; panel c) and, effective pond area (*EPA*; panel d). Average correlations are given in panels e and f.

Fig. 9 reveals important differences between Icepack-NN and Icepack-LR. Icepack-NN's errors in sea ice area fraction (Fig. 9a) are fairly small, with Icepack-NN on average slightly overestimating the sea ice area fraction. This is due to a slightly smaller area of melt ponds in summer meaning there is less melting occurring and less warming/absorption of incoming downwelling radiation to further warm the sea ice and reduce sea ice fraction. Icepack-LR errors, in these summer months are two to three times larger. Icepack-LR does a substantially worse job at representing melt ponds than Icepack-NN displaying very little summertime melt/feedback.

The implications of this behaviour can be seen whenever the average ice thickness are too large relative to Icepack-Phys and it begins already in the first summer of Icepack-LR simulations. By contrast, Icepack-NN has small errors in the first year (Fig. 9b), in either average ice thickness or sea ice area fraction. This means that Icepack-NN is able to generate melt ponds as it does not suffer from a comparative lack of summer sea ice melting and does not have an immediate build up of ice thickness relative to Icepack-Phys (cf the 0.5 metres mean error for Icepack-LR within one year of emulation).

Figs. 9c and 9d show important variables' evolution for the radiative scheme in Icepack: pond albedo and effective pond area. On the one hand, errors for both are relatively small for Icepack-NN indicating that the emulator is able to produce melt ponds when run in Icepack fairly consistently with the physical MPP. On the other hand, Icepack-LR does not perform equally well. For instance, Fig. 9d shows large areas of melt ponds that are covered with frozen lids (i.e. almost fully covered), whilst Icepack-Phys has positive values for effective pond

area in summer annually. Furthermore, we see annually negative error values for "effective pond area" (which corresponds to the exposed pond area, i.e. the area of a melt pond not covered with a frozen lid, but exposing open water) of Icepack-LR relative to Icepack-Phys, indicative of a very small or no effective pond area. Effective pond area in Icepack-LR in winter is the same as Icepack-Phys: to incoming radiation above it, ice is the same as a pond covered with ice. Accordingly, this yields a substantially increased average sea ice thickness of approximately 1 metre on average by the end of the simulations.

Correlation plots in Fig. 9 show that whilst there exist high correlations between Icepack-LR and Icepack-Phys for sea ice area fraction and average ice thickness (suggesting a large role of model forcing on retaining some relation between the Icepack model output) correlations are poor for melt pond related variables. In general linear regressions do a very poor job of being able to emulate a melt pond parametrisation, necessitating more powerful techniques than using Icepack-LR.

By contrast, Fig. 8 shows no such rapid increase and deviation of values in Icepack-NN relative to Icepack-Phys. Small to moderate errors appear in the summer/melt season, however essentially return back to zero ponds in the winter, unlike Icepack-LR. Icepack-NN slightly underestimates pond area fraction on average, with pond heights overestimated by only around 1 to 5 cm (Fig. 8). Correlation coefficients and the error time series show in general there is a good agreement between Icepack-NN and Icepack-Phys in both Fig. 8 and Fig. 9. A slightly smaller pond area (and effective pond area) contributes to slightly larger average ice thicknesses. Some drift after approximately 4–5 years does occur in average ice thickness in Icepack-NN as a gradual slow increase in relative sea ice thickness. We understand this is caused by a slightly smaller representation of melt ponds in Icepack-NN as

Mutual Information Scores

Fig. 10. Feature selection with Mutual Information - features are ordered with respect to their mutual information scores for each target (apnd, hpnd, ipnd and ffrac). For the definition of each input variable, see Table 3.

compared to Icepack-Phys. Importantly, Icepack-NN does not exhibit the rapid error shifts seen in Icepack-LR, and it attains error levels as small as fractions of those in Icepack-LR. The results suggest that NNs used here have the capability to emulate to a reasonable degree the MPP, with a slight underestimation of melt ponds, and indicate that NNs can be used to emulate a physical parametrisation where, for the most part, they do not cause any substantial drift or instabilities.

In the next Section we perform feature selection on the NN-based MPP in Icepack-NN, to determine if a model with a reduced number of the most important inputs can still give a reasonable performance, laying foundations for an (even further) computationally cheaper, possibly observationally trained, emulator where relations between inputs and outputs are clear and physically interpretable.

4.4. Feature selection using mutual information

Given widespread use, and powerful results in ML, such as the [UAT, 57] which states that there exists a multilayer feedforward NN able to approximate any Borel measurable function to any desired degree of accuracy, use of ML techniques is widespread. This has lead to NNs that can predict a model or function, but it do not provide guides on how to construct it from, for example, a physical perspective or on its interpretability. Machine learning can therefore be "quasi"-blindly used in many diverse problems across the sciences, leading to emulators that perform well but act as a "black-box" [e.g., 58], from where it is difficult to discern the underlying physical, societal, economical, and so on, drivers.

Feature selection helps devising a minimal model by identifying the inputs that have the largest impact on the output, thus evidencing the physical mechanism of greater importance. A reduced model is

easier to interpret and it is usually less computationally demanding. In this study, we perform feature selection using the *mutual information* approach on the training data for each MPP target relative to the input features. Mutual information [59] measures the amount of information one (or group of) random variable(s) contains about another. In essence, it measures the dependency existing between two (groups of) continuous or discrete random variables.

Let X and Y be random variables whose probability density functions are P_X and P_Y and whose domains are $\mathcal{X}$ and $\mathcal{Y}$, respectively. Let us also define their joint probability density function $P_{(X,Y)}$. The mutual information between X and Y is defined by

$$I(X;Y) = -\int_{\mathcal{X}} \int_{\mathcal{Y}} P_{(X,Y)}(x,y) \ln \left(\frac{P_{(X,Y)}(x,y)}{P_X(x)P_Y(y)} \right) \mathrm{d}x \mathrm{d}y, \qquad (5)$$

[e.g., 60]. Mutual information can thus be understood as the reduction of uncertainty on the values of Y once X is known, and vice-versa. It also has the key advantage of detecting non-linear relationships between variables, thus eliminating the need to make any assumptions about the linearity between features and targets.

Fig. 10 shows the mutual information scores for each target variable with respect to the features. For pond area (Fig. 10a), pond height (Fig. 10b) and refrozen lid height, Fig. 10(c) the largest mutual information score is between that target and the value of that same variable at the previous timestep (recall that they are among the features; see Table 3): these play the most dominant roles in determining the value at next time step. By contrast, for the fraction of atmosphere/ice heat flux used to melt frozen lids ("Frac A/I H flux", Fig. 10d) a more gradual decline is seen suggesting that predictive power for this variable comes from the cumulative total of including more features.

Identical NNs, as used in Icepack-NN (Fig. 6), are trained using $n = 1, \ldots, 15$ features (see Table 3) ranked by mutual information scores

Fig. 11. R2 scores for each target with respect to the number of features included in training. Neural networks with identical architecture are trained for 25 epochs using $n = 1, \dots, 15$ features, for targets pond area, pond height, ipond height and fraction of atmosphere-ice heat flux used to melt frozen lids. The logarithmic scale is used in the y-axis.

for each target. For example, for pond area, based on the results in Fig. 10, this implies training NNs as follows: for $n = 1$ using pond area at previous timestep as the sole feature, for $n = 2$, pond area and pond height at previous timestep, for $n = 3$ pond area, pond height at previous timestep together with surface temperature, and so on, until all features are included ($n = 15$). This is repeated for all four target variables, making a total of 60 NNs trained for 25 epochs each. The R2 scores of these NNs in the offline emulation setting are shown in Fig. 11.

Consistent with mutual information scores, we observe that for pond area, height and frozen lid height very good R2 scores are attained with as few as the (appropriate) one, two or three inputs. Recall from Fig. 10 that these very influential features are the same variables at the previous timestep. This suggests that one may produce a good emulator of melt ponds knowing their values at previous timestep and a few key variables such as surface temperature. However, for the fraction of atmosphere/ice heat flux (red line), many more factors concur to determining its value. As a consequence, R2 scores of NNs with few features (< 8) are very low, and effectively the full set of features is needed to reach R2 scores as good as those seen for the first three variables with only few (≤ 3) features.

We repeat now the same inspection on the needed features in the context of online emulation and, for 1 to 15 features we run Icepack-NN over all 20 test locations (i.e. Icepack-NN where four separate NNs use 1 different feature, 2 features, 3 features, and so on, in order of importance as given by the mutual information criterion). Note that the case $n = 15$ coincides with the Icepack-NN with full features discussed in Section 4.3. Fig. 12 shows the mean squared error (MSE; panels a and c) and Pearson correlation values (panels b and d) for the four melt pond target variables. Whilst two to three features have shown to suffice for returning offline results that are similar to the MPP, when run online this is not sufficient to replicate the performance of Icepack-Phys; 4 to 6 variables would be necessary to emulate and substitute a physical MMP.

5. Discussion and future work

By modifying the sea ice albedo, melt ponds play a central role in regulating the energy balance in the Arctic region [61]. Melt pond parametrizations (MPPs) have thus became crucial in present-day sea ice models [22,23]. Yet the multi scale processes leading the formation of a melt pond are not fully understood, nor they can all be explicitly resolved in sea ice models. Nowadays, and likewise most of the parametrizations in climate models, MMPs are derived from physical principles and have the virtue to also generally respect physical and dynamical balances. Nevertheless, they are also often over

parametrized, i.e. a large number of uncertain parameters have to be (possibly simultaneously) tuned. In this context, a further complication arises if the model exhibits high sensitivity to the parameters in the MMPs, whose accurate determination is then paramount, although extremely difficult.

We have studied a known advanced MMPs, the level-ice MPP [40], included in a prototypical 1d state-of-the-art model of the sea-ice thermodynamics, Icepack [34] We performed a quantitative examination of the its sensitivity to the MPP parameters using the Sobol Sensitivity Analysis [36]. Results indicated that the sensitivities of Icepack to its MPP is very large and inhomogeneous. Parameters significantly impact Icepack predicted output both in time and space. As a consequence, even little differences in the MPP parameter values lead to substantially different Icepack behaviours. The Icepack response is not only different in the long term but it also display a strong geographical dependency. The unavoidable uncertainty in the MPP parameters, together with the MPP high sensitivity to parameters themselves, can be the source of large sea-ice prediction errors.

In the second part of this study we have thus investigated whether machine learning (ML) and in particular neural networks (NNs) are (i) able to learn the physical processes in the physical MPPs, and, (ii) do that with a smaller amount of inputs than those needed by the original physical MPPs. Neural networks were shown to be capable of learning and emulating the MPP. In particular, although simpler linear regression models shown capabilities of learning the input-to-output processes in the MPP (offline emulation) with some skill, only nonlinear NNs were then able to successfully substitute the original physical MMP in the full Icepack (online emulation). In contrast, with the MPP emulator based on linear regression, Icepack fails in reproducing the original Icepack with physical MMP.

Our results indicate that NNs can learn and represent a viable subgrid-scale parametrisation, and also that a reduced, simpler data driven emulator can act as a MMP in the Icepack model. Feature selection and the associated trained NNs demonstrate that the melt pond parametrisation is too complex for successful online emulation with only one or two features, but is successful with fewer than all of the input variables requested in the physical MMP. We showed that performance improves when some melt pond properties (pond area fraction, pond height and refrozen lid height) at the previous timestep are known, along with a few other fundamental variables, such as near surface and air temperatures and snow melt rate. In summary, our results suggest that key thermodynamic variables such as temperature, conservation of mass related variables regarding available meltwater (*e.g.*, melt rate of snow), and knowledge of the melt pond at the previous timestep may be key when focusing on an observationally trained emulator.

Offline performance of ML models does not guarantee good online simulations, where instability and drift can arise in the online implementation (*e.g.*, Rasp et al. [26], Brenowitz et al. [62]) and approaches to address this have received substantial attention. Yuval and O'Gorman [63] and Watt-Meyer et al. [64] swapped NNs for random forests which helped them achieve stability in online settings. Kurz et al. [65] use reinforcement learning to help address stability in turbulence modelling of large eddy simulations. Rasp [66] used online learning to tackle substantial instabilities and biases in NN parametrisations in the Lorenz 96 model, and also showed that models may need to sample a larger model state space through an additional online training step to aid stability.

When Rasp et al. [26] investigated the potential of NNs to represent short-term cloud resolving simulations, they used a deeper network of around half a million parameters, motivated combat instabilities and unrealistic artifacts previously seen in shallower architectures (*e.g.*, Krasnopolsky et al. [29], Brenowitz and Bretherton [25]).

In this study we have shown that it is possible to achieve online emulation of a MPP in a sea ice model using relatively shallow NNs. Our NN-based emulation is stable for up to a decade without showing

Fig. 12. Feature selection performance for online simulation. MSE and Correlation Coefficient scores.

imbalances, instabilities or drift. While the absence of instabilities can in part be attributable to the slower evolving nature of the sea ice as compared to, e.g., chaotic dynamics as in Lorenz 96 model, the lack of drifts and imbalances signify that the NN-based emulator does not introduce bias and it is consistent with the underlying physics.

When seeking to emulate melt ponds and replace existing parametrisations, a final goal is not to merely seek to replace the parametrisation, but to improve upon it. This study is a first step towards the design of fully data-driven MMPs. Our results are encouraging and demonstrate, as others have for other parametrization in climate models (e.g. for clouds see Krasnopolsky et al. [29]), that emulators can replace MPPs in regional and global climate models. Our next steps are to build an emulator from observations and use this to substitute the parametrisation in the Icepack model with a data driven parametrisation. A key challenge that we anticipate is the availability of accurate raw data, whereby accuracy is intended here in term of their spatio-temporal distribution, their precision, as well as their direct relation to the physical quantities of interest. We envisage using recently developed combined data assimilation and ML methods that have shown power in handling sparse and noisy data [67].

CRediT authorship contribution statement

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

[43] J.R. Urrego-Blanco, N.M. Urban, E.C. Hunke, A.K. Turner, N. Jeffery, Uncertainty quantification and global sensitivity analysis of the Los Alamos sea ice model, J. Geophys. Res.: Oceans 121 (4) (2016) 2709–2732, http://dx.doi.org/10.1002/2015JC011558, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015JC011558.

[44] A. Saltelli, Making best use of model evaluations to compute sensitivity indices, Comput. Phys. Comm. (ISSN: 0010-4655) 145 (2) (2002) 280–297, http://dx.doi.org/10.1016/S0010-4655(02)00280-1, URL https://www.sciencedirect.com/science/article/pii/S0010465502002801.

[45] Y. Tang, P. Reed, T. Wagener, K. van Werkhoven, Comparing sensitivity analysis methods to advance lumped watershed model identification and evaluation, Hydrol. Earth Syst. Sci. 11 (2) (2007) 793–817, http://dx.doi.org/10.5194/hess-11-793-2007, URL https://hess.copernicus.org/articles/11/793/2007/.

[46] L. Shapley, A value for n-person games, in: Contributions to Theory Games, Vol. 2, Princeton University Press, 1953.

[47] T. Homma, A. Saltelli, Importance measures in global sensitivity analysis of non-linear models, Reliab. Eng. Syst. Saf. (ISSN: 0951-8320) 52 (1) (1996) 1–17, http://dx.doi.org/10.1016/0951-8320(96)00002-6, URL https://www.sciencedirect.com/science/article/pii/0951832096000026.

[48] J. Herman, W. Usher, SALib: An open-source python library for sensitivity analysis, J. Open Source Softw. 2 (9) (2017) http://dx.doi.org/10.21105/joss.00097.

[49] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, S. Tarantola, Global Sensitivity Analysis: The Primer, Wiley, ISBN: 978-0-470-05997-5, 2008.

[50] S. Saha, S. Moorthi, X. Wu, J. Wang, S. Nadiga, P. Tripp, D. Behringer, Y.-T. Hou, H.-Y. Chuang, M. Iredell, M. Ek, J. Meng, R. Yang, M. Peña, H. Dool, Q. Zhang, W. Wang, M. Chen, E. Becker, The NCEP climate forecast system version 2, J. Clim. 27 (2014) 2185–2208.

[51] S. Takahashi, T. Kosugi, H. Enomoto, Sea-ice extent variation along the coast of Hokkaido, Japan: Earth's lowest-latitude occurrence of sea ice and its relation to changing climate, Ann. Glaciol. 52 (58) (2011) 165–168, http://dx.doi.org/10.3189/172756411797252301.

[52] R. Caruana, Multitask learning, Mach. Learn. 28 (1997) http://dx.doi.org/10.1023/A:1007379606734,

[53] R. Assaad, S. Fayek, Predicting the price of crude oil and its fluctuations using computational econometrics: Deep learning, LSTM, and convolutional neural networks, Econometr. Res. Finance 6 (2021) 119–137, http://dx.doi.org/10.2478/erfin-2021-0006.

[54] F. Hutter, H. Hoos, K. Leyton-Brown, An efficient approach for assessing hyperparameter importance, in: E.P. Xing, T. Jebara (Eds.), Proceedings of the 31st International Conference on Machine Learning, in: Proceedings of Machine Learning Research, Vol. 32, (1) PMLR, Bejing, China, 2014, pp. 754–762, URL https://proceedings.mlr.press/v32/hutter14.html.

[55] L. Li, K. Jamieson, G. DeSalvo, A. Rostamizadeh, A. Talwalkar, Hyperband: A novel bandit-based approach to hyperparameter optimization, 2018.

[56] B. Krawczyk, Learning from imbalanced data: open challenges and future directions, Progr. Artif. Intell. 5 (4) (2016) 221–232, http://dx.doi.org/10.1007/s13748-016-0094-0,

[57] K. Hornik, M. Stinchcombe, H. White, Multilayer feedforward networks are universal approximators, Neural Netw. (ISSN: 0893-6080) 2 (5) (1989) 359–366, http://dx.doi.org/10.1016/0893-6080(89)90020-8, URL https://www.sciencedirect.com/science/article/pii/0893608089900208.

[58] O. Camburu, Explaining deep neural networks, (Ph.D. thesis), University of Oxford, 2020.

[59] C.E. Shannon, A mathematical theory of communication, Bell Syst. Tech. J. (379–423) (1948) 27.

[60] B. Frénay, G. Doquire, M. Verleysen, Is mutual information adequate for feature selection in regression? Neural Netw. (ISSN: 0893-6080) 48 (2013) 1–7, http://dx.doi.org/10.1016/j.neunet.2013.07.003, URL https://www.sciencedirect.com/science/article/pii/S0893608013001883.

[61] C. Polashenski, D. Perovich, Z. Courville, The mechanisms of sea ice melt pond formation and evolution, J. Geophys. Res.: Oceans 117 (C1) (2012) http://dx.doi.org/10.1029/2011JC007231, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2011JC007231.

[62] N.D. Brenowitz, T. Beucler, M. Pritchard, C.S. Bretherton, Interpreting and stabilizing machine-learning parametrizations of convection, J. Atmos. Sci. 77 (12) (2020) 4357–4375, http://dx.doi.org/10.1175/JAS-D-20-0082.1, URL https://journals.ametsoc.org/view/journals/atsc/77/12/jas-d-20-0082.1.xml.

[63] J. Yuval, P.A. O'Gorman, Stable machine-learning parameterization of subgrid processes for climate modeling at a range of resolutions, Nature Commun. 11 (1) (2020) 3295, http://dx.doi.org/10.1038/s41467-020-17142-3,

[64] O. Watt-Meyer, N.D. Brenowitz, S.K. Clark, B. Henn, A. Kwa, J. McGibbon, W.A. Perkins, C.S. Bretherton, Correcting weather and climate models by machine learning nudged historical simulations, Geophys. Res. Lett. 48 (15) (2021) http://dx.doi.org/10.1029/2021GL092555, URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021GL092555, e2021GL092555 2021GL092555.

[65] M. Kurz, P. Offenhäuser, A. Beck, Deep reinforcement learning for turbulence modeling in large eddy simulations, Int. J. Heat Fluid Flow (ISSN: 0142-727X) 99 (2023) 109094, http://dx.doi.org/10.1016/j.ijheatfluidflow.2022.109094, URL https://www.sciencedirect.com/science/article/pii/S0142727X2200162X.

[66] S. Rasp, Coupled online learning as a way to tackle instabilities and biases in neural network parameterizations: general algorithms and Lorenz 96 case study (v1.0), Geosci. Model Dev. 13 (5) (2020) 2185–2196, http://dx.doi.org/10.5194/gmd-13-2185-2020, URL https://gmd.copernicus.org/articles/13/2185/2020/.

[67] S. Cheng, C. Quilodran-Casas, S. Ouala, A. Farchi, C. Liu, P. Tandeo, R. Fablet, D. Lucor, B. Iooss, J. Brajard, D. Xiao, T. Janjic, W. Ding, Y. Guo, A. Carrassi, M. Bocquet, R. Arcucci, Machine learning with data assimilation and uncertainty quantification for dynamical systems: a review, 2023.

Simon Driscoll (s.driscoll@pgr.reading.ac.uk) is a Ph.D. student focusing on machine learning and thermodynamics of sea ice at the University of Reading. Trained as a mathematician, he completed a DPhil focusing on volcanic eruptions and their effects on climate, and then left academia briefly. He returns with a passion for machine learning.

Alberto Carrassi (alberto.carrassi@unibo.it), is a Professor at the University of Bologna. trained as a physicist, with a Ph.D. thesis on dynamical systems, has been always working at the crossroad between applied mathematics and climate. Research breakthroughs include contribution to the development of algorithms known as the assimilation in the unstable subspace unifying dynamical systems and data assimilation or of a novel paradigm for causal inference using data assimilation for detection and attribution of climate change. He contributed to pioneer powerful combined methods merging data assimilation and machine learning, making possible using machine learning with sparse and noisy datasets.

Julien Brajard (julien.brajard@nersc.no) is an associate professor at Sorbonne University (Paris, France) since 2009 and a researcher at NERSC (Nansen Environmental and Remote Sensing Center) since 2018. He works at proposing new methodologies at the crossroads between data assimilation and machine learning. His objective is to improve the forecast of key climate variables by correcting model errors using data, in particular from remote sensing.

Laurent Bertino (laurent.bertino@nersc.no) is a Research Leader at NERSC with a background in Geostatistics and data assimilation. He works on ensemble-based data assimilation techniques with applications to oceanography and sea ice forecasting in the Arctic. He has published 84 peer-reviewed papers.

Marc Bocquet (marc.bocquet@enpc.fr) is Professor at Ecole des Ponts (France) and deputy director of CEREA laboratory. He works on the methods of data assimilation, machine learning, inverse problems and environmental statistics, with applications to dynamical systems, atmospheric chemistry and transport, and meteorology. He is a fellow of ECMWF, and editor for the QJRMS, Foundation of Data Science, and Frontiers in Applied Mathematics and Statistics. He has published over 115 peer-reviewed papers.

Einar Olason (Einar.Olason@nersc.no) is Research Leader/Senior Researcher in the Sea Ice Modelling Group at NERSC. With a Doctorate in physical oceanography, he works on sea ice modelling and dynamics, numerical models and atmosphere-ocean-ice interactions. Author of over 40 publications, research discoveries and topics include drivers of variability in Arctic sea ice drift speed, multi-fractal scaling properties of sea ice deformation and development of a new Lagrangian sea ice model.