Robustness of Neural Network Emulations of Radiative Transfer Parameterizations in a State-of-the-Art General Circulation Model

Alexei Belochitski[1,2], Vladimir Krasnopolsky[2]

[1]IMSG, Rockville, MD 20852, USA
[2]NOAA/NWS/NCEP/EMC, College Park, MD 20740, USA

Correspondence to: Alexei Belochitski (alexei.a.belochitski@noaa.gov)

Abstract. The ability of Machine-Learning (ML) based model components to generalize to the previously unseen inputs, and its impact on stability of the models that use these components, has been receiving a lot of recent attention, especially in the context of ML-based parameterizations. At the same time, ML-based emulators of existing physically based parameterizations can be stable, accurate, and fast when used in the model they were specifically designed for. In this work we show that shallow-neural-network-based emulators of radiative transfer parameterizations developed almost a decade ago for a state-of-the-art general circulation model (GCM) are robust with respect to the substantial structural and parametric change in the host model: when used in two seven month-long experiments with a new GCM, they remain stable and generate realistic output. We concentrate on the stability aspect of the emulators' performance and discuss features of neural network architecture and training set design potentially contributing to the robustness of ML-based model components.

1. Introduction

One of the main difficulties in developing and implementing high-resolution environmental models is complexity of the physical processes involved. For example, the calculation of radiative transfer in a GCM often takes a significant part of the total model run time. From the standpoint of basic physics, radiative transfer is well understood. Very accurate, but computationally complex benchmark models exist (Oreopoulos et al, 2012) that demonstrate excellent agreement with observations (Turner et al, 2004). Parameterizations of radiative transfer seek a compromise between accuracy and computational performance. Arguably, the biggest simplification they make is treatment of radiative transfer as a 1-D as opposed to a 3-D process (Independent Column Approximation, ICA): both solar, or short-wave (SW), and terrestrial, or long-wave (LW), radiation is considered to flow within the local column of the model, up and down the local vertical (two stream approximation), but not between columns. This approximation works well at spatial resolutions characteristic of general circulation models of the atmosphere (Marshak and Davis, 2005). To integrate over the spectrum of radiation, parameterizations split it into several broad bands and a number of representative spectral intervals that are treated monochromatically (Fu and Liou, 1992). State-of-the-art parameterizations can reproduce benchmark calculations to a high degree of accuracy even with these simplifications, but they still require substantial computational expense.

Radiative transfer parameterizations supply their host model with broadband fluxes and heating rates, which are obtained by integration over time, space, and frequency. Therefore, a trade-off between accuracy and computational expense can be found in how finely these dimensions are discretized (Hogan et al, 2017).

- *Discretization in time.* All GCMs update their radiative heating/cooling rates less frequently than the rest of the model fields. For example, National Centers for Environmental Prediction (NCEP) Global Forecast System (GFS) v16 general circulation model (GCM) in its operational configuration updates its radiative fields once per model hour, while updates to temperature, moisture, most cloud properties etc. due to unresolved physics processes happen every 150 model seconds, or 24 times per single radiation call. Updates due to dynamical processes happen even more frequently, every 12.5 seconds (Kain et al, 2020). This approximation is good for slowly changing fields of certain radiatively active gases but is less justified for small-scale clouds with lifetimes of an hour or less.
- *Discretization in space.* Some GCMs calculate radiative fields on a coarser spatial grid and interpolate them onto a finer grid used for the rest of the model variables. For example, radiation grid in European Centre for Medium-Range Weather Forecasts (ECMWF) Integrated Forecast System (IFS) v43R3 in the ensemble mode is 6.25 times coarser than the physics grid (Hogan et al, 2017). This may cause 2m temperature errors in areas of surface heterogeneity, e.g. coasts (Hogan and Bozzo, 2015).
- *Discretization/sampling in frequency space.* The Rapid Radiative Transfer Model (RRTMG), a parameterization of radiative transfer for GCMs used in NCEP GFS and ECMWF IFS, utilizes 14 bands in the short wave (Mlawer et al, 1997), while the parameterization used at United Kingdom Met Office Unified Model utilizes 6 (Edwards and Slingo, 1996). Monte Carlo Spectral Integration (Pincus and Stevens, 2009) performs integration over only a part of the radiative spectrum, randomly chosen in each point in time and space, allowing to increase temporal/spatial resolution of radiation calculations. Monte Carlo Integration of the Independent Column Approximation (McICA) (Pincus et al, 2003), integrates over the entire spectrum, but samples subgrid-scale (SGS) cloud properties in a random, unbiased manner, in each grid column in time and space, instead of integrating over them.

All of the methods for improving computational efficiency of radiative transfer parameterizations outlined above are either numerical and/or statistical in nature. In recent years there has been a substantial increase in interest in adding machine learning (ML) techniques to the arsenal of these methods. It has been accomplished in at least two different ways: 1) as an emulation technique for accelerating calculations of radiative transfer parametrizations or their components, 2) as a tool for development of new parameterizations based on data simulated by more sophisticated models and/or reanalysis.

An ML-based emulator of a model physics parameterization is a functional imitation of this parameterization in a sense that the results of model calculations with the original parameterization and with its ML emulator are close to each other by a metric appropriate for an application at hand as to be identical for the practical purposes. From the mathematical point of view, model physics and individual parameterizations are mappings

$$Y = M(X);\ X \in \Re^n,\ and\ Y \in \Re^m \quad (1)$$

where *n* and *m* are the dimensionalities of the input and output vector spaces correspondingly. Therefore, emulating existing parameterizations using ML techniques is a mapping approximation problem. In practice, this mapping can be defined by a set of its input and output vectors that is obtained by running the original model with the parameterization that is to be emulated, and saving inputs and outputs of this parameterization with a frequency and spatiotemporal coverage sufficient to comprehensively cover domain and range of the mapping. This data is then used for the emulator training. This approach allows to achieve very high accuracy of approximation because model output, unlike empirical data, is neither noisy nor sparse.

The domain and range of the mapping is defined not only by the parameterization that is being emulated, but by the entirety of the atmospheric model environment: the dynamical core, the suite of physical parameterizations, and the set of configuration parameters for both. Once any of these components and/or parameters are modified, the set of possible model states is altered as well, possibly now including states that were absent in the emulator's training data set.

How accurately should the emulator approximate the original mapping? Unbiased, random, uncorrelated errors in radiative heating rates with magnitudes as large as the net cooling rate do not statistically affect forecast skill of an atmospheric model (Pincus et al, 2003). From the physical standpoint this can be understood in the following way: random small local heating rate errors in the bulk of the atmosphere lead to local small-scale instabilities that are mixed away by the flow; however, there no such mechanism for the surface variables, such as skin temperature, and errors in surface fluxes can be more consequential (Pincus and Stevens, 2013). Therefore, it may be useful to think of the above as necessary conditions on approximation error of an ML emulator of a radiative transfer parameterization for it to be a successful functional imitation of the original scheme.

Developing a stable and robust NN-based emulator is a multifaceted problem that requires deep understanding of multiple technical aspects of the training process and details of NN architecture. Many techniques for stabilization of hybrid statistical-deterministic models have been developed. Compound parameterization has been proposed for climate and weather modeling applications where an additional NN is trained to predict errors of the NN emulator, and, if the predicted error is above a certain threshold, compound parameterization falls back to calling the original physically-based scheme (Krasnopolsky et al, 2008). Stability theory was used to identify the causes and conditions for instabilities in ML parameterizations of moist convection when coupled to idealized linear model of atmospheric dynamics (Brenowitz et al, 2020). An NN optimization via random search over hyperparameter space resulted in considerable improvements in stability of subgrid physics emulators in the Super-parameterized Community Atmospheric Model version 3.0 (Ott et al, 2020). A coupled online learning approach was proposed where a high-resolution simulation is nudged to the output of a parallel lower-resolution hybrid model run, and the ML-component of the latter is retrained to emulate tendencies of the former, helping to eliminate biases and unstable feedback loops (Rasp, 2020). Random forests approach was successfully used to build a stable ML parameterization of

convection (Yuval and O'Gorman, 2020). Physical constraints were used to achieve stability of hybrid models (e.g., Yuval et al, 2021; Kashinath et al, 2021).

In this work we present robust and stable shallow-NN based emulators of radiative transfer parameterizations. We explore how much of a change in the model's phase space (as well as the original parameterization's domain and range) a statistical model like the NN can tolerate. We will approach this question by installing shallow-NN-based emulators of LW and SW RRTMG developed in 2011 for NCEP CFS (Krasnopolsky et al, 2010) into the new version 16 of NCEP GFS that became operational in March of 2021. Given the scope of changes in the host model (described in Section 3), we do not expect results of parallel runs to be identical, therefore, we will mostly concentrate on the stability aspect of the emulators' performance.

In Section 2, we briefly describe design aspects of these and other emulators of radiative transfer parameterizations reported in literature so far. In Section 3, we outline major differences between the 2011 version of CFS and the GFS v16, and describe numerical experiments with SW and LW emulators developed for the 2011 version of CFS (Krasnopolsky et al, 2010) and incorporated into GFS v16. Results of these experiments are examined in Section 4. Section 5 discusses aspects of neural network architecture and training set design potentially contributing to stability of ML-based model components. Conclusions are formulated in Section 6.

2. Survey on Technical Aspects of Existing ML Emulators of Radiative Transfer Parameterizations

NeuroFlux, a shallow-neural-network-based LW radiative transfer parameterization, developed at ECWMF, was in part an emulator and in part a new ML-based parameterization (Chevallier et al, 1998, 2000). It consisted of multiple neural networks (NNs), each utilizing a hyperbolic tangent as an activation function (AF), but using a varying number of neurons in the single hidden layer: two NNs were used to generate vertical profiles of, respectively, up- and down-welling clear sky LW fluxes per each vertical layer; and a battery of NNs, two per each vertical layer of the host model, was used to compute profiles of up- and down-welling fluxes due to blackbody cloud on a given layer, with overall fluxes calculated using the multilayer graybody model. Training set for clear-sky NNs contained 6000 cloudless profiles from global ECMWF short-range forecasts; one day worth of three-hourly data per month of a single year was utilized. From this set, multiple training sets for cloudy sky NNs were derived, each containing 6000 profiles as well: a cloud with the emissivity of unity was artificially introduced on a given vertical layer, and radiative transfer parametrization was used in the offline mode to calculate resulting radiative fields. NeuroFlux was accurate and about an order of magnitude as fast as the original parameterization in a model with 31 vertical layers. It had been used operationally within the ECMWF four-dimensional variational data assimilation system (Janiskova et al, 2002). However, in model configurations with 60 vertical layers and above, NeuroFlux could not maintain the balance between speed up and accuracy (Morcrette et al, 2008).

The approach based on pure emulation of existing LW- and SW-radiative transfer parametrizations using NNs has been pursued at NCEP Environmental Modeling Center (Krasnopolsky et al, 2008,2010, 2012; Belochitski et al, 2011). In this approach, two shallow NNs with hyperbolic tangent activation functions, one for LW and the other for SW radiative transfer, generate heating rate profiles as well as surface and top-of-the-atmosphere radiative fluxes, replacing the entirety of respective RRTMG LW and SW parameterizations. Not only radiative transfer solvers were emulated, but also the calculations of gas and cloud optical properties (aerosol optical properties were prescribed from climatology). Two different pairs of emulators were designed for two different applications: climate simulation and medium-range weather forecast, each differing in the training set design. The data base for the former application was generated by running NCEP Climate Forecast System (CFS), a state-of-the-art fully coupled climate model, for 17 years (1990-2006) and saving instantaneous inputs and outputs of RTTMG every three hours for one day on the 1st and the 15th of each month, to sample diurnal and annual cycles, as well as decadal variability and states introduced by time-varying greenhouse gases and aerosols. 300 global snapshots were randomly chosen from this database, and consequently split into three independent sets for training, testing, and validation, each containing about 200,000 input/output records (Krasnopolsky et al, 2010). The data set for the medium range forecast application was obtained from 24 10-day NCEP GFS forecasts initialized on the 1st and the 15th of each month of 2010, with each forecast saving instantaneous three-hourly data. Independent data sets were obtained following the same procedure as for the climate application (Krasnopolsky et al, 2012).

Dimensionality of data sets and NN input vectors for both applications was reduced in the following manner: some input profiles (e.g. pressure) that are highly correlated in the vertical were sampled on every other level without decrease in approximation accuracy; some inputs that are uniformly constant above certain level (water vapor) or below a certain level (ozone) were excluded from the training set on these levels; inputs that are given by prescribed monthly climatological look up tables (e.g. trace gases, tropospheric aerosols) were replaced by latitude and periodic functions of longitude and month number; inputs given by prescribed monthly time series (e.g. carbon dioxide, stratospheric aerosols) were replaced by the year number and periodic function of month number. No reduction of dimensionality was applied to outputs.

A very high accuracy and up to two orders of magnitude increase in speed as compared to the original parameterization for both NCEP CFS and GFS full radiation has been achieved for model configurations with 64 vertical levels. The systematic errors introduced by NN emulations of full model radiation were negligible and did not accumulate during the decadal model simulation. The random errors of NN emulations were also small. Almost identical results have been obtained for the parallel multi-decadal climate runs of the models using the NN and the original parameterization, and in the limited testing in the medium-range forecasting mode. Regression trees were explored as an alternative to NNs and were found to be nearly as accurate in a 10-years long climate run while requiring much more computer memory due to the fact that the entire training data set has to be stored in memory during model integration (Belochitski et al, 2011).

Using the approach developed at NCEP, an emulator of RRTMG consisting of a single shallow NN that replaces both LW and SW parameterizations at once was developed at Korean Meteorological Agency for the short-range weather forecast model Korea Local Analysis and Prediction System in an idealized configuration with 39 vertical layers (Roh and Song, 2020). Inputs and outputs to RRTMG were saved on each 3 second time step of a 6 hour-long simulation of a squall line, and about 270,000 input/output pairs were randomly chosen from this data set to create training, validation, and testing sets. Dimensionality reduction was performed by removing constant inputs. Several activation functions were tested (tanh, sigmoid, softsign, arctan, linear) with hyperbolic tangent providing best overall accuracy of approximation. The emulator was two orders of magnitude as fast as the original parameterization, and was stable in a 6 hour-long simulation.

Two dense, fully-connected, feed-forward deep-NN-based emulators with three hidden layers, one emulator per parametrization, were developed for LW and SW components of RRTMG-P for the Department of Energy's Super-Parametrized Energy Exascale Earth System Model (SP-E3SM) (Pal et al, 2019). In SP-E3SM, radiative transfer parameterizations act in individual columns of a 2-D cloud resolving model with 31 vertical levels embedded into columns of the host GCM. Calculation of cloud and aerosol optical properties were not emulated, instead, original RRTMG-P subroutines were used. Inputs and outputs of radiative parameterizations were saved on every time step of a year-long model run, with 9% of this data randomly chosen to form a data set of 12,000,000 input/output records for LW, and of 6,000,000 input/output records for SW emulator training and validation. 90% of the data in these sets was used for training, and 10% for validation and testing. No additional dimensionality reduction was performed. Sigmoid AF was chosen as it was found to provide slightly better training convergence than the hyperbolic tangent. The emulator was an order of magnitude faster than the original parametrization and was stable in a year-long run.

Recently, an entire suite of model physics in NCEP GFS, that among other parameterizations includes LW and SW RRTMG schemes, was emulated with a single shallow NN. It was shown that the emulator is accurate and stable in multiple 10-day forecasts and in one-year continuous run (Belochitski and Krasnopolsky, 2021).

A number of ML-based radiative transfer parameterizations or their components have been developed, but, to our knowledge, have not yet been tested in an online setting, or in interactive coupling to an atmospheric model. Among them are deep-NN-based parameterizations of gas optical properties for RTTMG-P (Ukkonen et al, 2020; Veerman et al, 2021), and a SW radiative transfer parameterization based on convolutional deep neural networks (Lagerquist et al, 2021).

3. Design of Numerical Experiments with GFS v16

GFS v16 differs from the 2011 version of the atmospheric component of NCEP CFS in a number of ways, most relevant of which are summarized in Table 1.

	CFS 2011	GFS 2021
Dynamical core	Spectral Eulerian	Finite Volume Cubed Sphere
Horizontal resolution	T126 (~100 km)	C768 (~13 km)
Vertical res. and coordinate	64 levels, hybrid sigma-p	127 levels, hybrid sigma-p
Physics Grid	Gaussian	Cubed Sphere
Radiation	RRTMG v2.3	RRTMG LW v4.82, SW v3.8
Microphysics	Zhao-Carr, single moment, two species, one prognostic variable	GFDL, single moment, five species, five prognostic variables
Planetary Boundary Layer	K-profile	Hybrid TKE-EDMF
Middle atm. H_20 photochemistry	None	Climatological
0_3 photochemistry	None	Climatological
Stratospheric aerosols	Time-dependent, prescribed	
Tropospheric aerosols	Climatological	
CO_2	Time-dependent, prescribed	
Trace gases	Climatological	

Table 1. Differences between the atmospheric component of 2011 NCEP CFS and 2021 version of NCEP GFS

From the standpoint of implementation of radiative transfer emulators developed in 2011 into the modern generation of GFS, the most consequential change in the model is the near doubling of the number of vertical layers because it has a direct impact on the size of the input layer of the NN-based emulator. Therefore, we reconfigure GFS v16 to run with 64 layers in the vertical.

Another consequential change in the model appears to be replacement of the Zhao-Carr microphysics (Zhao and Carr, 1993) with the GFDL scheme (Lin et al, 1983; Chen and Lin, 2011; Zhou et al. 2019). Using the latter in combination with 2011 RTTMG emulators resulted in unphysical values of outgoing LW radiation at the top of the atmosphere (TOA) (not shown). Potential explanation is that the change in microphysical parameterization leads to an increase in the number of the model's prognostic variables. Both the spectral and the finite volume dynamical cores include zonal and meridional wind components,

pressure, temperature, water vapor and ozone mixing ratios as prognostic variables. Zhao-Carr microphysics adds only one more prognostic to this list: mixing ratio of total cloud condensate (defined as the sum of cloud water and cloud ice mixing ratios). GFDL microphysics adds 6 prognostic variables: cloud water, cloud ice, rain, snow, and graupel mixing ratios, as well as cloud fraction. The near doubling of the number of prognostic variables from 7 to 12 leads to the proportional increase in dimensionality of the physical phase space of the model. As a result, the set of possible model states in GFS v16 is very different, from a mathematical standpoint, than in the 2011 CFS. Even though the vector of inputs to the LW parameterization remains the same in the new model, it is obtained by mapping from a very different mathematical object, potentially increasing the probability that a given input vector lies outside of the NN's original training data set domain. For our experiments, we replaced the GFDL microphysical parametrization with the Zhao-Carr scheme.

The new hybrid TKE-EDMF planetary boundary layer (PBL) parameterization (Han and Bretherton, 2019) also introduces a new prognostic variable, sub-grid scale turbulent kinetic energy, that was absent in the 2011 version of CFS. Even though w did not see adverse effects stemming from the use of the new PBL scheme in preliminary testing, we replaced it with the original K-profile/EDMF scheme (Han and Pan, 2011) out of abundance of caution.

Concentrations of radiatively active gases are important inputs to radiative transfer schemes, and, more generally, are important parameters of the Earth system. From the standpoint of emulator training, change in these parameters leads to a change in phase space of the host model, potentially necessitating retraining of the emulator to ensure its accuracy and stability. CO_2 concentration values used during training of 2011 emulators ranged from 350 to 380 ppmv between the years 1990 and 2006, respectively. In our current experiments spanning 2018, CO2 concentration was about 409 ppmv, or about 10% higher on average than in the training set.

There were incremental updates and parametric changes to all other components of the suite of physical parameterizations, too numerous to be listed here; in addition, model's software infrastructure was completely overhauled, including new ESMF-based modeling framework, coupler of the dynamical core to the physics package, input/output system, and workflow scripts (For more detail, see the document GFS/GDAS Changes Since 1991, https://www.emc.ncep.noaa.gov/gmb/STATS/html/model_changes.html)

For experiments presented in this paper, we configure GFS v16 to run at C96 horizontal resolution (~100 km) to reduce computational expense of the model. This configuration will be referred to GFS in the following discussion and was used in control runs. We then replaced both modern versions of LW and SW RTTMG parametrizations in GFS with radiative transfer emulators developed in in Krasnopolsky et al (2010). This version of the model will be referred to as hybrid deterministic-statistical GFS, or HGFS. Two 7-month long runs were performed with each model configuration: one initialized on 1/1/2018 and the other one on 7/1/2018, both using 2018 values of radiative forcings, with the instantaneous output saved 3-hourly. Sea

surface temperatures (SSTs) in GFS forecasts are initialized from analysis and exponentially relax to climatology on a 90-day time scale as forecast progresses. First 30 days of each of the two 7 month-long runs were discarded, and remaining 6 months of data in each experiment were combined into a single data set mimicking a 12 month-long run forced by climatological SSTs.

4. Results

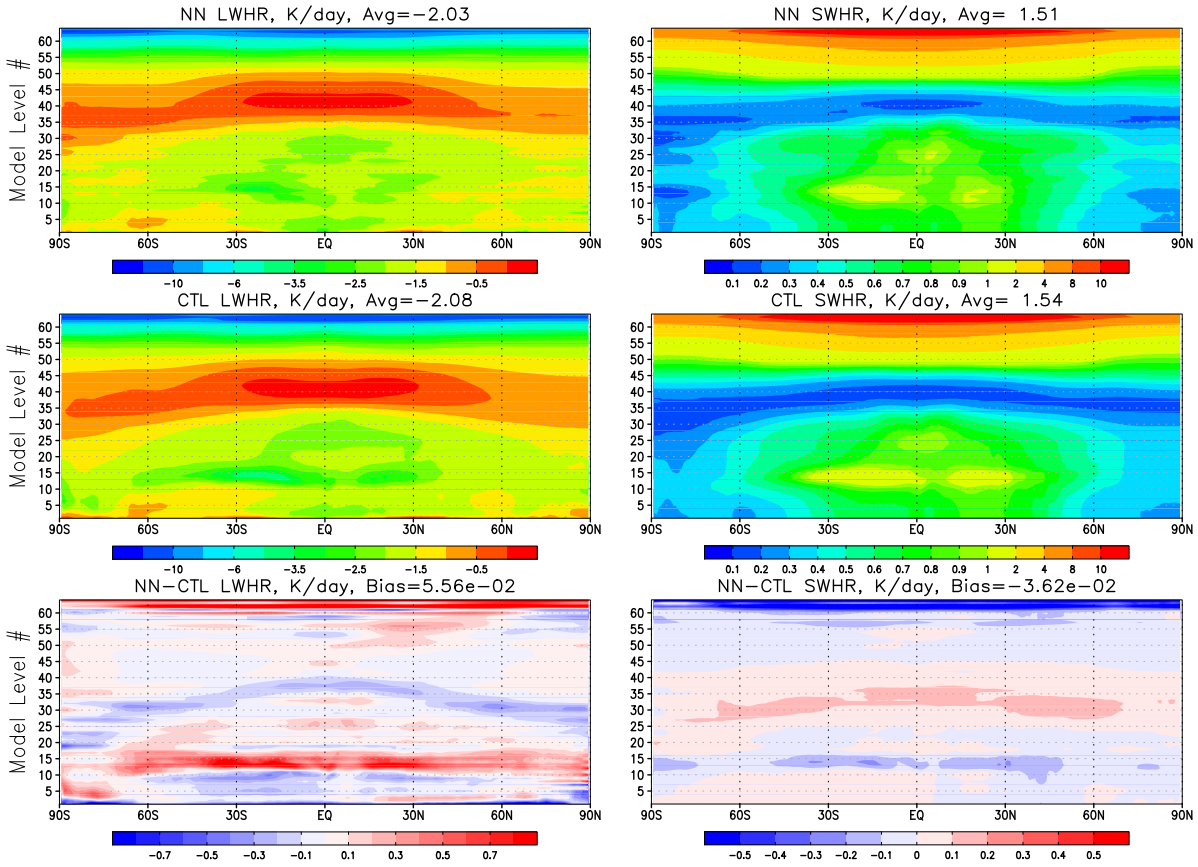

Figure 1. Zonal and time mean over 12-month worth of model output covering 2/1/2018-2/1/2019 for: (a) Long-wave heating rate, K/day; (b) Short-wave heating rate, K/day. Upper row – results produced by HGFS, medium – by GFS, and the lower row the difference (HGFS – GFS). Vertical coordinate shows model level number.

Figure 1 shows zonal and time mean over 12 months worth of model output, covering the period of 2/1/2018-2/1/2019 for LW (left panel) and SW (right panel) heating rates. Global biases are small for both heating rates and constitute about 2-3% of the global mean value. Decrease in LW radiative cooling at the top of tropical and subtropical boundary layer is compensated by the corresponding decrease in SW radiative heating, and consistent with decrease in low cloud cover in these areas (not shown).

Biases in the stratopause may be related to the new parameterizations of O_3 and H_2O photochemistry that were not present in the 2011 version of the model.

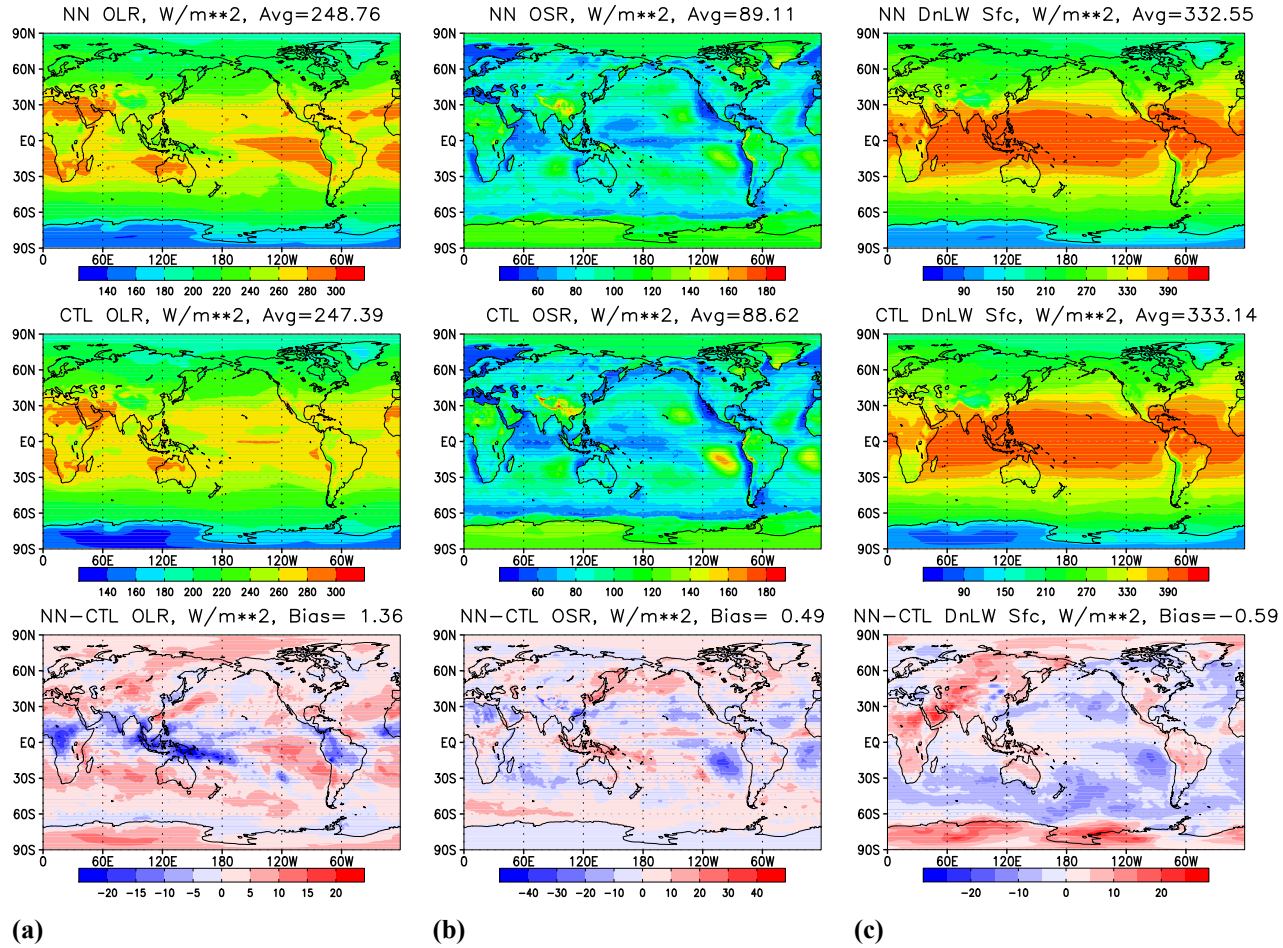

Figure 2. Time mean over 12-month worth of model output covering 2/1/2018-2/1/2019 for: (a) Outgoing LW radiation at the TOA; (b) Outgoing SW radiation at the TOA; (c) Downwelling SW radiation at the surface. Upper row – results produced by HGFS, medium – by GFS, and the lower row the difference (HGFS – GFS).

Figure 2 panel (a) shows outgoing long wave radiation (OLR) at TOA, and panel (b) shows outgoing SW radiation (OSR) at TOA. Global biases are below %1 of the global time mean; however, local biases are more pronounced. Decrease in OLR and increase in OSR over the Maritime continent is consistent with increase in high cloud cover in the region (not shown). Increase in OLR and decrease in OSR in the subtropical areas off western coasts of continents is consistent with decrease in stratocumulus cloud cover (not shown). These changes in cloud cover are also consistent with increase in downwelling SW at the surface in the stratocumulus regions and decrease over the Maritime continent, shown on the panel (c) of Figure 2, with global time mean biases being about 0.2% of the global average.

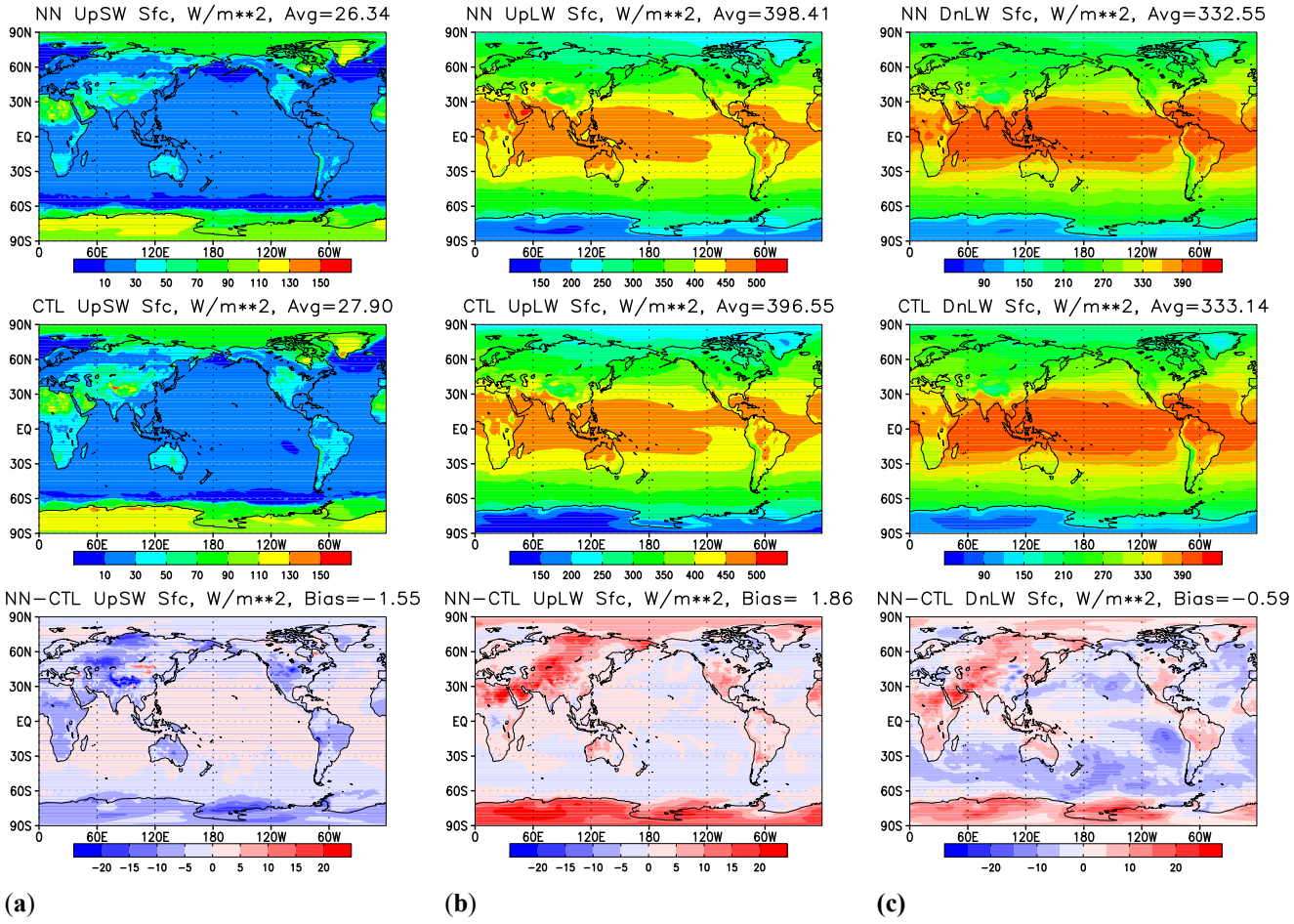

Figure 3. Time means over 12-month worth of model output covering 2/1/2018-2/1/2019 for: (a) Upwelling SW radiation at the surface; (b) Upwelling LW radiation at the surface; (c) Downwelling LW radiation at the surface. Upper row – results produced by HGFS, medium – by GFS, and the lower row the difference (HGFS – GFS).

Figure 3(a) shows upwelling SW radiation flux at the surface. Global mean negative bias is almost 5% of the global average value, with negative biases prevalent over continents and extratropical oceans and positive biases over tropical oceans. Upwelling LW at the surface (Figure 3(b)) is biased high by about 0.5% of the global mean value, with positive biases over most of the continents, polar areas and most of tropical oceans, and negative biases in the midlatitude oceans, northern Canada and Alaska, as well as Barents and Norwegian Seas. Downwelling LW at the surface (Figure 3(b)) is biased low globally by approximately 0.5%. Table 2 summarizes time and global mean biases for the heating rates and radiative fluxes predicted by the emulators.

Figure 4 shows time series of a 10-day running mean of globally averaged LW and SW fluxes at the surface and the TOA generated by HGFS (black curves) and GFS (green curves) for the last 6 months (2/1/18-8/1/18) of a 7-month long run initialized on 1/1/2018. Time series of the same quantities for the run initialized on 7/1/18 exhibit similar properties and

285 therefore are not shown. Magnitudes and signs of biases of each emulator predicted variable are consistent with their time and globally averaged values shown in Table 2. HGFS run captures seasonal cycle and the amplitude of seasonal and sub-seasonal variability reasonably well. As to be expected from long term free running experiments with a GCM, details of individual weather systems differ between the two runs even when considered through the lens of a 10-day running mean. This is manifested most starkly in short-wave fluxes leaving the atmosphere, outgoing SW at TOA (Figure 4b) and downwelling SW
290 at the surface (Figure 4d), which are very sensitive to the instantaneous cloud distributions.

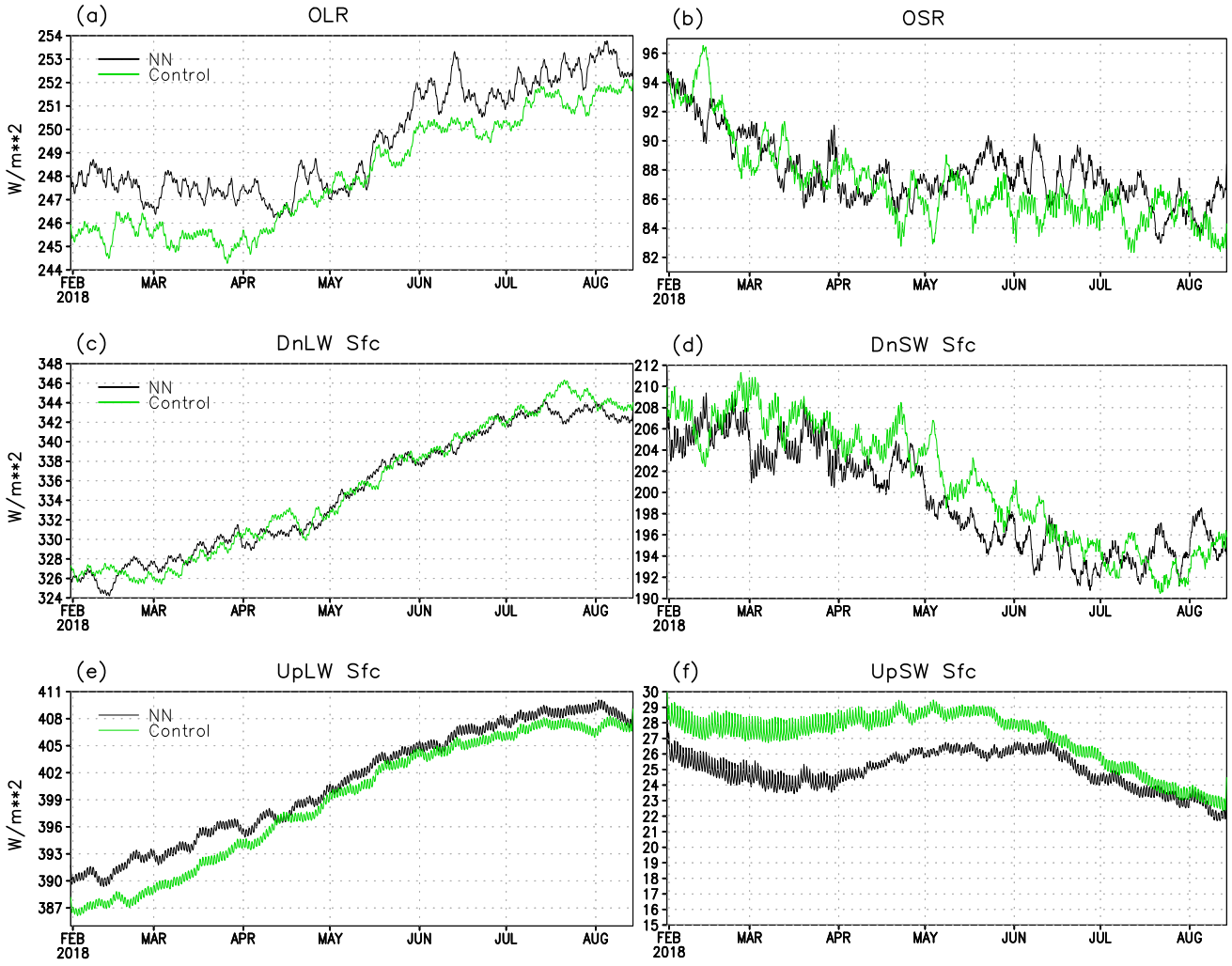

Figure 4. Time series of a running 10-day mean covering 2/1/2018-8/1/2018 for: (a) Outgoing LW at TOA; (b) Outgoing SW at TOA; (c) Downwelling LW radiation at the surface ; (d) Downwelling SW radiation at the surface; (e) Upwelling LW radiation at the surface; (f) Upwelling SW radiation at the surface. Black curves – results produced by HGFS, green – by GFS.

295

Variable	LWHR	SWHR	OLR	OSR	LW Up Sfc	LW Dn Sfc	SW Up Sfc	SW Dn Sfc
Bias	5.56e-2 K/day	-3.62e-2 K/day	1.36 W/m**2	0.49 W/m**2	-0.59 W/m**2	-1.55 W/m**2	1.86 W/m**2	-0.59 W/m**2

Table 2. Time means and global mean biases in radiative heating rates and fluxes over 12-month worth of model output covering 2/1/2018-2/1/2019.

5. Discussion

What could be the factors contributing to stability of the emulators presented in this paper? In the following, we highlight and discuss aspects of the machine learning technique choice (shallow vs deep neural network, activation function selection) and training set design that distinguish the emulators developed in Krasnopolsky et al (2010).

5.1 Shallow vs. deep neural networks: complexity and nonlinearity

Application of shallow NNs (SNNs) to the problem of mapping approximation has thorough theoretical support. The universal approximation theorem proves that an SNN is a generic and universal tool for approximating any continuous and almost continuous mappings under very broad assumptions and for a wide class of activation functions (e.g., Hornik et al, 1990; Hornik, 1991). Similarly broad results for deep NNs (DNNs) do not exist as of yet (Vapnik, 2019), however specific combinations of DNN architectures and activation functions have theoretical support (e.g., Leshno et al, 1993; Lu et al, 2017; Elbrachter et al 2020). Until there is a universal theory, it has been suggested to consider DNN a heuristic approach since, in general, "from the theoretical point of view, deep network cannot guarantee a solution of any selection problem that constitute complete learning problem" (Vapnik, 2019). These considerations are important to keep in mind when selecting NN architecture for the emulation of model physics or its components

Next, we compare some properties of DNNs and SNNs to further emphasize their differences and to point out some properties of DNNs that may lead to instabilities in deterministic models coupled to DNN-based model components.

To avoid overfitting and instability, complexity, and nonlinearity of approximating/emulating NN should not exceed complexity and nonlinearity of the mapping to be approximated. A measure of the SNN complexity can be written as (see below for explanation),

$$\mathbb{C}_{SNN} = k \cdot (n + m + 1) + m \qquad (2)$$

where n and m are the numbers of the SNN inputs and outputs, and k is the number of neurons in a single hidden layer. The complexity of the SNN (Equation 2) increases linearly with the number of neurons in the hidden layer, k. For given numbers of inputs and outputs there is only one SNN architecture/configuration with a specified complexity $\mathbb{C}_{SNN}$.

For the DNN complexity, a similar measure of complexity can be written as (again, see below for explanation),

$$\mathbb{C}_{DNN} = \sum_{i=0}^{K} k_{i+1}(k_i + 1) \qquad (3)$$

where k_i is the number of neurons in the layer i ($i = 0$ and $i = K$ correspond to the input and output layers, respectively). The complexity of the DNN (Equation 3) increases geometrically with the increasing number of layers, K.

Both $\mathbb{C}_{SNN}$ and $\mathbb{C}_{DNN}$ are simply the number of parameters of the NN that are trained or fit during SNN/DNN training. While there exists a one to one correspondence between the SNN complexity, $\mathbb{C}_{SNN}$, and the SNN architecture, given the fixed number of neurons in the input and output layers, correspondence between the DNN complexity, $\mathbb{C}_{DNN,}$ and the DNN architecture is multivalued: many different DNN architectures/configurations have the same complexity $\mathbb{C}_{DNN}$ given the same size of input and output layers. Overall, controlling complexity of DNNs is more difficult than controlling complexity of SNN.

For an SNN given by the expression

$$y_j = b_j^1 + \sum_{i=1}^{k} a_{ji}^1 \cdot t_i \ , \ \ j = 1, \cdots, m \ ,$$

where n, m, and k are the same as in Equation (2), nonlinearity increases arithmetically or linearly with addition of each new hidden neuron, $t_i = \phi(b_i^0 + \sum_{s=1}^{n} a_{is}^0 \cdot \ x_s)$, to the single hidden layer of the NN.

For a DNN, symbolically written as

$$Y = \ X^{n+1} = B^n + A^n \cdot \phi\left(B^{n-1} + \ A^{n-1} \cdot \phi\left(B^{n-2} + \ A^{n-2} \cdot \phi\left(B^{n-3} + \cdots \phi(B^0 + \ A^0 \cdot X)\right)\right)\right),$$

each new hidden layer/neuron introduces additional nonlinearity on top of nonlinearities of the previous hidden layers; thus, the nonlinearity of the DNN increases geometrically with addition of new hidden layers, much quicker than the nonlinearity of the SNN. Thus, controlling nonlinearity of DNNs is more difficult than controlling nonlinearity of SNNs. The higher the nonlinearity of the model the more unstable and unpredictable generalization is (especially nonlinear extrapolation that is an ill-posed problem).

DNNs are a very powerful and flexible technique that is extensively used for emulation of model physics and its components (Kasim et al, 2020). Discussion of its limitations can be found in Thompson et al (2020). The arguments listed here are intended to point out possible sources of instability of DNNs in the models and the need for careful handling this very sensitive tool.

5.2 Preparation of training sets

Specifics of training set design may impact stability of the NN as well. We would like to point out a few aspects of training set preparation that, in our experience, are of relevance to development of robust ML-based components of geophysical models.

A general rule of thumb when it comes to fitting statistical models to data is that the number of records in the training set should be at least as large as the number of the model parameters, or, in the context of current discussion, as the NN complexity introduced in Section 5.1. As a consequence, NNs of larger complexity require larger training sets to approximate a given mapping. To use DNN as an example, as the complexity of DNN, $\mathbb{C}_{DNN}$ (Equation 3), increases geometrically with the number of DNN layers, so does the amount of data required for the DNN training (Thompson et al, 2020)

We also find that comprehensiveness of the training set is an important contributing factor to the generalization capability of the NN. In the context of application at hand, comprehensiveness of the training set means that it should encompass as much of complexity of the underlying physical system as permitted by the numerical model that hosts the NN. In practice, it translates into sampling diurnal, seasonal, and annual variability, as well as states introduced by boundary conditions, e.g greenhouse gas and aerosol concentrations, realistic orography, and surface state. Inclusion of events of special interest, e.g hurricanes, snow storms, droughts, extreme precipitation events etc., is beneficial as well.

Care should be taken of proper sampling of the training data. For example, saving the training data set on a Gaussian longitude-latitude grid will result in overrepresentation of polar areas, and data must be resampled to get more uniform representation over the globe.

Purging and normalization of inputs and outputs are important. Constant inputs and outputs must be removed: from the standpoint of mapping emulation, constants carry no information about the input-to-output relation; however, with incorrect normalization, they may become a source of noise during training. Normalization of inputs and outputs strongly affects NN training. More specifically to the present application, if some inputs or outputs of an NN are vertical profiles of a physical variable, as is common in geophysical models, the profiles should be normalized as a whole, as opposed to as a collection of independent variables, for the NN to better capture correlations and dependencies between the levels of the profile (Krasnopolsky, 2013).

5.3 Continuously vs. not continuously differentiable activation functions

Universal approximation theorem for SNNs is satisfied for a wide class of bounded, non-linear AFs. Note, that many popular AFs used in DNN applications, e.g. variants of ReLU, do not belong to this class. However, for a specific problem of mapping approximation, it may be useful to consider additional restrictions on AFs.

If the AF is almost continuous, or, in other words, has only finite discontinuities (e.g, step function), the first derivative (Jacobian) of the NN using this AF will be singular. If the AF is not continuously differentiable (e.g, ReLU), its first derivative will not be continuous (will have finite discontinuities), and so will be the NN Jacobian. Using a non-continuously

differentiable NN as a model component may lead to instability, especially if the Jacobian of this component is calculated in the model. Using gradient-based optimization algorithms for training such NNs may be challenging due to discontinuities in gradients.

If the AF is monotonic, the error surface associated with a single-layer model is guaranteed to be convex, simplifying the training process (Wu, 2009). When AF approximates identity function near the origin (i.e. $\phi(0) = 0., \phi'(0) = 1$, and ϕ' is continuous at 0), the neural network will learn efficiently when its weights are initialized with small random values. When the activation function does not approximate identity near the origin, special care must be used when initializing the weights (Susillo and Abbott, 2014).

It is noteworthy that the sigmoid and hyperbolic tangent AFs, popular in SNN applications, meet all aforementioned criteria. Additionally, in the context of emulation of model physics parameterizations, these AFs provide one of the lowest training losses, as compared to other AFs (Chantry et al, 2020).

6. Conclusions

One of the major challenges in development of ML/AI-based parameterizations for multi-dimensional non-linear forward environmental models is ensuring stability of the coupling between deterministic and statistical components. This problem is particularly acute for neural network-based parameterizations since, in theory, generalization to out-of-sample data is not guaranteed, and, in practice, previously unseen inputs may lead to unphysical outputs of the NN-based parameterization, often destabilizing the hybrid model even in idealized simulations.

Shallow NN-based emulators of radiative transfer parameterizations developed almost a decade ago for a state-of-the-art GCM are stable with respect to substantial structural and parametric change in the host model: when used in two seven months long experiments with the new model, they not only remain stable, but generate realistic output. Two types of modifications of the host model that NN emulators cannot tolerate are the change of the model vertical resolution, and the change in number of model prognostic variables due to, in both cases, alteration of dimensionality of phase space of the mapping (parameterization) and of the emulating NN. After the changes of this nature are introduced into the host model NN emulators must be retrained.

We conjecture that careful control of complexity and nonlinearity of an AI/ML model component, along with comprehensiveness and realism of its training data set, are important factors contributing both to the component's generalization capability and to stability of the model hosting it.

Code availability. The source code of NCEP GFS atmospheric model component using the full physics NN emulator, including the file with trained NN coefficients, is available in the GitHub repository https://github.com/AlexBelochitski-NOAA/fv3atm_old_radiation_nn_emulator, and is also archived on Zenodo: doi: 10.5281/zenodo.4663160

Author contributions. Conceptualization, A.B. and V.K.; methodology, A.B. and V.K.; data for training, A.B.; NN training V.K.; NN validation, A.B.; analysis of results A.B. and V.K.; writing A.B. and V.K. All authors have read and agreed to the published version of the manuscript.

Competing interests. The authors declare no competing interests.

Acknowledgments. The authors would like to thank Dr. Edoardo Bucchignani and the anonymous reviewer for their thoughtful and incisive comments that helped to improve the manuscript. We also thank Drs. Ruiyu Sun and Jun Wang for valuable help with practical use of NCEP GFS and for useful discussions and consultations. We thank Drs. Jack Kain, Fanglin Yang, and Vijay Tallapragada for their support.

References

Belochitski, A., Binev, P., DeVore, R., Fox-Rabinovitz, M., Krasnopolsky, V., and Lamby, P.: Tree approximation of the long wave radiation parameterization in the NCAR CAM global climate model. *Journal of Computational and Applied Mathematics*, 236, 447–460. https://doi.org/10.1016/j.cam.2011.07.013, 2011

Belochitski, A.A., and Krasnopolsky, V.: Stable Emulation of an Entire Suite of Model Physics in a State-of-the-Art GCM using a Neural Network, https://arxiv.org/ftp/arxiv/papers/2103/2103.07028.pdf

Boukabara S-A., Krasnopolsky V., Stewart J. Q., Maddy E. S., Shahroudi N., and Hoffman R. N.: Leveraging modern artificial intelligence for remote sensing and NWP: Benefits and Challenges. BAMS, 100, pp. ES473–ES491; DOI: https://doi.org/10.1175/BAMS-D-18-0324.1, 2019.

Brenowitz, N. D., Beucler, T., Pritchard, M., and Bretherton, C. S.: Interpreting and Stabilizing Machine-Learning Parametrizations of Convection, Journal of the Atmospheric Sciences, 77(12), 4357-4375, https://doi.org/10.1175/JAS-D-20-0082.1, 2020

Chantry M., Hatfield S., Dueben P., Polichtchouk I., and Palmer T.: Machine learning emulation of gravity wave drag in numerical weather forecasting, https://arxiv.org/abs/2101.08195, 2020.

Chen, J.-H., & Lin S.-J.: The remarkable predictability of inter-annual variability of Atlantic hurricanes during the past decade. Geophysical Research Letters, 38, L11804, doi:10.1029/2011GL047629. 2011

Chevallier, F., Cheruy F., Scott N. A., and Chedin A.: An neural network approach for a fast and accurate computation of longwave radiative budget, *J. Appl. Meteorol.*, 37, 1385 – 1397, https://doi.org/10.1175/1520-0450(1998)037%3C1385:ANNAFA%3E2.0.CO;2, 1998

Chevallier, F., Morcrette, J.-J., Chéruy, F., and Scott, N. A. : Use of a neural-network-based longwave radiative transfer scheme in the ECMWF atmospheric model. Quarterly Journal of the Royal Meteorological Society, 126, 761-776, https://doi.org/10.1002/qj.49712656318, 2000.

Elbrachter D., Perekrestenko D., Grohs P., and Bölcskei H.: Deep Neural Network Approximation Theory. arXiv:1901.02220, 2020

Edwards, J. M., and Slingo A.: Studies with a flexible new radiation code: 1. Choosing a configuration for a large-scale model. *Q. J. R. Meteorol. Soc*., 122, 689–719, https://doi.org/10.1002/qj.49712253107, 1996

Fu, Q., and Liou, K. N.: On the correlated k-distribution method for radiative transfer in nonhomogeneous atmospheres. *Journal of the Atmospheric Sciences*, 49(22), 2139–2156, https://doi.org/10.1175/1520-0469(1992)049%3C2139:OTCDMF%3E2.0.CO;2, 1992

Han, J., and Bretherton, C. S.: TKE-based moist eddy-diffusivity mass-flux (EDMF) parameterization for vertical turbulent mixing. Weather and Forecasting, 34, 869–886. https://doi.org/10.1175/WAF-D-18-0146.1, 2019

Han, J., and Pan H.-L.:. Revision of convection and vertical diffusion schemes in the NCEP global forecast system. Weather and Forecasting, 26, 520–533. https://doi.org/10.1175/WAF-D-10-05038.1., 2011

Hogan, R, Ahlgrimm, M, Balsamo, G, Beljaars, A, Berrisford, P, Bozzo, A, et al: Radiation in numerical weather prediction. *ECMWF Technical Memorandum.,* doi: 10.21957/2bd5dkj8x*,* 2017

Hogan, R.., and Bozzo A.: Mitigating errors in surface temperature forecasts using approximate radiation updates. *J. Adv. Model. Earth. Syst.,* **7,** 836–853, https://doi.org/10.1002/2015MS000455, 2015

Hornik, K., Stinchcombe, M., and White, H.: Universal approximation of an unknown mapping and its derivatives using multilayer feedforward network. Neural Networks, 3, 551-560, https://doi.org/10.1016/0893-6080(90)90005-6, 1990

Hornik, K.: Approximation Capabilities of Multilayer Feedforward Network. Neural Networks, 4, 251-257. https://doi.org/10.1016/0893-6080(91)90009-T, 1991

Janiskova M., Mahfouf J-F., Morcrette J-J., and Chevallier F.: Linearized radiation and cloud schemes in the ECMWF model: Development and evaluation. *Quart. J. Roy. Meteor. Soc*., 128 , 1505–1528, https://doi.org/10.1002/qj.200212858306, 2002

Kain, J. S., Moorthi, S., Yang, F., Yang, R., Wei, H., Wu, Y., Hou, Y.-T., Lin, H.-M., Yudin, V. A., Alpert, J. C., Tallapragada, V., and Sun R.: Advances in model physics for the next implementation of the GFS (GFSv16). AMS Annual Meeting, Boston, MA, 6A.3., 2020

Kashinath K et al.: Physics-informed machine learning: case studies for weather and climate modelling. Phil. Trans. R. Soc. A 379: 20200093. https://doi.org/10.1098/rsta.2020.0093, 2021

Kasim M. F. et al.:. Building high accuracy emulators for scientific simulations with deep neural architecture search https://arxiv.org/pdf/2001.08055.pdf, 2020

Krasnopolsky, V.: The Application of Neural Networks in the Earth System Sciences. Neural Network Emulations for Complex Multidimensional Mappings. Atmospheric and Oceanic Science Library. (Vol. 46). Dordrecht, Heidelberg, New York, London: Springer. DOI 10.1007/978-94-007-6073-8, 2013.

Krasnopolsky, V., Belochitski, A. A., Hou, Y.-T., Lord S., and Yang F.: Accurate and fast neural network emulations of long and short-wave radiation for the NCEP Global Forecast System model., NCEP Office Note, 471, https://repository.library.noaa.gov/view/noaa/6951, 2012

Krasnopolsky, V., Fox-Rabinovitz, M. S., and Belochitski, A. A.: Decadal climate simulations using accurate and fast neural network emulation of full, longwave and shortwave, radiation. *Mon. Wea. Rev.*, 136, 3683–3695, https://doi.org/10.1175/2008MWR2385.1, 2008

Krasnopolsky, V. M., Fox-Rabinovitz, M. S., and Belochitski, A. A.: Using Ensemble of Neural Networks to Learn Stochastic Convection Parameterization for Climate and Numerical Weather Prediction Models from Data Simulated by Cloud Resolving Model, Advances in Artificial Neural Systems, 2013, Article ID 485913, doi:10.1155/2013/485913, 2013

Krasnopolsky, V. M., Fox-Rabinovitz, M. S., Hou, Y.-T., Lord, S. J., and Belochitski, A. A.: Accurate and Fast Neural Network Emulations of Model Radiation for the NCEP Coupled Climate Forecast System: Climate Simulations and Seasonal Predictions, Monthly Weather Review, 138, 1822-1842. doi: 10.1175/2009MWR3149.1, 2010.

Krasnopolsky, V. M., Fox-Rabinovitz, M.S., Tolman, H.L., and Belochitski, A. A.: Neural network approach for robust and fast calculation of physical processes in numerical environmental models: Compound parameterization with a quality control of larger errors, Neural Networks, 21, 535–543; doi:10.1016/j.neunet.2007.12.019, 2008

Lagerquist, R., Turner ,D. D., Ebert-Uphoff, I., Hagerty, V., Kumler, C., and Stewart, J.: Deep Learning for Parameterization of Shortwave Radiative Transfer, *20th Conference on Artificial Intelligence for Environmental Science*, 101st AMS Annual Meeting, Virtual, 10-15 January 2021, 6.1, 2021

Leshno M., Lin V. Ya., Pinkus A., and Schocken S.: Multilayer Feedforward Networks With a Nonpolynomial Activation Function Can Approximate Any Function, Neural Networks, Vol. 6, pp. 861-867, https://doi.org/10.1016/S0893-6080(05)80131-5, 1993

Lin, Y.-L., Farley, R. D., and Orville, H. D.: Bulk parameterization of the snow field in a cloud model. Journal of Climate and Applied Meteorology, 22, 1065–1092. https://doi.org/10.1175/1520-0450(1983)022%3C1065:BPOTSF%3E2.0.CO;2, 1983

Lu, Z, Pu, H., Wang, F, Hu Z; Wang, L: The Expressive Power of Neural Networks: A View from the Width. Advances in Neural Information Processing Systems 30. Curran Associates, Inc.: 6231–6239. arXiv:1709.02540, 2017

Marshak, A., and Davis, A. B. (Eds.): 3D Radiative Transfer in Cloudy Atmospheres. Springer, 686 pp., 2005

Morcrette, J-J., Mozdzynski, G. , and Leutbecher, M.: A reduced radiation grid for the ECMWF Integrated Forecasting System. *Mon. Wea. Rev*., 136 , 4760–4772, https://doi.org/10.1175/2008MWR2590.1, 2008

Mlawer, E. J., Taubman S. J., Brown P. D., Iacono M. J., and Clough S. A.: Radiative transfer for inhomogeneous atmospheres: RRTM, a validated correlated-k model for the longwave. Journal of Geophysical Research: Atmospheres, 102, 16663–16682. https://doi.org/10.1029/97JD00237, 1997

Oreopoulos, L., Mlawer, E., Delamere, J., Shippert, T., Cole, J., Fomin, B., et al.: The continual intercomparison of radiation codes: Results from phase I. *Journal of Geophysical Research*, 117, D06118. https://doi.org/10.1063/1.3117076, 2012

Ott J, Pritchard M, Best N, Linstead E, Curcic M, and Baldi P.: A Fortran-Keras deep learning bridge for scientific computing, Scientific Programming, vol. 2020, Article ID 8888811, https://doi.org/10.1155/2020/8888811, 2020

Pal, A., Mahajan, S., and Norman, M. R.: Using deep neural networks as cost-effective surrogate models for Super-Parameterized E3SM radiative transfer. *Geophysical Research Letters*, 46, 6069– 6079, https://doi.org/10.1029/2018GL081646, 2019

Pincus, R., Barker H. W., and Morcrette J.-J.: A fast, flexible, approximate technique for computing radiative transfer in inhomogeneous clouds. *J. Geophys. Res. Atmos*., 108, 4376, doi:10.1029/2002JD003322, 2003

Pincus, R., and Stevens, B.: Monte Carlo Spectral Integration: a Consistent Approximation for Radiative Transfer in Large Eddy Simulations, *J. Adv. Model. Earth Syst*., 1, 1, doi:10.3894/JAMES.2009.1.1., 2009

Pincus, R., and Stevens, B.: Paths to accuracy for radiation parameterizations in atmospheric models, *J. Adv. Model. Earth Syst.,* 5, 225– 233, https://doi.org/10.1002/jame.20027, 2013

Rasp, S.: 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, 2185–2020. doi:10.5194/gmd-13-2185-2020, 2020

Roh, S., and Song, H.-J.: Evaluation of neural network emulations for radiation parameterization in cloud resolving model. *Geophysical Research Letters*, 47, e2020GL089444, https://doi.org/10.1029/2020GL089444, 2020

Sussillo, D., and Abbott, L. F.: Random Walk Initialization for Training Very Deep Feedforward Networks, arXiv:1412.6558, 2014

Thompson N.C., Greenewald K., Lee K., and Manso G.F.: The Computational Limits of Deep Learning, https://arxiv.org/abs/2007.05558, 2020

Turner, D. D., Tobin, D. C., Clough, S. A., Brown, P. D., Ellingson, R. G., Mlawer, E. J., et al.: The QME AERI LBLRTM: A closure experiment for downwelling high spectral resolution infrared radiance. *Journal of the Atmospheric Sciences*, 61(22), 2657–2675, https://doi.org/10.1175/JAS3300.1, 2004

Ukkonen, P., Pincus, R., Hogan, R. J., Nielsen, K. P., and Kaas, E.: Accelerating radiation computations for dynamical models with targeted machine learning and code optimization. *Journal of Advances in Modeling Earth Systems*, 12, e2020MS002226, https://doi.org/10.1029/2020MS002226, 2020

Vapnik, V. N.: Complete Statistical Theory of Learning, Automation and Remote Control, 80, 1949–1975. https://doi.org/10.1134/S000511791911002X, 2019

Veerman MA, Pincus R, Stoffer R, van Leeuwen CM, Podareanu D, and van Heerwaarden CC: Predicting atmospheric optical properties for radiative transfer computations using neural networks. *Phil. Trans. R. Soc. A* 379: 20200095, https://doi.org/10.1098/rsta.2020.0095, 2021

Wu, H: Global stability analysis of a general class of discontinuous neural networks with linear growth activation functions, Information Sciences. 179 (19): 3432–3441. doi:10.1016/j.ins.2009.06.006, 2009.

Yuval J, and O'Gorman P.: Stable machine-learning parameterization of subgrid processes for climate modeling at a range of resolutions. Nat. Commun. 11, 3295. doi:10.1038/s41467-020-17142-3, 2020

Yuval, J., O'Gorman, P. A., and Hill, C. N.: Use of neural networks for stable, accurate and physically consistent parameterization of subgrid atmospheric processes with good performance at reduced precision. Geophysical Research Letters, 48, e2020GL091363, https://doi.org/10.1029/2020GL091363, 2021

Zhao, Q., and Carr F. H.: A Prognostic Cloud Scheme for Operational NWP Models. Monthly Weather Review,125, 1931–1953. https://doi.org/10.1175/1520-0493(1997)125%3C1931:APCSFO%3E2.0.CO;2, 1997

Zhou, L., Lin, S., Chen, J., Harris, L. M., Chen, X., & Rees, S. L.: Toward Convective-Scale Prediction within the Next Generation Global Prediction System, *Bulletin of the American Meteorological Society*, *100*(7), 1225-124. https://doi.org/10.1175/BAMS-D-17-0246.1, 2019