









## RESEARCH ARTICLE

10.1029/2023MS003800

# Efficient Super-Resolution of Near-Surface Climate Modeling Using the Fourier Neural Operator

 Peishi Jiang<sup>1</sup> , Zhao Yang<sup>1</sup> , Jiali Wang<sup>2</sup> , Chenfu Huang<sup>3</sup>, Pengfei Xue<sup>2,3</sup> , T. C. Chakraborty<sup>1</sup>, Xingyuan Chen<sup>1</sup> , and Yun Qian<sup>1</sup> 
<sup>1</sup>Pacific Northwest National Laboratory, Richland, WA, USA, <sup>2</sup>Argonne National Laboratory, Lemont, IL, USA, <sup>3</sup>Michigan Technological University, Houghton, MI, USA
**Special Section:**

Machine learning application to Earth system modeling

**Key Points:**

- We used the Fourier neural operator (FNO) to perform zero-shot super-resolution on near-surface heat-related estimates
- Incorporating a physics-constrained loss based on the Clausius–Clapeyron equation improves the emulation performance of the trained FNO
- Trained on a 4-km Weather Research and Forecasting simulation, the FNO generates a 1-km emulation that captures fine-grained climate features induced by topography

**Supporting Information:**

Supporting Information may be found in the online version of this article.

**Correspondence to:**P. Jiang,  
[peishi.jiang@pnnl.gov](mailto:peishi.jiang@pnnl.gov)**Citation:**

Jiang, P., Yang, Z., Wang, J., Huang, C., Xue, P., Chakraborty, T. C., et al. (2023). Efficient super-resolution of near-surface climate modeling using the Fourier neural operator. *Journal of Advances in Modeling Earth Systems*, 15, e2023MS003800. <https://doi.org/10.1029/2023MS003800>

Received 11 MAY 2023

Accepted 9 JUL 2023

**Abstract** Downscaling methods are critical in efficiently generating high-resolution atmospheric data. However, state-of-the-art statistical or dynamical downscaling techniques either suffer from the high computational cost of running a physical model or require high-resolution data to develop a downscaling tool. Here, we demonstrate a recently proposed *zero-shot super-resolution* method, the Fourier neural operator (FNO), to efficiently perform downscaling without the need for high-resolution data. Because the FNO learns dynamics in Fourier space, FNO is a resolution-invariant emulator; it can be trained at a coarse resolution and produces emulation at any high resolution. We applied FNO to downscale a 4-km resolution Weather Research and Forecasting (WRF) Model simulation of near-surface heat-related variables over the Great Lakes region. The FNO is driven by the atmospheric forcings and topographic features used in the WRF model at the same resolution. We incorporated a physics-constrained loss in FNO by using the Clausius–Clapeyron relation to better constrain the relations among the emulated states. Trained on merely 600 WRF snapshots at 4-km resolution, the FNO shows comparable performance with a widely-used convolutional network, U-Net, achieving averaged *modified Kling–Gupta Efficiency* of 0.88 and 0.94 on the test data set for temperature and pressure, respectively. We then employed the FNO to produce 1-km emulations to reproduce the fine climate features. Further, by taking the WRF simulation as ground truth, we show consistent performances at the two resolutions, suggesting the reliability of FNO in producing high-resolution dynamics. Our study demonstrates the potential of using FNO for zero-shot super-resolution in generating first-order estimation on atmospheric modeling.

**Plain Language Summary** High-resolution climate estimates are crucial for applications like examining heat exposure in populated areas. Downscaling from lower-resolution estimates is one popular technique for obtaining these estimates. However, developing downscaling tools usually requires expensive computation or difficult-to-obtain data; they rely on either numerically solving a physical model or using a pair of existing high- and low-resolution data. To address this issue, we demonstrated a novel deep learning model that can generate high-resolution results from only low-resolution data. Our study successfully applied this method to downscale near-surface heat-related variables of the Weather Research and Forecasting simulation to 1-km resolution over the Great Lakes region.

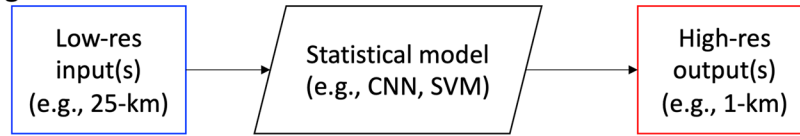
## 1. Introduction

There is an increasing need for high-fidelity atmospheric data at or below kilometer-scale in Earth science (Lucas-Picher et al., 2021). These high-resolution data, oftentimes serving as meteorological forcings, are critical to a variety of regional studies, such as climate risk assessments (Orr et al., 2021), surface heat-exposure analyses (Zakšek & Oštir, 2012; Y. Jiang et al., 2015; X. Pan et al., 2018), hydrological modeling (Fowler et al., 2007; Teutschbein & Seibert, 2010), etc. Although advances in observation techniques and rapidly increasing computing power enable more access to a variety of observed data or reanalysis products (Abatzoglou et al., 2018; Karger et al., 2017; Muñoz Sabater et al., 2021), these high-resolution data are not readily available everywhere.

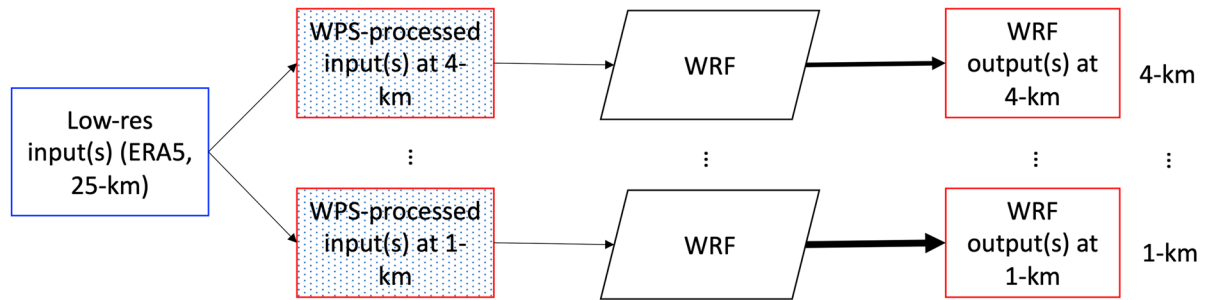
Downscaling is a procedure that generates high-resolution data from a low-resolution reanalysis product or global climate model using either statistical or dynamical methods. Shown in Figure 1a, statistical downscaling leverages statistical or machine learning tools to learn the mappings between existing simulations or observations on low and high resolutions and then produce high-resolution estimates from low-resolution products. Dynamical downscaling (Giorgi, 2019), on the other hand, generate refined estimates from global reanalysis

© 2023 UChicago Argonne, LLC and Battelle Memorial Institution. This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs License](https://creativecommons.org/licenses/by-nc-nd/4.0/), which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

(a) Statistical downscaling

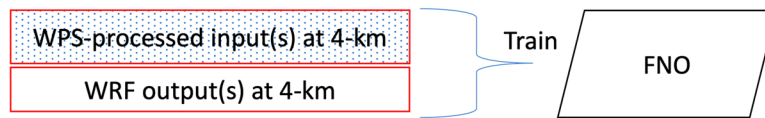


(b) Dynamical downscaling (e.g., through physics-based WRF)

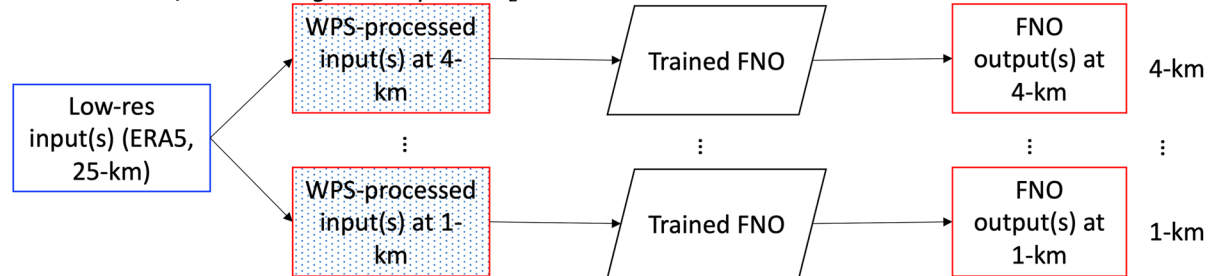


(c) Zero-shot super-resolution (e.g., through FNO)

FNO Training on a given-res (e.g., 4-km) WRF at time period  $T_1$



FNO emulation/downscaling at time period  $T_2$



**Figure 1.** Three types of downscaling techniques: (a) statistical downscaling that directly maps from low- to high-resolution estimations based on a statistical or machine learning model (e.g., convolutional neural net [CNN]); (b) dynamical downscaling that generates high-resolution estimates from low-resolution products (e.g., ERA5) by leveraging a physics-based model such as WRF; and (c) zero-shot super-resolution that develops a resolution-invariant emulator through the Fourier neural operator (FNO) on training data with a relatively coarse resolution (e.g., generated by Weather Research and Forecasting [WRF] simulation) and performs the emulation on a high-resolution in a different time period (WRF Preprocessing System [WPS] is used to generate the input forcings at both resolutions).

products by using a physical model, such as the Weather Research and Forecasting model (WRF) (Skamarock et al., 2019), which has been extensively employed in performing vulnerability, impact, and adaptation assessments (Coppola et al., 2018; Im et al., 2017; Liu et al., 2017; Rockel & Woth, 2007; Teutschbein & Seibert, 2010; Tobin et al., 2016; Xue et al., 2020). In contrast to its statistical counterpart, dynamical downscaling is able to generate simulations at any resolution using process-based rather than data-driven constraints; however, it suffers from high computational cost (Figure 1b).

The computational efficiency of statistical downscaling has enabled numerous successes based on a variety of data-driven methods, including linear regression (Von Storch, 1999), support vector machine (Le Roux et al., 2018; Pour et al., 2018), random forest (Chen et al., 2021; Hutengs & Vohland, 2016), etc. Recently, deep learning (DL) for super-resolution (Z. Wang et al., 2021) has proven to be a better downscaling technique because of the generally better capability of DL in learning nonlinear features than traditional machine learning methods (Goodfellow et al., 2016). Convolutional neural networks (CNNs), which learn spatial structures by using convolutional kernels (Atlas et al., 1987), have been extensively employed to obtain super-resolution estimates from Earth system datasets, including precipitation (B. Pan et al., 2019; Shi, 2020; Kumar et al., 2021), air temperature (Baño Medina et al., 2020; Huang, 2020), and soil moisture (SM) (Xu et al., 2021). Researchers have further embedded CNNs in generative adversarial networks, which show improved performance over CNN-only methods in a variety of studies that downscaled precipitation (J. Wang et al., 2021; Watson et al., 2020) and wind field (Manepalli et al., 2020; Stengel et al., 2020).

Nevertheless, most conventional statistical downscaling methods are limited to the spatial resolution of the training data set and require sufficient high-resolution data for developing the downscaling tool. First, to the best of our knowledge, merely a fixed high resolution can be achieved by the current downscaling applications in atmospheric science. For the needs of a different high-resolution product, the mapping between low and high resolutions has to be redeveloped, and thus new training data are required. This is much less flexible than dynamical downscaling, which can resolve any downscaled resolution through physical models. Second, high-resolution data are required to train the statistical or machine learning-based mapping between low and high resolutions. These high-fidelity data can be either unavailable due to limited observations or computationally expensive to obtain through numerical models (Jeffers et al., 2016; Shainer et al., 2009). These limitations impede the preparation of downscaling tools, preventing researchers from performing high-resolution downscaling in an efficient way.

The emergence of *zero-shot super-resolution* techniques provides a pathway to generate fast simulations on varying resolutions that do not require high-resolution data to train the super-resolution tool. Zero-shot learning is a learning setup where the target predictor (e.g., high-resolution features) is unseen during training, but can be estimated by the learner during evaluation (Xian et al., 2019). In the context of image super-resolution, it provides a first-order estimation of high-resolution data when these data are too expensive to obtain. For this purpose, zero-shot super-resolution is a more flexible way of leveraging a limited computational budget. The FNO has been proposed as a zero-shot learning method to downscale the simulations generated from numerically solving partial differential equations (PDEs) (Li et al., 2021). Different from classical DL models for image analysis that learn dynamics in Euclidean space at a fixed discretized domain (e.g., CNN), FNO learns in Fourier space through an integral kernel and as a result is able to emulate the PDE solver itself regardless of the domain discretization. So far, FNO has been employed in emulating a variety of Earth system processes, such as subsurface multiphase flow (Wen et al., 2022), sea surface height (P. Jiang et al., 2021), and wind velocity (Pathak et al., 2022). Despite its early success in developing surrogate models, the super-resolution feature of FNO has not yet been fully explored in Earth science applications.

Here, we leverage the capability of FNO in performing zero-shot super-resolution to efficiently downscale regional climate model simulations to convection-permitted scales (i.e., 1~3 km). We focus on the use case of near-surface heat-exposure relevant variables because of their increased significance for better understanding and predicting heatwaves and urban heat islands (Chakraborty et al., 2020; Qian et al., 2022; Song et al., 2022; Yang et al., 2019; Zhang et al., 2021). To this end, we perform the downscaling by leveraging an existing 3-month WRF simulation with a spatial resolution of 4-km over the Great Lakes region (J. Wang et al., 2022), where resolutions at kilometer scale are critical to uncover the topography and land use impact. We divided the 3-month 4-km simulation into training, validation, and test periods to develop the FNO, which was later compared with U-Net to assess its emulation performance at the 4-km resolution. A physics-constraint loss was employed to honor the physical relationships among the emulated variables during the training of FNO. To validate the downscaling capabilities of FNO, we further generated two additional sets of WRF simulations at the 1-km resolution. The first set of 1-km WRF simulations overlaps with the test period and is used to assess the downscaling results generated by the FNO trained at 4-km resolution. The second data set, including nearly a week of simulations, is used to train another FNO in order to both show the flexibility of FNO and compare FNO's performance trained at varying resolutions.

Below, Section 2 formulates the zero-shot super-resolution in a mathematical context. Section 3 details the WRF simulation data as well as how FNO is used to perform the zero-shot super-resolution for the near-surface heat estimates in this study. During FNO development, we incorporate a physics-based constraint by using the CC relation in the loss objective function to better restrain the emulations on the temperature, humidity, and pressure fields. In Section 4, we present the training result of the FNO and the impact of the physics-based constraint, followed by an in-depth assessment of the performance of the FNO. The performance assessment includes both the emulation at 4-km resolution and the downscaling at 1-km resolution. A brief conclusion is drawn in Section 5.

## 2. Problem Formulation

In this study, we are interested in emulating the spatially distributed near-surface heat-exposure-related fields (i.e., air temperature, relative/specific humidity (SH), and pressure) at a time step  $t$  given the corresponding WRF forcings. Specifically, we consider a two-dimensional space  $\mathbb{X} \subset \mathbb{R}^2$ . The physics-based model WRF  $G$  takes  $N_i$  inputs and generates  $N_o$  outputs at any location  $x$  ( $\forall x \in \mathbb{X}$ ), such that  $G : \mathcal{I} \rightarrow \mathcal{O}$  where  $\mathcal{I} = \mathcal{I}(\mathbb{X}; \mathbb{R}^{N_i})$  and  $\mathcal{O} = \mathcal{O}(\mathbb{X}; \mathbb{R}^{N_o})$ . Here, we aim to develop a numerical approximation  $G_\theta : \mathcal{I} \rightarrow \mathcal{O}$ , where  $\theta$  is a set of learnable parameters, that is able to perform *zero-shot super-resolution modeling* such that:

1.  $G_\theta$  is trained on a set of model realizations at a given spatial discretization with  $N_i$  pairs of inputs  $I_{t_1} \in \mathbb{R}^{N_{x_1} \times N_{x_2} \times N_i}$  and outputs  $O_{t_1} \in \mathbb{R}^{N_{x_1} \times N_{x_2} \times N_o}$ , where  $t_1$  belongs to the training period  $T_1$  and  $N_{x_1}/N_{x_2}$  are the numbers of uniformly discretized points at each spatial dimension.
2. Once trained,  $G_\theta$  can be readily used to evaluate a new input at a new discretized domain  $I'_{t_2} \in \mathbb{R}^{N'_{x_1} \times N'_{x_2} \times N_i}$  and generate the corresponding output  $O'_{t_2} \in \mathbb{R}^{N'_{x_1} \times N'_{x_2} \times N_o}$ , where  $t_2$  belongs to the period  $T_2$  that can differ from  $T_1$  and  $N_{x'_1}/N_{x'_2}$  are from a new domain discretization (typically  $N_{x'_1} > N_{x_1}$  and  $N_{x'_2} > N_{x_2}$  for the downscaling task).

Figure 1c illustrates the application of the FNO in zero-shot super-resolution. The FNO is first trained on simulations during time period  $T_1$  generated from WRF at a given resolution (e.g., 4-km) that is coarser than the desired resolution (e.g., 1-km). The trained FNO, which is in essence a mesh-invariant emulator of WRF, can be readily used to perform emulation at any higher resolution (e.g., 1-km) and at a different time period  $T_2$ .

## 3. Data and Methods

In this section, we first describe the WRF simulations that are used to develop the zero-shot super-resolution tool, FNO. Then, we briefly review the mathematical framework of the FNO that is used for the downscaling and introduce a physics-constrained loss function that restrains the relationship among the downscaled surface heat variables. After that, we describe the training and configuration of FNO as well as other benchmark DL models. Finally, we present the performance evaluation metrics.

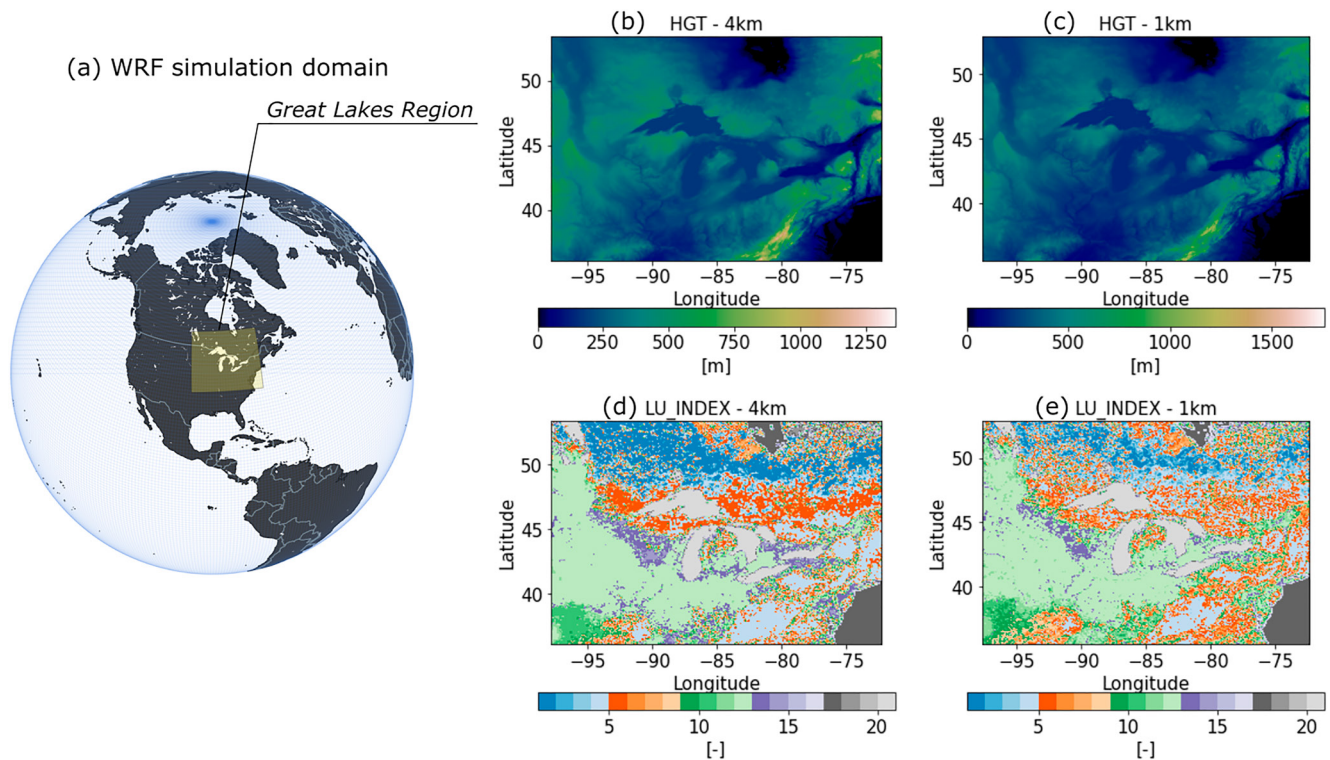
### 3.1. WRF Simulation Product

The WRF simulation at the Great Lakes region provides the training and evaluation datasets of the near-surface heat-exposure-related fields. The model is mainly driven by the following three types of inputs: (a) the atmospheric vertical profile including relative humidity (RH), wind components (UU and VV), air temperature (TT), pressure (PRES), geopotential height (GHT), etc; (b) land surface states including skin temperature (SKIN-TEMP), SM at multiple soil layers, etc.; and (c) topographical features including both terrain heights (HGT) and land use index (LU\_INDEX). In this study, we leverage the WRF simulations at both 4-km and 1-km resolutions at the Great Lakes region to develop and evaluate the performance of FNO.

#### 3.1.1. 4-km WRF Simulations

The 4-km WRF simulation was generated by J. Wang et al. (2022) for a recent regional modeling study. Figure 2a shows the modeling domain, which was set up at a spatial resolution of 4-km and contains a total of  $485 \times 543$  grid cells. The model ran from 1 June 2018 through 31 August 2018. The WRF model was driven by 3-hourly  $0.25^\circ$  European Center for Medium-Range Weather Forecasts atmospheric reanalysis of the global climate, version 5 (ERA5; Bell et al., 2020), leading to a total of 735 time steps. The details of the WRF physical configuration can be found in J. Wang et al. (2022). The 4-km WRF simulations are used to develop the FNO.





**Figure 2.** The topographic features of the Weather Research and Forecasting modeling domain: (a) the geographic location of the modeled Great Lakes regions. (b) and (c) The terrain heights (HGT) of the domain at resolutions of 4-km and 1-km, respectively. (d) and (e) The corresponding land use index (LU\_INDEX) at resolutions of 4 and 1-km, respectively (see Table 1 for detailed land use index description).

### 3.1.2. 1-km WRF Simulations

We further generate two sets of WRF simulations at a spatial resolution of 1-km, to evaluate the downscaling performance of the FNO that is trained on 4-km data set. The 1-km simulation contains  $1,939 \times 2,175$  grid cells over the same Great Lakes region. Instead of running WRF for the entire 3 months, which would be too computationally demanding, we performed the 1-km simulation for two separate weeks (i.e., 15 July 2018 through 20 July 2018 and 23 August 2018 through 31 August 2018). Note that due to the finer resolution, the computational cost of each week at the 1-km resolution is comparable to that of the 3-month 4-km simulation. Figures 2b–2e show the spatial patterns of HGT and LU\_INDEX (see Table 1) at the two resolutions.

### 3.2. Zero-Shot Super-Resolution Modeling

We develop the zero-shot super-resolution modeling of near-surface heat estimates as follows (Figure 3a):

1. We train the FNO model by using the WRF simulation and the corresponding static and dynamic inputs at a 4-km resolution (The static inputs: topography and land use. The dynamic inputs: the ERA5 data consisting of both atmospheric forcings and land surface states).
2. We evaluate the performance of FNO on 4-km emulation in a separate time period that is not used for training (The performance assessment is done by comparing it with the corresponding WRF simulation).
3. The developed FNO is used to perform downscaling at 1-km resolution using the 1-km inputs generated by WRF Preprocessing System (WPS).
4. We generate WRF simulation at 1-km spatial resolution and compare it with the FNO downscaling result.

We use FNO to emulate the current time step of four near-surface variables (i.e.,  $N_o = 4$ ), including 2-m temperature ( $T_2$ ), 2-m relative humidity ( $RH_2$ ),  $SH$ , and surface pressure ( $PSFC$ ) (Figure 3b). The emulator takes as inputs  $I_t$  the following dynamics or states: (a) the atmospheric forcings including RH, wind component (zonal and meridional), air temperature, pressure, and geopotential heights at three vertical levels (i.e., 925, 850, and 100 hPa); (b) the land surface states including skin temperature and the SM at the top soil layer (0–10 cm); and

**Table 1**

The Descriptions of the Land Use Indices (LU\_INDEX) Used in Weather Research and Forecasting Modeling (See Figures 2d and 2e for Its Spatial Distribution)

LU_INDEX	Description	LU_INDEX	Description
1	Evergreen needleleaf forest	12	Croplands
2	Evergreen broadleaf forest	13	Urban and built-Up
3	Deciduous needleleaf forest	14	Cropland/natural vegetation mosaic
4	Deciduous broadleaf forest	15	Snow and ice
5	Mixed forests	16	Barren or sparsely vegetated
6	Closed shrublands	17	Water
7	Open shrublands	18	Wooded tundra
8	Woody savannas	19	Mixed tundra
9	Savannas	20	Barren tundra
10	Grasslands	21	Lakes
11	Permanent wetlands		

(c) the topography features including terrain heights and land use index (as a result,  $N_i = 22$ ). The inputs  $I_l$  and the outputs  $O_l$  are in the same two-dimensional spatial grids.

### 3.3. Fourier Neural Operator (FNO)

The architecture of FNO is shown in the middle box of Figure 3b. FNO consists of one or multiple Fourier layers that learns and emulates the interactions among the variables of interest in Fourier space, sandwiched by two linear transformation layers that convert the dimensions between inputs, hidden, and output layers. Such sequential architecture can be formulated as  $I_l \mapsto V_0 \mapsto V_1 \dots \mapsto V_L \mapsto O_l$ , where  $L$  is the number of stacked FNO layers, and  $V_l$  are the hidden states  $\in \mathbb{R}^{N_{x_1} \times N_{x_2} \times N_v}$  where  $0 \leq l \leq L$  and  $N_v$  is the number of the hidden features. FNO evaluates the dynamics at each location  $x \in \mathbb{X}$  as follows:

1. Perform the linear transform at each location  $x$  from  $I_l(x) \in \mathbb{R}^{N_i}$  to  $V_0(x) \in \mathbb{R}^{N_v}$  as:

$$V_0(x) = W_0 I_l(x) + b_0, \quad (1)$$

where  $W_0 \in \mathbb{R}^{N_v \times N_i}$  and  $b_0 \in \mathbb{R}^{N_v}$  are the learnable weights and biases, respectively.

2. Execute the Fourier layer  $V_l \mapsto V_{l+1}$  at each location  $x$  as:

$$V_{l+1}(x) = \sigma(W_l V_l(x) + (\mathcal{K}(\phi)V_l)(x)), \quad (2)$$

where  $\sigma$  is a component-wise non-linear activation function;  $W_l \in \mathbb{R}^{N_v \times N_v}$  are the learnable parameters; and  $(\mathcal{K}(\phi)V_l)(x)$  is the kernel integral operator. FNO adopts a convolution operator defined in Fourier space as the kernel integral operator as below:

$$(\mathcal{K}(\phi)V_l)(x) = \mathcal{F}^{-1}(R_\phi \cdot (\mathcal{F}V_l))(x), \quad (3)$$

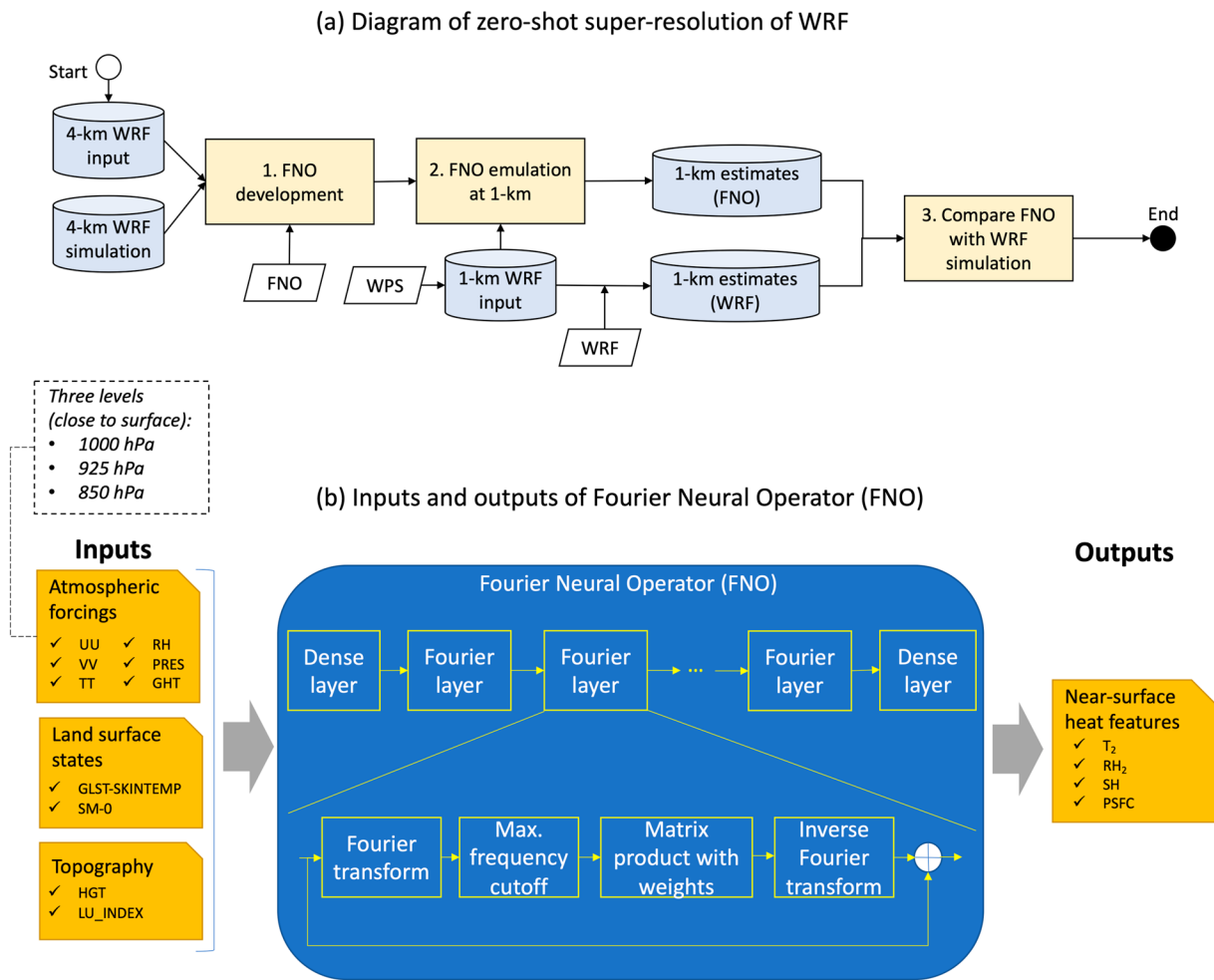
where  $\mathcal{F}V_l \in \mathbb{C}^{N_{x_1} \times N_{x_2} \times N_v}$  is the Fourier transform of  $V_l$ ;  $\mathcal{F}^{-1} \in \mathbb{C}^{N_{x_1} \times N_{x_2} \times N_v}$  is the corresponding inverse Fourier transform; and  $R_\phi$  is a set of learnable complex-valued tensor which truncates the dominant Fourier modes such that  $R_\phi \in \mathbb{R}^{k_m \times k_m \times N_v \times N_v}$  ( $k_m$  is the maximum truncated Fourier mode in each dimension). The Fourier transform and its inverse,  $\mathcal{F}$  and  $\mathcal{F}^{-1}$ , are implemented using the Fast Fourier Transform.

3. Perform the linear transform from the last Fourier layer to the output space as:

$$O_l(x) = W_2 V_L(x) + b_2, \quad (4)$$

where  $W_2 \in \mathbb{R}^{N_v \times N_i}$  and  $b_2 \in \mathbb{R}^{N_v}$  are the learnable weights and biases, respectively.

The first linear transform in Equation 1 serves as a general dimension projection tool that can either reduce ( $N_v < N_i$ ) or increase ( $N_v > N_i$ ) the feature dimensions. After that, the projected data is input into the Fourier layers in Equation 2 which is the main workhorse learning the interactions among the data. Different from CNNs that



**Figure 3.** Zero-shot super-resolution Weather Research and Forecasting (WRF) near-surface heat simulation using Fourier Neural Operator (FNO): (a) the diagram that develops FNO using 4-km WRF simulation and performs the downscaling using FNO at 1-km scale. (b) The inputs and outputs design of FNO that takes in a variety of atmospheric vertical profiles, land surface states, and topography features and generates the corresponding four near-surface heat dynamics at the same time step (FNO consists of one or multiple Fourier layers that learn and perform the dynamics of interest in Fourier space, sandwiched by two dense layers that convert the dimensions among inputs, hidden, and output features).

learn local relations through spatial convolution on image data, FNO performs such convolution (Equation 3) in the frequency domain. In doing so, FNO honors the way of the spectral method in solving the differential equations (Karniadakis & Sherwin, 2005) and thus is best suited to data generated from PDE-based systems. The output from the Fourier layers is finally converted back to its original dimension through another dense layer Equation 4.

One unique feature, as a result of learning in Fourier space, is that FNO is not restricted to a particular discretization and is considered as a type of neural operator that learns mappings between function spaces (Goswami et al., 2022). Equation 3 shows that FNO is resolution-independent because it mainly learns the interactions at the truncated Fourier space through learnable parameters  $R_\phi$ , and therefore, FNO can be evaluated in a space that is discretized in an arbitrary way. At a given location  $x$ , its estimation is composed of two parts: (a) the kernel operator  $(\mathcal{K}(\phi)V_l)(x)$  informing the domain and its nearby interactions; and (b) a local adjustment through  $W_1V_l(x)$ , as shown in Equation 2.

### 3.4. The Clausius–Clapeyron Relation-Based Physics-Constrained Loss

We integrate the Clausius–Clapeyron (CC) relation as the physics constraint of the near-surface heat-exposure estimates in the loss function  $\mathcal{L}$  as below:

$$\mathcal{L} = \mathcal{L}_{MSE} + \alpha\mathcal{L}_{CC}, \quad (5)$$

**Table 2**

The Deep Learning Models Described in Section 3.5 to Emulate or Downscale Near-Surface Heat States With Different Training Periods and Resolution, Loss Function Configurations, and Abilities in Generating 1-km and 4-km Results

Model	Training			Test		
	Period (data set)	With CC loss	Hard constrain $RH_2$	Period	1-km	4-km
FNO-base	1 July 0 hr–14 August 21 hr (4-km WRF)	No	No	23 August 3 hr–31 August 21 hr	Yes	Yes
FNO-cc	1 July 0 hr–14 August 21 hr (4-km WRF)	Yes	No	23 August 3 hr–31 August 21 hr	Yes	Yes
FNO- $RH_2$	1 July 0 hr–14 August 21 hr (4-km WRF)	No	Yes	23 August 3 hr–31 August 21 hr	Yes	Yes
FNO-1kmonly	15 July 0 hr–20 July 8 hr (1-km WRF)	No	No	23 August 3 hr–31 August 21 hr	Yes	Yes
U-Net	1 July 0 hr–14 August 21 hr (4-km WRF)	No	No	23 August 3 hr–31 August 21 hr	No	Yes

where  $\mathcal{L}_{MSE} = \frac{1}{N_x N_t} \sum_t \sum_x [O_t(x) - O_t^{WRF}(x)]^2$  is the mean square error (MSE) that measures the difference between the true WRF simulation  $O_t^{WRF}(x)$  and the FNO emulation  $O_t(x)$ , with  $N_x = N_{x_1} \cdot N_{x_2}$ ; and  $\mathcal{L}_{CC}$  is the CC relation-based constraint regularized by a hyperparameter  $\alpha$ .  $\alpha$  controls the influence of the physics constraint in the overall loss such that  $\alpha = 0$  means there is no physics constraint in the loss;  $\alpha = 1$  refers to the equal importance of the MSE and physics constraint; and an extremely large  $\alpha$  (e.g., 100) stands for the dominant role of the CC relation in the loss function.

The CC equation, along with other hydroclimatology relations, links the variables  $T_2$ ,  $RH_2$ ,  $SH$ , and  $PSFC$  such that:

$$w = \frac{SH}{1 - SH}, \quad (6a)$$

$$e = \frac{w}{w + 0.622} \times PSFC, \quad (6b)$$

$$e_{sat} = 610.78 \times \exp\left(\frac{17.27 \times T_2}{237.3 + T_2}\right), \quad (6c)$$

$$RH_2 = \frac{e}{e_{sat}} \times 100, \quad (6d)$$

where  $w$  is the water vapor mixing ratio (kg/kg);  $e_{sat}$  is the saturated vapor pressure (Pa); and  $e$  is the vapor pressure (Pa). We thus design  $\mathcal{L}_{CC}$  in a way that is consistent with Equations 6a–6d as:

$$\mathcal{L}_{CC} = \frac{1}{N_x N_t} \sum_t \sum_x [RH_{2,cc,t}(x) - RH_{2,t}(x)]^2, \quad (7)$$

where  $RH_{2,t}(x)$  is the  $RH_2$  emulated by FNO at location  $x$  for a given time step  $t$ ; and  $RH_{2,cc,t}(x)$  is the corresponding  $RH_2$  computed by Equation 6 using the FNO emulated  $T_2$ ,  $SH$ , and  $PSFC$ . Note that though the original values of the simulated variables are represented in Equations 5 and 6, we use the normalized values of the simulated variables in both  $\mathcal{L}_{MSE}$  and  $\mathcal{L}_{CC}$  to reduce the impact of magnitude during model training as described in the next section.

### 3.5. Deep Learning Model Developments

We developed the zero-shot super-resolution tool, FNO, as follows: (a) performing hyperparameter tuning on FNO without physics constraint (i.e.,  $\alpha = 0$ ) to find the best FNO configuration; and (b) training FNOs on varying  $\alpha$  values by using the best model configuration identified in the previous step. To assess the performance of the FNOs, we developed U-Net at the 4-km resolution and additional FNOs that is either trained at 1-km resolution or using a hard constraint to predict  $RH_2$ . The developments of all the models listed in Table 2 are described below:

#### 3.5.1. FNO-Base: Trained Using 4-km Simulation Without CC Constraint

The best FNO configuration was obtained through a grid search performed on FNO without physics constraint (i.e.,  $\alpha = 0$  or using  $\mathcal{L}_{MSE}$  as the loss function). We varied the number of Fourier layers  $L = [1, 3, 5, 7]$ , the number of hidden variables  $N_v = [5, 20, 40]$ , and the maximum cutoff frequency mode  $k_m = [3, 10, 20, 40]$ . We then temporally split the 3-month 4-km WRF simulation into 600/65/70 time steps as the training/validation/test



datasets, whose periods correspond to 1 July 0 hr–14 August 21 hr, 15 August 0 hr–23 August 0 hr, and 23 August 3 hr–31 August 21 hr, respectively.

### 3.5.2. FNO-cc: Trained Using 4-km Simulation With CC Constraint

Once finding the best FNO architecture, we then trained FNO by using different  $\alpha = [0.005, 0.01, 0.03, 0.05, 0.08, 0.1, 0.3, 0.6, 0.8, 1, 1.5, 2, 3, 5, 8, 10, 50, 100]$  using Equation 5 as the loss function during the training to assess the impact of the CC-based physics constraint.

### 3.5.3. FNO-RH<sub>2</sub>: Trained Using 4-km Simulation With a Hard Constraint on RH<sub>2</sub>

An additional FNO, that provides hard constraining on  $RH_2$ , was developed to verify the performance of the proposed physics constraint. Adopting the same configuration as FNO-base, FNO-RH<sub>2</sub> only predicts  $T_2$ ,  $SH$ , and  $PSFC$  at each time step and then calculates  $RH_2$  directly based on Equation 6.

### 3.5.4. FNO-1kmonly: Trained Using 1-km Simulation

To assess the performance of FNO trained with varying resolutions, we developed FNO-1kmonly solely based on the set of 1-km WRF simulations from 15 July 0 hr to 20 July 8 hr. While the computational costs are comparable between the 1 and 4-km WRF simulations, the differing performances of FNO-1kmonly and the FNOs trained using 4-km simulations (i.e., FNO-base and FNO-cc) are expected to uncover the impacts of the spatial and temporal contents of the training data on FNO development.

### 3.5.5. U-Net: Trained Using 4-km Simulation

U-Net, a widely used convolutional network (Ronneberger et al., 2015), is developed to assess the emulation of FNO at the 4-km resolution. While known for its capability in image processing, U-Net is not resolution-invariant and thus can not provide 1-km simulations. Here, we developed the U-Net by using three blocks of convolution in both contracting and expansive paths with the remaining architecture equivalent to Ronneberger et al. (2015).

During the training, we adopted the same strategy to develop each model described above as Li et al. (2021) such that the training was performed on 500 epochs by using the Adam optimizer (Kingma & Ba, 2014) with an initial learning rate of 0.001 that decayed at a rate of 0.5 at every 100 epochs. The raw WRF simulation was normalized by removing the mean and scaling to unit variance before the training. All the training was performed on one Nvidia A100 GPU with 64 GB memory. Depending on the model size, the training time varies from less than 8 min (for the most lightweight FNO with  $L = 1$ ,  $N_v = 5$ , and  $k_m = 3$ ) to nearly 7 hr (for the most complicated FNO with  $L = 7$ ,  $N_v = 40$ , and  $k_m = 40$ ).

## 3.6. Performance Evaluation

We employ two metrics to evaluate the performance of an emulated output  $o \in [T_2, RH_2, SH, PSFC]$  at each location, that is, the Nash–Sutcliffe model Efficiency ( $NSE$ ) and the modified Kling–Gupta Efficiency ( $mKGE$ ) (Kling et al., 2012) coefficients.  $NSE$  evaluates the impact of the variance of the estimation bias, thus putting more emphasis on large deviations such as extremes. On the other hand,  $mKGE$  provides an overall assessment of modeling performance by considering a variety of metrics including correlation, variability, and bias. The calculations of the two metrics are given below:

$$NSE = 1 - \frac{\sum_n (o_n - o_n^{WRF})^2}{\sum_n (\bar{o}^{WRF} - o_n^{WRF})^2}, \quad (8a)$$

$$mKGE = 1 - \sqrt{(r - 1)^2 + (\gamma - 1)^2 + (\beta - 1)^2}, \quad (8b)$$

where  $n$  is the time index of the test period;  $o^{WRF}$  is the corresponding WRF simulation of  $o$ ;  $\bar{o}^{WRF}$  refers to the average value of  $o^{WRF}$  across the index  $n$ ;  $r$  is the correlation coefficient between  $o$  and  $o^{WRF}$ ;  $\gamma$  is the variability ratio; and  $\beta$  is the bias ratio. Both  $NSE = 1$  and  $mKGE = 1$  indicate the perfect matching between  $o^{WRF}$  and  $o$ , and smaller values mean more mismatches between the emulation and the true simulation. To further assess the overall performance, we computed both their spatial mean and median, denoted by  $\bar{\cdot}$  ( $\bar{NSE}$  and  $\bar{mKGE}$ ) and  $\sim$  ( $\sim{NSE}$  and  $\sim{mKGE}$ ), respectively.

**Table 3**

The Performance of the Trained Deep Learning Models Listed in Table 2 in Generating the Four Near-Surface Heat States at 4 and 1-km Resolutions on the Test Period, Using the Median of Nash–Sutcliffe Model Efficiency (NSE) and Modified Kling–Gupta Efficiency (mKGE), With  $\alpha = 1$  Used in FNO-cc

	$T_2$	$RH_2$	$SH$	$PSFC$
$\overline{NSE/mKGE}$ —4-km				
FNO-base	0.892/0.913	0.685/0.781	0.815/0.846	0.962/0.953
FNO-cc	0.897/0.897	0.710/0.791	0.854/0.876	<b>0.965/0.950</b>
FNO- $RH_2$	0.888/0.890	0.665/0.781	0.828/0.850	0.960/0.955
FNO-1kmonly	0.367/0.595	0.356/0.658	0.227/0.482	0.889/0.853
U-Net	<b>0.904/0.926</b>	<b>0.713/0.799</b>	<b>0.880/0.904</b>	0.958/0.954
$\overline{NSE/mKGE}$ —1-km				
FNO-base	0.883/0.917	0.699/0.814	0.835/0.865	0.962/0.958
FNO-cc	<b>0.896/0.928</b>	<b>0.725/0.822</b>	<b>0.868/0.896</b>	<b>0.964/0.960</b>
FNO- $RH_2$	0.884/0.923	0.666/0.813	0.837/0.875	0.960/0.960
FNO-1kmonly	0.372/0.628	0.383/0.690	0.265/0.525	0.895/0.876
U-Net	n/a	n/a	n/a	n/a

Note. The corresponding result on the mean of NSE and mKGE are shown in Table S1 in Supporting Information S1. Bold values represent the best model for each of the four emulated variables.

## 4. Results and Discussion

In this section, we first present the DL model training result, including the hyperparameter tuning on FNO-base, the impact of the physics constraint using FNO-cc, and the comparison between FNO-cc and FNO- $RH_2$ . Then, by taking WRF simulation as ground truth, we evaluate the performance of FNO-cc in emulating the 4-km near-surface variables on the test data set. To this end, we assess its performance on each spatial grid, compare its emulation with that of U-Net, and explore how the emulation is affected by the land use index.

After that, we conduct the performance evaluation of FNOs in downscaling 1-km dynamics. Particularly, we assess the 1-km performance by comparing the FNOs trained using various resolutions, that is, 4-km (FNO-cc) and 1-km (FNO-1kmonly). Then, the emulation performances of FNO-cc at both resolutions are compared to assess to what extent the model is able to keep its emulation capability when evaluated at a higher and untrained resolution. Last, we select four subregions with varying topographic features to provide a focused analysis of FNO performance.

### 4.1. Deep Learning Model Training Result

An overview of the performances of all the developed DL models is presented in Table 3, based on the NSE and mKGE computed at each spatial grid of every model using the test data set. The table shows the  $\overline{NSE}$  and  $\overline{mKGE}$  of the models in generating all four near-surface heat states at both resolutions, with FNO-cc using  $\alpha = 1$ . For each emulated variable, the best performance with the highest metric is bolded. Generally, except for FNO-1kmonly, most

models are able to produce reasonably well emulations at the two resolutions with the medians of both metrics greater than 0.65. U-Net and FNO-cc turn out to be the best models in 4-km and 1-km, respectively.

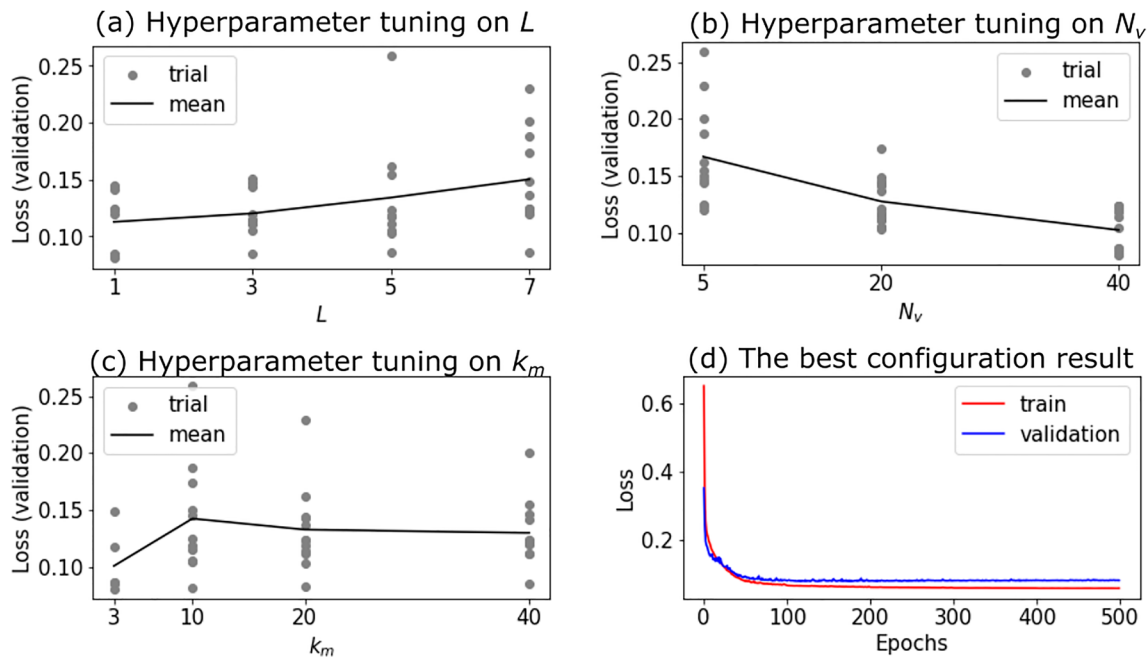
#### 4.1.1. Hyperparameter Tuning Result of FNO-Base

The best hyperparameter configuration is found for all FNO models by tuning FNO-base. Figures 4a–4c show its loss metrics on the validation data set against the three hyperparameters, where the black line is the averaged loss over the trials given one hyperparameter value. The best FNO configuration that yields the minimum loss turns out to be  $L = 1$ ,  $N_v = 40$ , and  $k_m = 3$ . The validation loss generally decreases with the increase of the hidden states  $N_v$  (Figure 4b), whereas more Fourier layers  $L$  and a larger frequency cutoff  $k_m$  do not necessarily improve the performance (Figures 4a and 4c). This is probably because of the limited number of the training data set (600 data points), which are too few to be able to fully utilize a more complicated FNO architecture. Indeed, the best-configured FNO-base, using only one hidden Fourier layer and three cutoff frequency models, has only 31,524 trainable parameters, which is much fewer than U-Net and the heaviest FNO in the tuning that contain 483,908 and 5,122,724 trainable parameters, respectively.

Despite the limited training data, the best-configured FNO-base exhibits reasonable training results according to the plot of training and validation losses against training epochs shown in Figure 4d. It can be observed that the losses decrease with epochs and converges at around 300 epochs. Further, there is little discrepancy between training and validation losses at convergence, suggesting the limited impact of the temporal extrapolation from the validation period.

#### 4.1.2. The Impact of $\alpha$ on FNO-cc

We used the architecture of the best-configured FNO-base to further train FNO-cc on varying  $\alpha$ . Note that FNO-base is a special case of FNO-cc with  $\alpha = 0$ . Figure 5 plots  $\overline{NSE}$  and  $\overline{mKGE}$  against  $\alpha$  for each of the four emulated near-surface variables on the test data set. For  $T_2$ ,  $RH_2$ , and  $SH$ ,  $\overline{NSE}$  slightly increases with  $\alpha$  and peaks at around  $\alpha = 1$ , while  $\overline{mKGE}$  is generally unchanged for  $\alpha$  less than 1. Different from the other three variables, the emulation of  $PSFC$  does not get improved over  $\alpha$  due to its excellent performance even without the physics constraint with both  $\overline{mKGE}$  and  $\overline{NSE}$  greater than 0.9. A similar conclusion on the impact of  $\alpha$  can be drawn based on the corresponding results of  $\overline{NSE}$  and  $\overline{mKGE}$  in Figure S2 in Supporting Information S1.

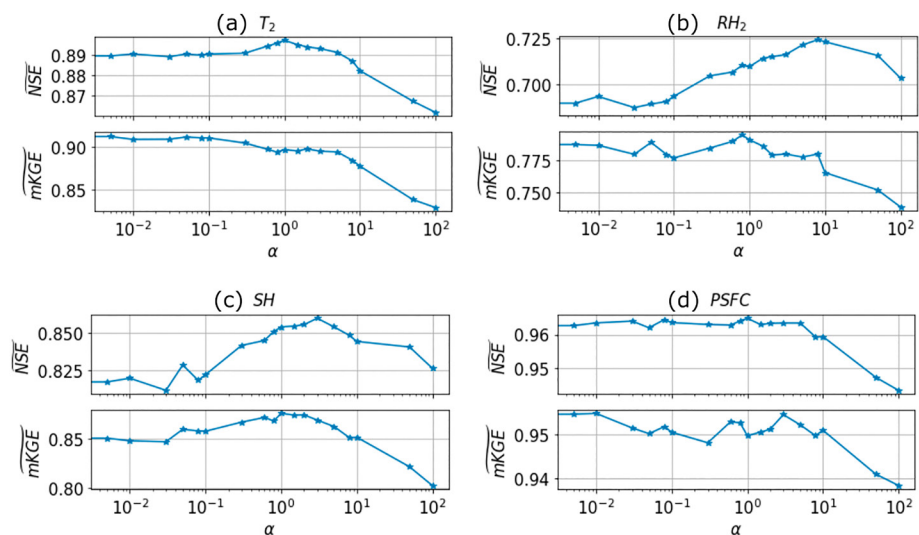


**Figure 4.** The hyperparameter tuning result of Fourier Neural Operator (FNO) using  $\alpha = 0$  in Equation 5: (a)–(c) The loss computed at the validation set against the three FNO hyperparameters: the number of Fourier layers  $L$ , the number of hidden features  $N_v$ , and the cutoff frequency mode  $k_m$ , respectively (the gray dots and black lines refer to all the trails and the mean of a given hyperparameter value). (d) The loss against the training epoch of the best hyperparameter configuration (i.e.,  $L = 1$ ,  $N_v = 40$ , and  $k_m = 3$ ).

In short, using physics constraint marginally improves the performance of FNO, particularly on  $T_2$ ,  $RH_2$ , and  $SH$ , when  $\alpha$  is less than around 1 ~5. Therefore, we use the FNO-cc trained on  $\alpha = 1$  for the remaining analysis.

#### 4.1.3. The Performance Gain by the Proposed Physics-Based Constraint

Thanks to the physics-based constraint, FNO-cc outperforms its vanilla version (i.e., FNO-base) in emulating all four variables at both resolutions, as shown in Table 3. Further, FNO-cc performs better than FNO- $RH_2$  that adopts a hard constraint by directly calculating  $RH_2$  from the other three emulated variables. For instance, the



**Figure 5.** The impact of the physics-constraint loss in Equation 5: (a)–(d) The computed  $\overline{NSE}$  and  $\overline{mKGE}$  against different  $\alpha$  for the 2-m temperature ( $T_2$ ), 2-m relative humidity ( $RH_2$ ), specific humidity ( $SH$ ), and surface pressure ( $PSFC$ ), respectively, with  $\alpha = [0.005, 0.01, 0.03, 0.05, 0.08, 0.1, 0.3, 0.6, 0.8, 1., 1.5, 2., 3., 5., 8., 10., 50., 100.]$  (see the corresponding results  $\overline{NSE}$  and  $\overline{mKGE}$  in Figure S2 in Supporting Information S1).

performance of FNO- $RH_2$  in emulating  $RH_2$  is worse than that of FNO-cc, with  $\widetilde{NSE}$  decreasing from 0.710 (FNO-cc) to 0.665 (FNO- $RH_2$ ). The inferior performance of FNO- $RH_2$  is probably due to its fewer training variables (i.e., three) than FNO-cc. This result underscores the significance of leveraging all the existing information in developing the FNO model, including both training simulation and known physical relationships.

#### 4.2. FNO Emulation at 4-km Scale

Figure 6 shows the heatmaps of spatially-distributed  $NSE$  and  $mKGE$  of the four emulated variables for the test period, using the two best-performed models at 4-km resolution, that is, FNO-cc and U-Net. A diverging colormap is used ranging from  $-1$  (red) to  $0$  (yellow) to  $+1$  (blue) to display the performance metrics. The mean and median of the two metrics are labeled on top of each subplot. To evaluate the impact of land use index (Figure 2d), we use boxplots to visualize the distribution of the two metrics of each emulated variable for each land use shown in Figures 7 and 8. In each subplot, the land use indices are ranked in the  $x$ -axis according to their percentages in the overall grids.

##### 4.2.1. Overall Performance of FNO-cc

Figures 6a–6h show that FNO-cc's emulations on  $T_2$  and  $PSFC$  are better than that of  $RH_2$  and  $SH$ . For  $T_2$ , the  $NSE$  and  $mKGE$  are close to 1 in most regions, while the yellow and red in areas such as Lake Superior and the Pacific ocean indicates inferior performance.  $T_2$  is able to achieve an average of 0.819 and 0.879 for  $NSE$  and  $mKGE$ , respectively, and both the averaged metrics of emulating  $PSFC$  are larger than 0.9. In fact, except for the ridge mountain area at the lower right of the domain (Figure 2b), the dominant “blue” in the remaining area of  $PSFC$  illustrates the nearly perfect match between the FNO emulation and the WRF simulation. On the other hand, the metric heatmaps of both  $RH_2$  and  $SH$  exhibit more yellow and red regions than that of  $T_2$  and  $PSFC$ , indicating their inferior performances. Though the mean of  $mKGE$  of  $RH_2$  is still able to be greater than 0.75, its averaged  $NSE$  drops to only 0.627 with most of the low  $NSE$  occurring in lake areas (e.g., Superior Lake in the middle and Hudson Bay in the north). Different from  $RH_2$ , the low averaged metrics of  $SH$  are mainly due to the vast red spots occurring across the whole domain whose  $NSE$  and  $mKGE$  can be as low as negative several hundred. While the medians of the two metrics are greater than 0.85, the extremely low values in the red spots lowered their mean to  $-171.95$  and  $0.477$  for  $NSE$  and  $mKGE$ .

##### 4.2.2. Comparable Performance With U-Net

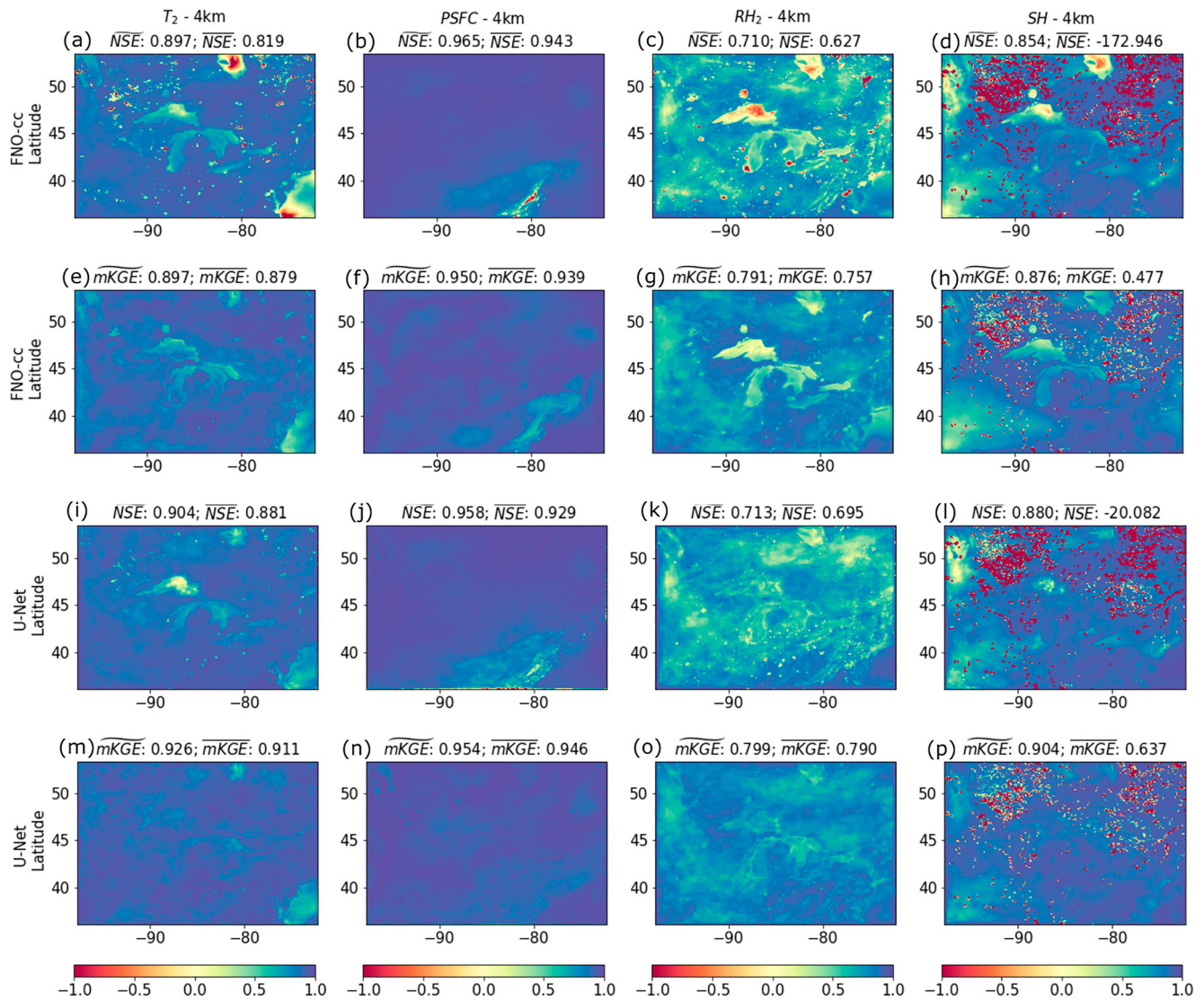
Table 3 shows that both FNO-base and FNO-cc generate emulations with comparable performance with U-Net. The two FNOs have marginally higher performance metrics in  $PSFC$  but slightly lower metrics in the other three variables. A further assessment of the spatially distributed  $NSE$  and  $mKGE$  of FNO-cc and U-Net in Figure 6 reveals that the marginal differing performance between the two models mainly occurs in the lake regions and water bodies. Take the example of Superior Lake and Hudson Bay. While both models show inferior performances in the two regions over the remaining areas, FNO-cc exhibits more yellow and reddish spots (i.e., lower  $NSE$  and  $mKGE$ ) than U-Net. The general decreased emulation performance on the lake area of FNO-cc requires a further assessment of the impact of spatially-varying land uses on the FNO performance which is discussed in the next paragraphs.

##### 4.2.3. The Impact of Land Uses on the Performance

Nearly 77.5% of the land uses are contributed by five types of croplands and forests, including 12 (croplands), 5 (mixed forests), 1 (evergreen needleleaf forest), 4 (deciduous broadleaf forest), and 14 (cropland/natural vegetation mosaic). As illustrated in Figures 7 and 8, the emulations generated from FNO-cc, FNO-base, and U-Net perform well over these natural lands. The corresponding metric distributions of all four variables are either close to 1 or fall into the range of 0.5–1.0. In fact, the metrics are also high at other natural lands that are outside of the top five, such as grassland (10), wooded tundra (18), woody savannas (8), deciduous needleleaf forest (3), etc., indicating that FNO provides reasonably well emulations over natural vegetation.

Nevertheless, the emulation performances over the lake (21) and water bodies (17) are worse, which are ranked as the sixth and seventh in the overall land uses and attributed to 13.7% of the whole area. Except for  $PSFC$ , the metric distributions of the other three variables have a lower spread. This inferior performance over surface water happens to all three models, to different degrees, with U-Net less adversely affected than the two FNOs. For FNOs, although most metrics over the lake and water grids still hit around 0.5–1.0, the long tails of these



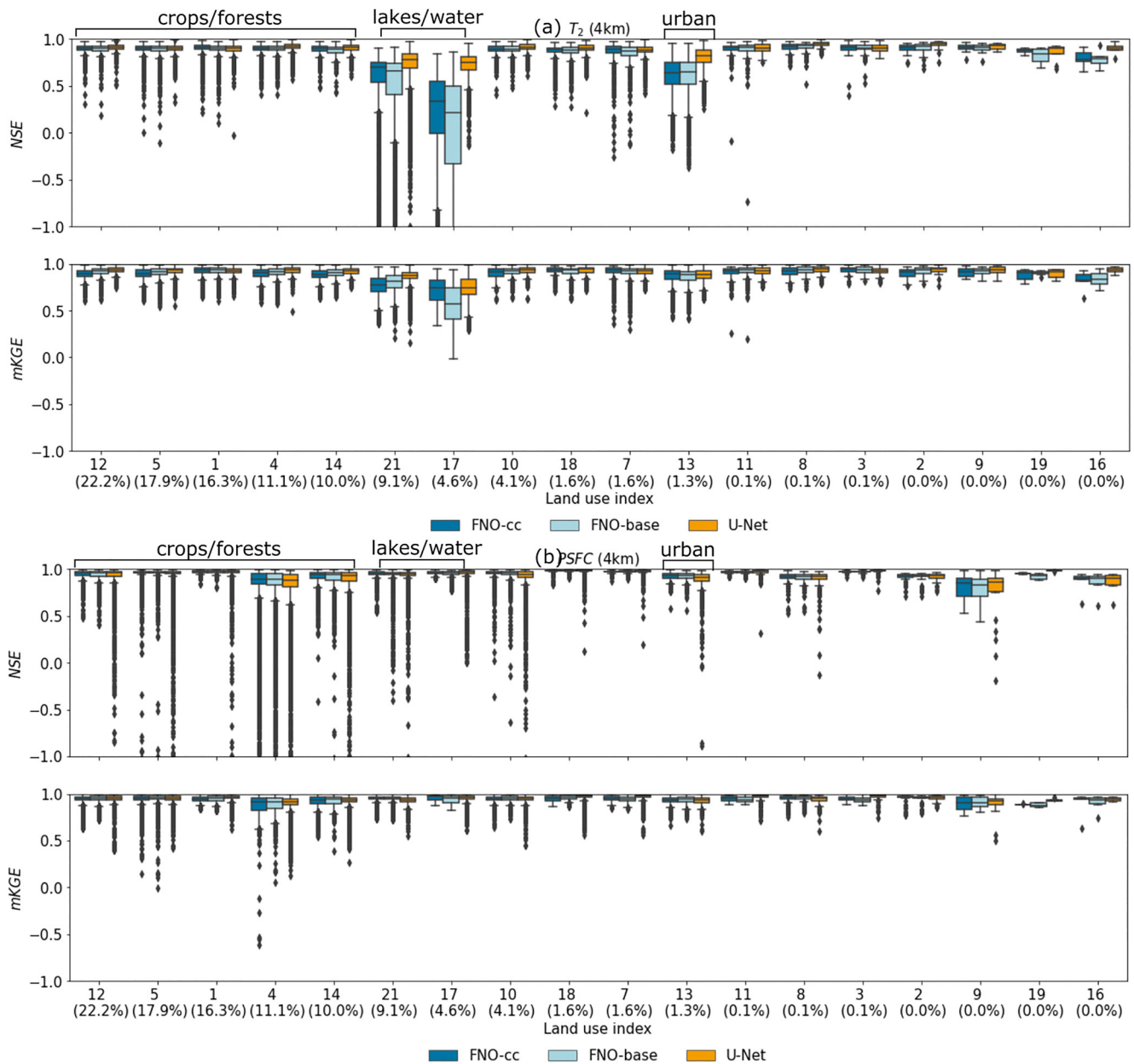


**Figure 6.** The emulation performance of FNO-cc (with  $\alpha = 1$ ) and U-Net at 4-km scale: (a)–(d) The spatially distributed Nash–Sutcliffe model Efficiency (NSE) computed by Equation 8a for the four emulated surface heat variables at the test period using FNO-cc. (e)–(h) The corresponding modified Kling–Gupta Efficiency (mKGE) computed by Equation 8b using FNO-cc. (i)–(l) The corresponding NSE computed by Equation 8a using U-Net. (m)–(p) The corresponding mKGE computed by Equation 8b using U-Net (The corresponding performances of the other models in Table 2 are plotted in Figures S3 and S4 in Supporting Information S1).

distributions toward low values (i.e., less than 0.5) illustrate the poor performance of FNOs over water. In fact, the spreads of distributions of NSE and mKGE of SH are so small that most of the values are actually below  $-1$ , resulting in the low and even negative mean shown in Figure 8b.

The discrepancy between FNO's performances of natural lands and water regions is probably due to the nature of very different atmospheres over land and lake regions (e.g., the atmosphere over lakes is much cooler and wetter in summer months, and is warming up much slower than that overland). FNO, which is developed for solving a fixed PDE, might not be able to fully capture the dynamics driven by two clusters of data that behave differently. This is different from a CNN like U-Net which learns the local interactions in Euclidean space. In fact, past studies found that FNO outperforms U-Net in datasets encoding “homogeneous” dynamics, such as ocean modeling (P. Jiang et al., 2021). One evidence of the reduced ability of FNO in capturing different dynamics is its inferior performance in emulating the  $T_2$  and the two humidity fields than PSFC in the lake region. While the pressure field is affected more by topography and less by land use, the near-surface temperature and humidity are more subjective to land use and thus differ more between land and water regions. That causes the difficulty of FNO in learning the varying interactions occurring in different land uses. Similarly, the FNO's performance in the urban





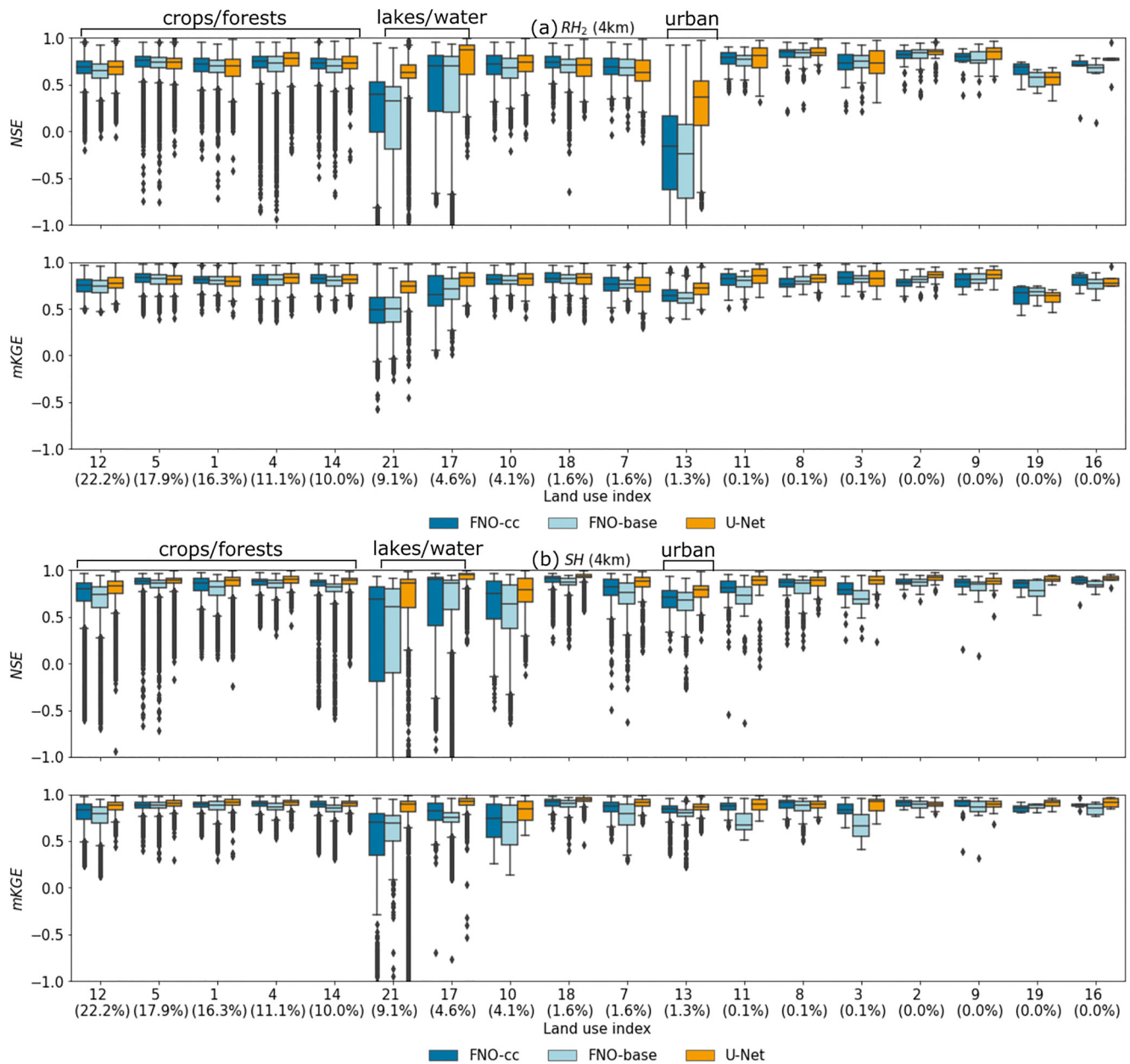
**Figure 7.** The impact of land use on the 4-km emulation of FNO-cc, FNO-base, and U-Net on the temperature and pressure fields (with  $\alpha = 1$ ): (a) the boxplots plots of Nash–Sutcliffe model Efficiency and modified Kling–Gupta Efficiency values of 2-m temperature ( $T_2$ ) against each land use index ranked by its proportion in the overall number of spatial grids. (b) The corresponding plots for surface pressure (PSFC) (See Table 1 for the detailed description of each land use index).

area (13) is also not as good as that over the natural lands, because the atmosphere over urban is much warmer and drier than surroundings.

Despite FNO's deteriorated performance over surface water, its similar performances with U-Net on most land covers (e.g., forests/crops) show the capability of FNO in capturing the dominant dynamics over the land surface, even in the presence of different clusters of data. This would add the benefit of capturing the heterogeneity of heat exposure over land, relevant for the health impacts of weather and climate.

### 4.3. FNO Downscaling at 1-km Scale

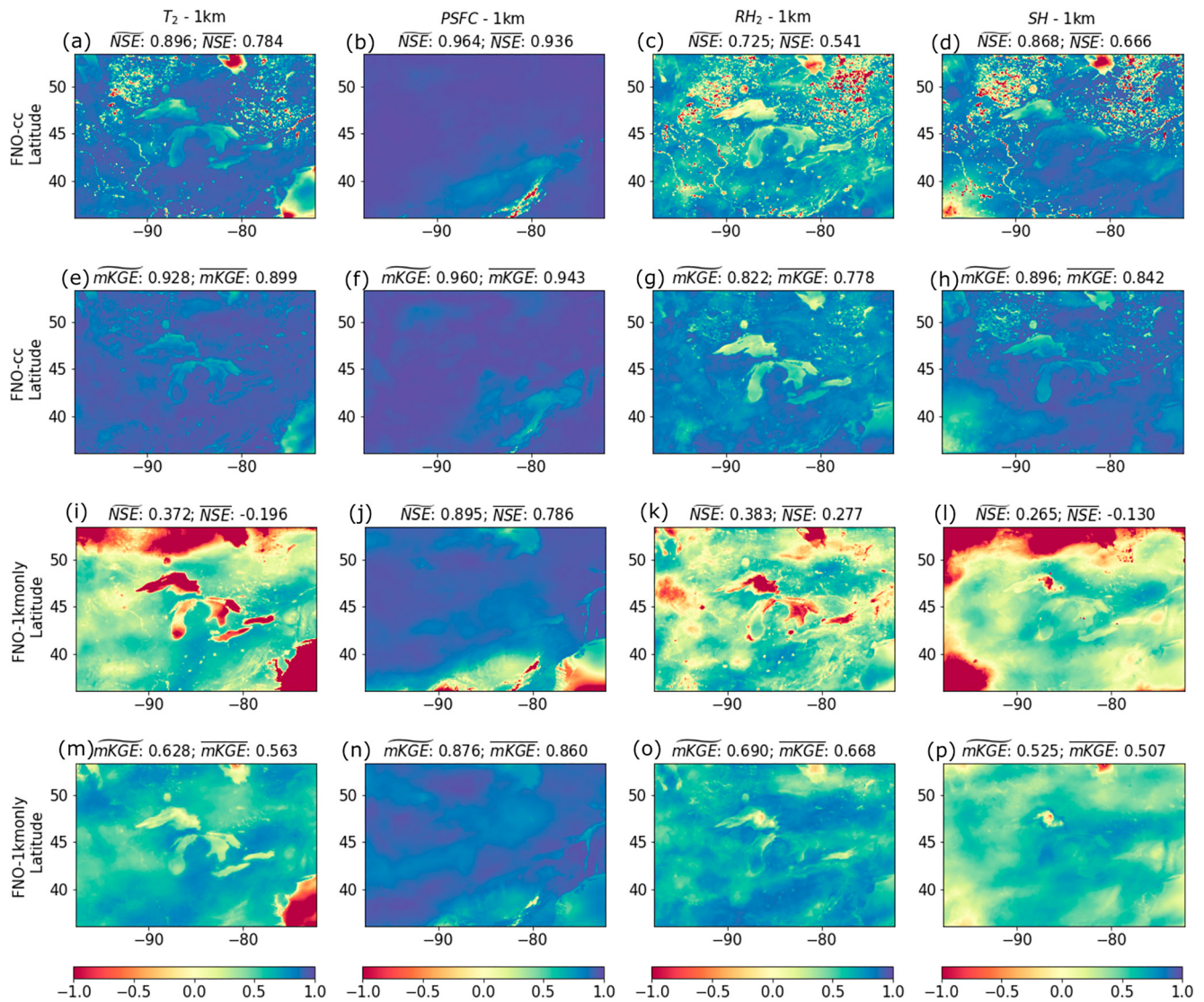
Figure 9 plots the spatial heatmaps of the NSE and mKGE at the 1-km resolution of the two FNOs trained using simulations with different spatial scales, that is, 4-km (FNO-cc) and 1-km (FNO-1kmonly).



**Figure 8.** The impact of land use on the 4-km emulation of FNO-cc, FNO-base, and U-Net on the humidity fields (with  $\alpha = 1$ ): (a) the boxplots of Nash–Sutcliffe model Efficiency and modified Kling–Gupta Efficiency values of 2-m relative humidity ( $RH_2$ ) against each land use index ranked by its proportion in the overall number of spatial grids. (b) The corresponding plots for specific humidity ( $SH$ ) (See Table 1 for the detailed description of each land use index).

### 4.3.1. Overall Performance of FNO-cc

Similar to the 4-km performance, FNO-cc provides better downscaling results of  $T_2$  and  $PSFC$  than that of  $RH_2$  and  $SH$  at 1-km resolution. The emulations of both  $T_2$  and  $PSFC$  achieve the averaged  $mKGE$  as high as 0.899 and 0.943, respectively. For  $NSE$ , while  $PSFC$  still keeps a high averaged value up to 0.936, that of  $T_2$  is slightly lower than 0.8, which is mainly due to the relatively low  $NSE$  scattering in and around the Great Lakes. Meanwhile, the averaged  $NSE$  of the two humidity fields are no greater than 0.7, though their mean  $mKGE$  is able to score up to 0.778 ( $RH_2$ ) and 0.842 ( $SH$ ). Akin to  $T_2$ , the low  $NSE$  of the humidity mainly occurs in the lake region and its surrounding areas. The high  $mKGE$  and relatively low  $NSE$  indicate that the downscaled  $T_2$ ,  $RH_2$ , and  $SH$  perform by and large well but do not fully capture the temporal variability (particularly in water regions).



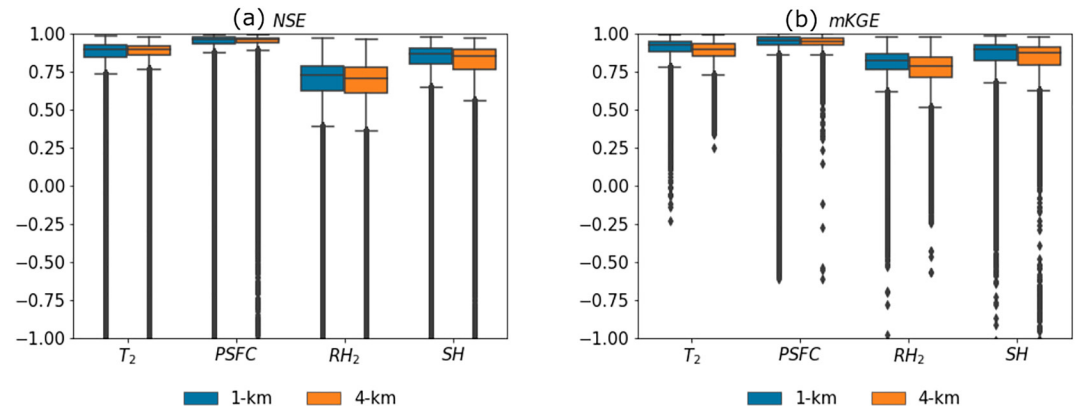
**Figure 9.** The emulation performance of FNO-cc (with  $\alpha = 1$ ) and FNO-1kmonly at 1-km scale: (a)–(d) the spatially distributed Nash–Sutcliffe model Efficiency (NSE) computed by Equation 8a for the four emulated surface heat variables at the test period using FNO-cc. (e)–(h) The corresponding modified Kling–Gupta Efficiency (mKGE) computed by Equation 8b using FNO-cc. (i)–(l) The corresponding NSE computed by Equation 8a using FNO-1kmonly. (m)–(p) The corresponding mKGE computed by Equation 8b using FNO-1kmonly (The corresponding performances of the other models in Table 2 are plotted in Figures S5 and S6 in Supporting Information S1).

#### 4.3.2. Comparison Between the FNOs Trained at Varying Resolutions

Table 3 shows that FNO-1kmonly, though directly trained at 1-km simulations, performs much worse than its counterparts trained at 4-km simulations (i.e., FNO-base and FNO-cc) such that some median metrics of FNO-1kmonly are as low as 0.2–0.5. Indeed, Figure 9 shows more reddish areas with lower metric values in the spatial metrics plot of FNO-1kmonly than that of FNO-cc, particularly in the lake regions. The inferior performance of FNO-1kmonly is probably because the limited 1-week training period does not capture sufficient temporal dynamics over the whole summer. Correspondingly, the trained FNO-1kmonly struggles to reproduce the WRF simulation at the test period.

This comparison demonstrates the tradeoff between the spatial and temporal contents of the training simulation used in developing FNO. Though generating the 1-km, 1-week WRF simulation costs comparable computational resources to that of the 4-km, 3-monthly WRF simulation, FNO-1kmonly does not compete with FNO-base and FNO-cc in emulating the 1-km dynamics. In other words, low-resolution simulation with a long period can be more informative to the dynamics in an unseen time regime than high-resolution simulation with a short period.





**Figure 10.** Performance comparison between 4-km emulation and 1-km downscaling of FNO-cc (with  $\alpha = 1$ ): (a) the box plots of Nash–Sutcliffe model Efficiency (NSE) of the four surface heat variables with blue and orange boxes referring to the values of 1-km and 4-km scales, respectively. (b) The corresponding box plots of modified Kling–Gupta Efficiency (mKGE) value.

This result highlights the potential of leveraging low-resolution data in developing such a resolution-invariant emulator to perform zero-shot super-resolution.

#### 4.3.3. 1-km Versus 4-km Emulation Performance of FNO-cc

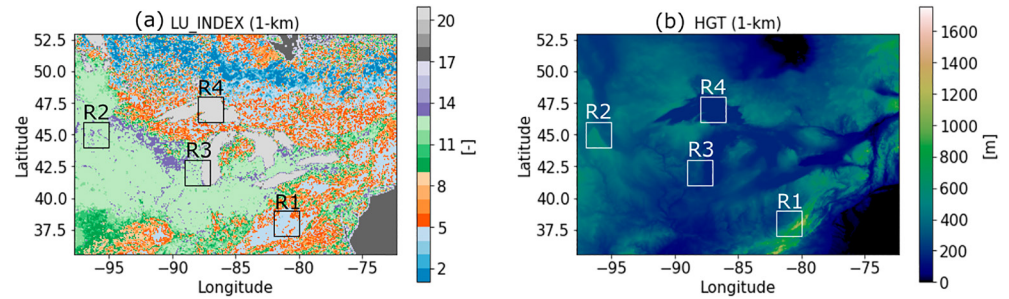
The better performance of FNO-cc over FNO-1konly further necessitates comparing the emulation performances of FNO-cc at 4 and 1-km. The objective is to assess whether FNO is able to keep a similar emulation performance against the WRF simulation when being applied at a finer resolution. We show the box plots of the metric distributions of each variable in Figure 10, where blue and orange represent the metrics of 1 and 4-km, respectively. Surprisingly, the box distribution plots are mostly consistent between the two scales with few variations. For  $T_2$ ,  $RH_2$ , and  $SH$ , we observe that the medians of both metrics at 1-km scale are slightly higher than that of 4-km scale. Another variation is that there are marginally more “outliers” with low  $mKGE$  values (less than zero) at the 1-km scale than 4-km for all four variables. Despite these variations, the general consistency of the metric distributions at the two scales suggests that FNO downscaling is able to keep the performance of the training scale. The similar performances of the two scales are also evidenced by the decreasing downscaling performances of FNO from natural land, urban, to the lake/water region (see the land use-based metrics distribution plot at 1-km scale in Figures S7 and S8 in Supporting Information S1).

#### 4.4. Subregion Analysis of FNO Performance

We selected four  $2^\circ \times 2^\circ$  rectangular subregions that show varying topography and land uses to further visualize and analyze how these factors affect the emulation and downscaling performances of the four variables. Figure 11 shows the geographical locations of the four regions:

- R1: cropland/natural vegetation land (with varying topography);
- R2: cropland (with a valley);
- R3: Chicago urban area (mixed with cropland and Lake Michigan);
- R4: east of Lake Superior.

The terrain height and land use indices of the selected regions can be referred to in Figure S9 in Supporting Information S1. For illustration purposes, we computed the average of the emulated and downscaled near-surface variables at 3 p.m. during the test period. Figures 12–15 show the spatial maps of the temporally averaged variables at 3 p.m. for the four regions, with the results of the entire domain plotted in Figure S10 in Supporting Information S1. In each figure, we plot both the FNO emulation and WRF simulation at 1-km and 4-km scales of each variable as well as the topographic features at the two scales (Note that we also performed the subregion analysis at 3 a.m. during the test period, shown in Figures S12–S17 in Supporting Information S1, with similar conclusions drawn below).



**Figure 11.** The geographical locations of the four subregions (i.e., R1–R4) with regards to the spatial maps of the land use index (a) and the terrain height (b).

#### 4.4.1. R1 (Figure 12)

The first region is located in the southeast of the simulation domain, with part of the Appalachian Mountains in the lower right (elevation less than 1,000 m). The WRF simulations at both scales show that, with the increase of the terrain height,  $T_2$ ,  $PSFC$ , and  $SH$  fields decrease from northwest to southeast of the subregion whereas  $RH_2$  is relatively high at the west of the mountain and low at the other side. Further, the zoom-in plot illustrates the finer delineation of 1-km simulation than that of 4-km. For instance, the 1-km  $T_2$  and  $PSFC$  at the eastern part exhibit an improved representation of network structure, which clearly results from the enhanced representation of the terrain elevation.

Compared with the WRF simulations, the FNO emulations at the two resolutions generally capture the trend of all the four variables though the 1-km performance is limited by the 4-km performance. The emulations of  $PSFC$  are nearly perfect, consistent with its high performance shown in Figure 9f. For  $T_2$ , FNO emulation is able to reproduce the decreasing trend when moving toward the southeast and yield the finer network structure at the 1-km resolution. However, the 4-km emulated  $T_2$  at the western part of the subregion is generally lower than that of WRF, which impacts the corresponding 1-km downscaling result. Likewise, the performance of the down-scaled  $RH_2$  and  $SH$  are greatly restrained by the 4-km emulation though the downscaling result yields a detailed representation. In addition to the improved tree structure due to the topography, we observe more white spots of 1-km  $SH$  than that of 4-km, consistent with the WRF simulation. These scattered low  $SH$  mostly result from the 1-km land use map that contains more spatially distributed forest land cover types.

#### 4.4.2. R2 (Figure 13)

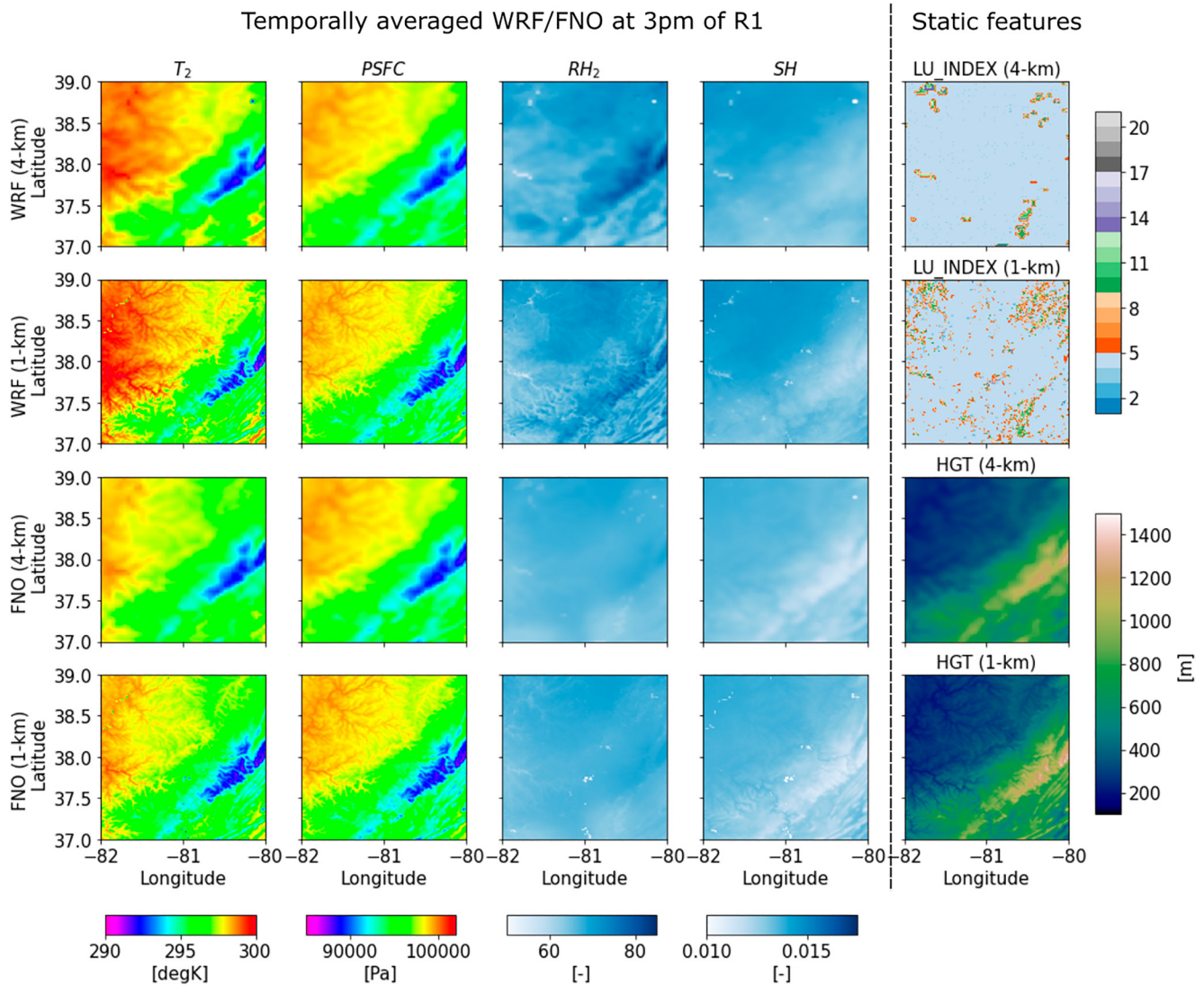
The second region is mostly composed of croplands with a valley in the east. WRF simulation generally exhibits low  $T_2$  with the exception of the south of the valley. The valley also divides the spatial distribution of  $PSFC$  and  $RH_2$  into the west part with lower values and the east part with higher magnitude. Similar to R1, the performance of FNO 1-km downscaling inherits from that of 4-km in this area. The  $PSFC$  downscaled by FNO still turns out to be able to nicely reproduce that of WRF simulation partially due to the perfect performance of FNO at the 4-km resolution.  $T_2$ , on the other hand, is overestimated by FNO at the western side of the valley at both resolutions. The emulations of  $RH_2$  and  $SH$  are worse. Though we can still observe the spatial changes of near-surface fields due to the valley, the spatial variabilities of both  $RH_2$  and  $SH$  are reduced in the FNO emulation compared to the WRF simulation.

The 1-km land use results in more scattered spots of the two humidity fields, but in different ways for FNO and WRF. The downscaled FNO produces mostly “white” spots over the lake areas (gray dots in Figure 13) similar to the “white” spots generated at the 4-km resolution, indicating lower humidity. Meanwhile, the 1-km WRF simulation generates more high humidity fields over the lake (“blue” spots), even though the WRF simulation at the 4-km resolution also yields low values that are actually captured by the 4-km FNO emulation. The discrepancy between 1-km and 4-km WRF simulation makes FNO struggle to reproduce the high humidity at the 1-km resolution over the water surface.

#### 4.4.3. R3 (Figure 14)

The WRF simulation shows that there is a sharp transition of these near-surface variables from Lake Michigan in the right of the subregion to the land area on the left. While  $T_2$  increases, the pressure and humidity fields decrease when moving from lake to surface land. Urbanization, which is the purple dots on the land surface maps,





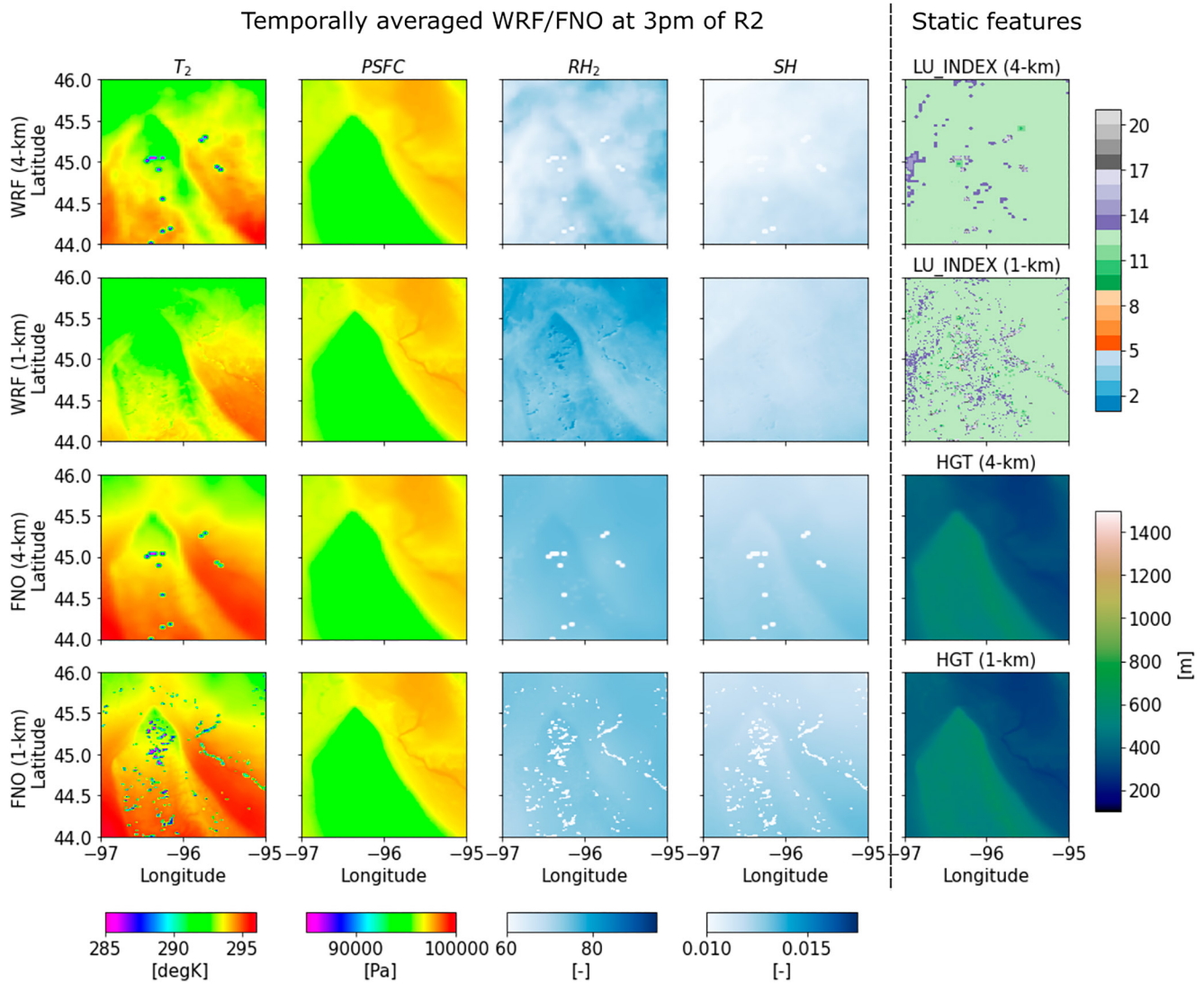
**Figure 12.** Temporally-averaged Weather Research and Forecasting (WRF) and Fourier neural operator (FNO)-cc (with  $\alpha = 1$ ) surface heat simulation at 3 p.m. of the subregion R1 during 23 August 2018 through 31 August 2018 (i.e., the test period). The right column is the land use index (LU\_INDEX) and terrain height (HGT) at both 1 and 4-km scales of R1. The left four columns are the simulations on the four surface heat variables at R1, including both WRF and FNO simulations at the two scales.

has a great impact on the near-surface variables. The temperature over urban land is generally higher than that of natural land, whereas urbanization reduces the pressure and humidity.

The FNO emulation reproduces the lake-land gradient and captures the urbanization impact to some extent. Although FNO fails to capture the decrease of humidity fields in the urban area, partially due to its reduced capability in emulating the humidity, the land use-induced spatial variabilities of pressure and temperature are well reflected in FNO emulations. The WRF-simulated PSFC field is accurately emulated by the FNO though  $T_2$  is slightly overestimated in the middle of the region. Furthermore, in the 1-km  $T_2$  field of FNO, there emerge many relatively low temperature spots (i.e., yellow spots) spreading across the high temperature field (in red) in the middle of the region, consistent with the corresponding downscaled WRF simulation. These scattered cooler locations result from the finer delineation of land use where many vegetation lands now occur and are mixed with the urban areas.

#### 4.4.4. R4 (Figure 15)

Similar to the simulation at R3, the existence of Lake Superior causes sharp decreases in  $T_2$  and  $SH$  and increases in  $PSFC$  and  $RH_2$ . Nevertheless, different from the other subregions, the homogeneity of topographic features in this subregion results in few changes of the 1-km simulation when compared to that of the 4-km. Regarding the

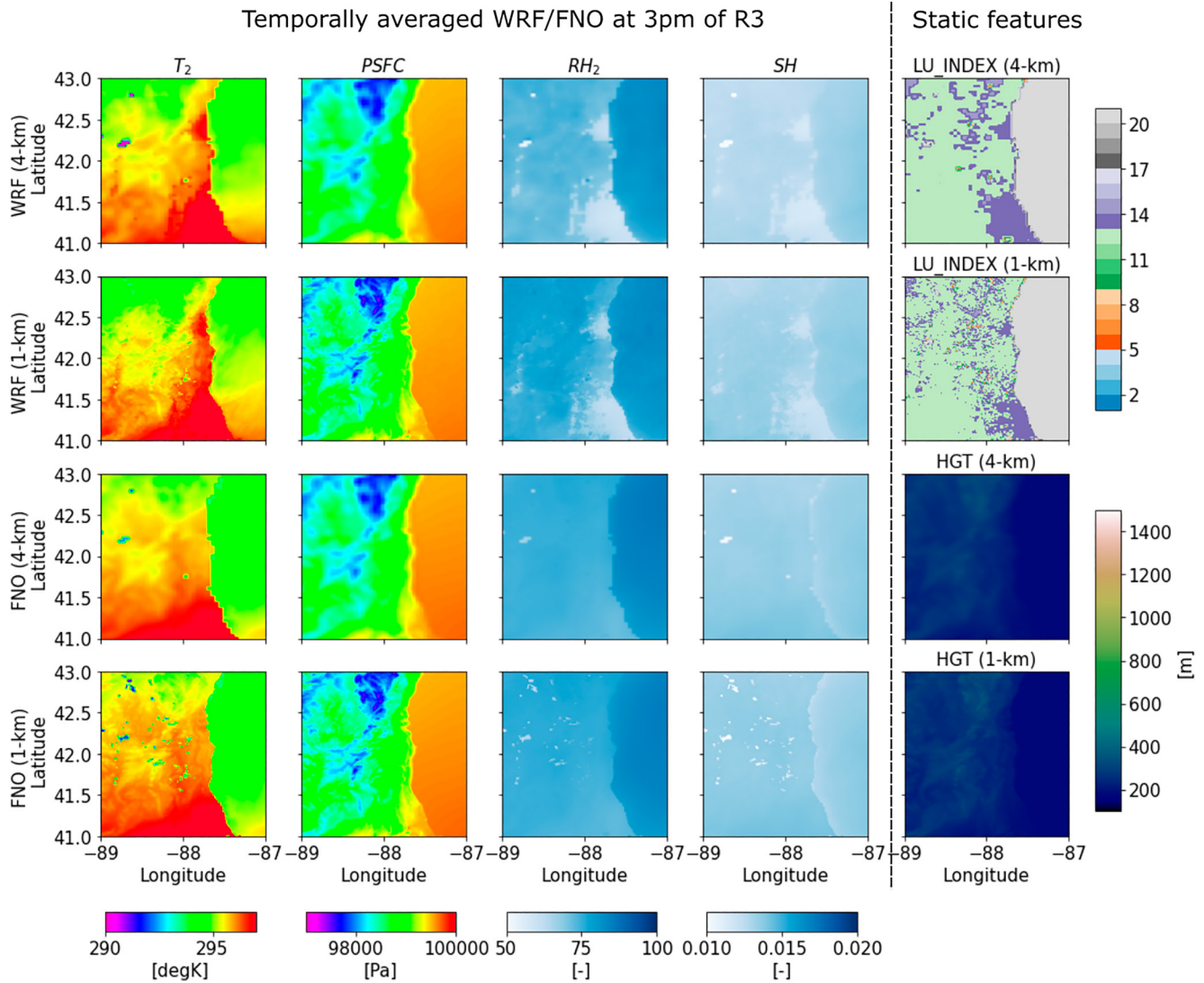


**Figure 13.** Temporally-averaged Weather Research and Forecasting (WRF) and Fourier neural operator (FNO)-cc (with  $\alpha = 1$ ) surface heat simulation at 3 p.m. of the subregion R2 during 23 August 2018 through 31 August 2018 (i.e., the test period). The right column is the land use index (LU\_INDEX) and terrain height (HGT) at both 1 and 4-km scales of R2. The left four columns are the simulations on the four surface heat variables at R1, including both WRF and FNO simulations at the two scales.

FNO emulation, the machine learning algorithm yields comparable results of all four variables with the WRF simulations, including both the land-lake changes and the magnitude. This close match between FNO and WRF illustrates that FNO is able to regenerate the averaged magnitude during the hottest hour of the day, despite its relatively poor performance in capturing their temporal variations over lake grids (i.e., low *NSE*). In fact, these near-surface climates generally show fewer variations in temperature/humidity fields over lakes than land due to higher heat capacity and thermal inertia of water (see the standard deviations of the four estimates in Figure S11 in Supporting Information S1).

### 5. Conclusions and Future Work

Our study showcases the usage of FNO, a state-of-the-art DL technique, to emulate and downscale the standard near-surface meteorological variables. To the best of our knowledge, this is the first implementation of zero-shot super-resolution in regional-scale climate modeling. Trained on merely 600 snapshots of WRF simulation at 4-km resolution, the developed FNO model is able to generate 1-km downscaled product from the WPS-processed ERA5 forcings that have comparable performances with the corresponding 4-km emulation at the test period



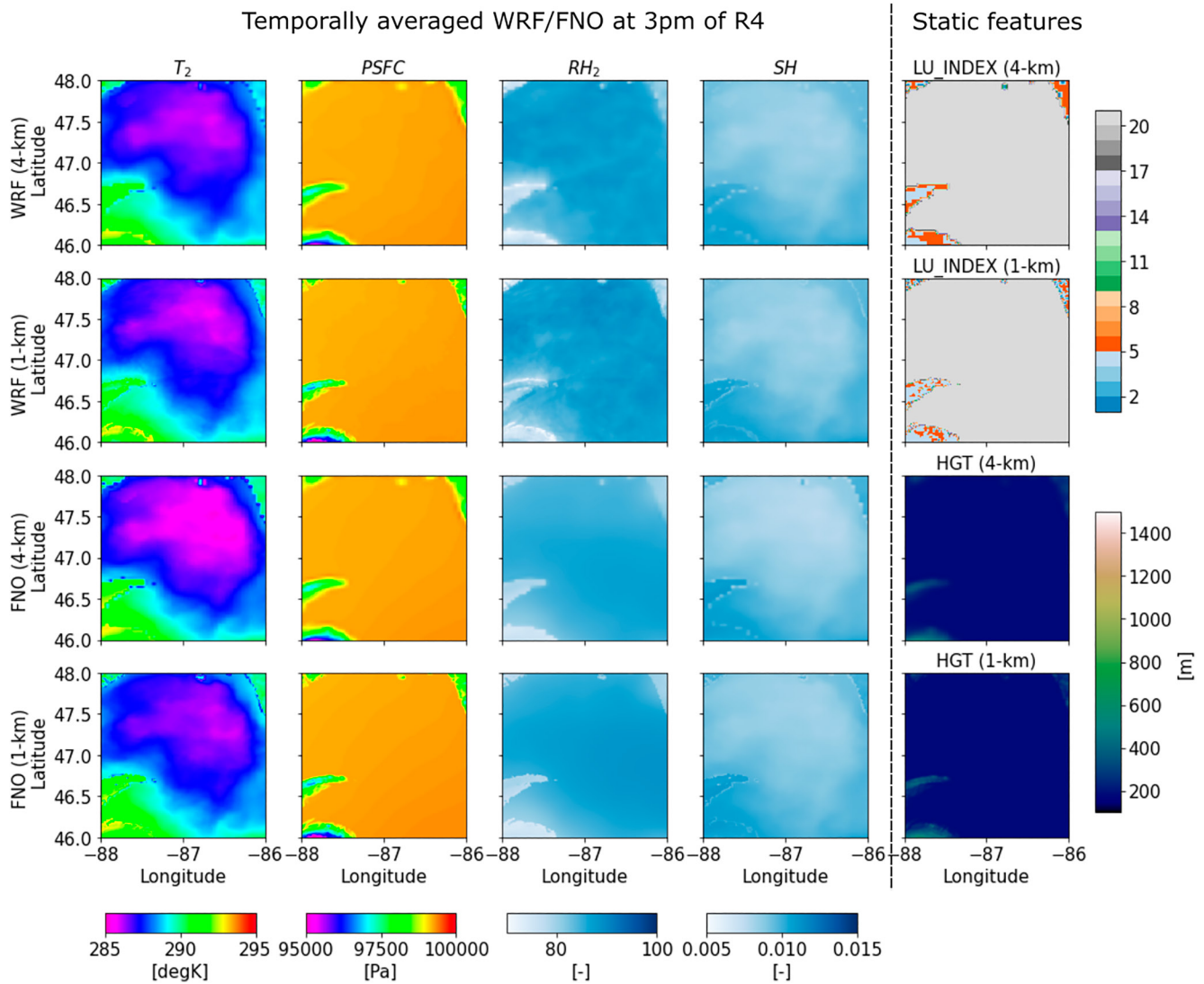
**Figure 14.** Temporally-averaged Weather Research and Forecasting (WRF) and Fourier neural operator (FNO)-cc (with  $\alpha = 1$ ) surface heat simulation at 3 p.m. of the subregion R3 during 23 August 2018 through 31 August 2018 (i.e., the test period). The right column is the land use index (LU\_INDEX) and terrain height (HGT) at both 1 and 4-km scales of R3. The left four columns are the simulations on the four surface heat variables at R1, including both WRF and FNO simulations at the two scales.

(Figure 10). Though only being tested at 1-km resolution, the FNO model theoretically can be used to perform the emulation at any spatial resolution.

We observe a tradeoff of the training data between its spatial and temporal contents through the comparison between FNO-1kmonly and FNO-base/FNO-cc. Given a limited computational budget, high resolution simulation does not necessarily encode sufficient temporal dynamics to be learned by a DL emulator. It is thus critical, in the future, to develop a strategic way of selecting the simulation period and the spatial resolution of the WRF simulation. Besides using simulations at one resolution, future work can focus on a multi-fidelity modeling approach (Lu et al., 2022; Meng & Karniadakis, 2020) of conducting zero-shot or even few-shot super-resolution by leveraging simulations with varying spatial scales to train a neural operator.

The performance of FNO in downscaling and emulating the surface heat dynamics is ranked from good to bad as the pressure *PSFC*, the 2-m temperature  $T_2$ , and the two humidity variables. Particularly, FNO does an excellent job in reproducing *PSFC* with both the average of *NSE* and *mKGE* metrics above 0.9. Its reduced performance in generating the temperature and humidity fields is probably because FNO, which is designed to





**Figure 15.** Temporally-averaged Weather Research and Forecasting (WRF) and Fourier neural operator (FNO)-cc (with  $\alpha = 1$ ) surface heat simulation at 3 p.m. of the subregion R4 during 23 August 2018 through 31 August 2018 (i.e., the test period). The right column is the land use index (LU\_INDEX) and terrain height (HGT) at both 1 and 4-km scales of R4. The left four columns are the simulations on the four surface heat variables at R1, including both WRF and FNO simulations at the two scales.

emulate a PDE system, might not be well suited to capture a dynamical system that is partially represented by semi-empirical equations (e.g., CC equation and other surface ecosystem processes). To solve this issue, we employed a physics-based constraint loss that restrains the training loss by the CC relation. The performance of physics-constrained FNO slightly improves and peaks when the physics-based constraint shares equal importance with the MSE loss (i.e.,  $\alpha = 1$  in Figure 5). Overall, the emulated heat dynamics sufficiently capture the CC-relation as  $\alpha$  increases (see Figure S1 in Supporting Information S1).

Terrain height and land use both have great impacts on the downscaling results and the overall performance of FNO. Through a subregion analysis (Section 4.4), we observe the finer features of the near-surface variables at the 1-km resolution due to either the enhanced representation of mountain ridge networks or more detailed scatters of different land uses. Meanwhile, we also find that FNO performs best over natural vegetation lands, followed by the urban area and the lake/water bodies (Figures 7 and 8). The varying performances at different land uses are probably owing to (a) the different parameters adopted in the PDE system over different land types such that FNO struggles to simultaneously learn all of them, and (b) that the land use index alone might be insufficient to represent all the physical constraints induced by the lower boundary condition. However, the

performance of the current FNO, trained on only 600 WRF snapshots, is expected to improve when given WRF simulation with a longer time period. This is evidenced by the superior performances of FNO-base and FNO-cc trained on 3-month simulations over FNO-1k trained on 1-week simulations. Future work can focus on both improving the representations of land use as the inputs of FNO and increasing training datasets in order to enhance neural network learning.

Generating fast and accurate climate dynamics at increasing spatial resolutions is always a long-standing challenge and need in Earth Sciences. Although our main objective is to downscale WRF simulation, the developed FNO is technically a generalized tool that is capable of performing both downscaling and emulation, the latter of which is to some extent explored in evaluating the 4-km emulation performance. Thus, future directions will be aligned with the following two pathways. The first direction will continue the downscaling efforts by applying FNO on more estimates (e.g., wind speed) and using more temporal data points. Given a longer simulation period (e.g., an entire year), one potential direction is to take the historical states into the FNO inputs to improve the predictability of the emulator and perform forecasting at the future time steps. Second, a well-developed FNO will be a perfect surrogate model used to evaluate different climate change-induced hypotheses by performing emulations driven by varying climate scenarios. With rapid developments of DL techniques in solving PDE-based systems, these two efforts can potentially yield high-fidelity climate products in an efficient manner and thus facilitate addressing grand challenges associated with climate change.

### Data Availability Statement

The scripts used in this study can be found on Zenodo: doi: [10.5281/zenodo.7927199](https://doi.org/10.5281/zenodo.7927199). The WRF simulations are currently hosted on the HPSS Archive System of National Energy Research Scientific Computing (NERSC) due to the large file size (~1.1 TB) and are available upon request.

### Acknowledgments

This research was supported by the U.S. Department of Energy (DOE), Office of Science (SC) Biological and Environmental Research (BER) program, as part of BER's Environmental System Science (ESS) program. This contribution originates from the Coastal Observations, Mechanisms, and Predictions Across Systems and Scales—Great Lakes Modeling (COMPASS-GLM) Project at Pacific Northwest National Laboratory (PNNL). PNNL is operated for DOE by Battelle under contract DE-AC05-76RL01830. This paper describes objective technical results and analyses. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government.

### References

- Abatzoglou, J. T., Dobrowski, S. Z., Parks, S. A., & Hegewisch, K. C. (2018). Terraclimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Scientific Data*, 5(1), 1–12. <https://doi.org/10.1038/sdata.2017.191>
- Atlas, L., Homma, T., & Marks, R. (1987). An artificial neural network for spatio-temporal bipolar patterns: Application to phoneme classification. In D. Anderson (Ed.), *Neural information processing systems* (Vol. 0). American Institute of Physics. Retrieved from <https://proceedings.neurips.cc/paper/1987/file/98f13708210194c475687be6106a3b84-Paper.pdf>
- Baño Medina, J., Manzanar, R., & Gutiérrez, J. M. (2020). Configuration and intercomparison of deep learning neural models for statistical downscaling. *Geoscientific Model Development*, 13(4), 2109–2124. <https://doi.org/10.5194/gmd-13-2109-2020>
- Bell, B., Hersbach, H., Berrisford, P., Dahlgren, P., Horányi, A., Muñoz Sabater, J., et al. (2020). *Era5 hourly data on pressure levels from 1950 to 1978 (preliminary version)*. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). Retrieved from <https://cds.climate.copernicus-climate.eu/cdsapp#!/dataset/reanalysis-era5-pressure-levels-preliminary-back-extension?tab=overview/>
- Chakraborty, T., Hsu, A., Many, D., & Sheriff, G. (2020). A spatially explicit surface urban heat island database for the United States: Characterization, uncertainties, and possible applications. *ISPRS Journal of Photogrammetry and Remote Sensing*, 168, 74–88. <https://doi.org/10.1016/j.isprsjprs.2020.07.021>
- Chen, C., Hu, B., & Li, Y. (2021). Easy-to-use spatial random-forest-based downscaling-calibration method for producing precipitation data with high resolution and high accuracy. *Hydrology and Earth System Sciences*, 25(11), 5667–5682. <https://doi.org/10.5194/hess-25-5667-2021>
- Coppola, E., Raffaele, F., & Giorgi, F. (2018). Impact of climate change on snow melt driven runoff timing over the alpine region. *Climate Dynamics*, 51(3), 1259–1273. <https://doi.org/10.1007/s00382-016-3331-0>
- Fowler, H. J., Blenkinsop, S., & Tebaldi, C. (2007). Linking climate change modelling to impacts studies: Recent advances in downscaling techniques for hydrological modelling. *International Journal of Climatology*, 27(12), 1547–1578. <https://doi.org/10.1002/joc.1556>
- Giorgi, F. (2019). Thirty years of regional climate modeling: Where are we and where are we going next? *Journal of Geophysical Research: Atmospheres*, 124(11), 5696–5723. <https://doi.org/10.1029/2018JD030094>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. Retrieved from <http://www.deeplearningbook.org>
- Goswami, S., Bora, A., Yu, Y., & Karniadakis, G. E. (2022). Physics-informed deep neural operator networks. arXiv. <https://doi.org/10.48550/ARXIV.2207.05748>
- Huang, X. (2020). Deep-learning based climate downscaling using the super-resolution method: A case study over the western us. *Geoscientific Model Development Discussions*, 2020, 1–18. <https://doi.org/10.5194/gmd-2020-214>
- Hutengs, C., & Vohland, M. (2016). Downscaling land surface temperatures at regional scales with random forest regression. *Remote Sensing of Environment*, 178, 127–141. <https://doi.org/10.1016/j.rse.2016.03.006>
- Im, E.-S., Pal, J. S., & Eltahir, E. A. B. (2017). Deadly heat waves projected in the densely populated agricultural regions of south Asia. *Science Advances*, 3(8), e1603322. <https://doi.org/10.1126/sciadv.1603322>
- Jeffers, J., Reinders, J., & Sodani, A. (2016). Chapter 22 - Weather research and forecasting (WRF). In J. Jeffers, J. Reinders, & A. Sodani (Eds.), *Intel Xeon phi processor high performance programming* (2nd ed., pp. 499–510). Morgan Kaufmann. <https://doi.org/10.1016/B978-0-12-809194-4.00022-3>
- Jiang, P., Meinert, N., Jordão, H., Weisser, C., Holgate, S., Lavin, A., et al. (2021). Digital twin Earth—coasts: Developing a fast and physics-informed surrogate model for coastal floods via neural operators. In *Neurips workshop on machine learning and the physical sciences*. Jiang, Y., Fu, P., & Weng, Q. (2015). Downscaling goes land surface temperature for assessing heat wave health risks. *IEEE Geoscience and Remote Sensing Letters*, 12(8), 1605–1609. <https://doi.org/10.1109/LGRS.2015.2414897>



- Karger, D. N., Conrad, O., Böhrer, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., et al. (2017). Climatologies at high resolution for the Earth's land surface areas. *Scientific Data*, 4(1), 1–20. <https://doi.org/10.1038/sdata.2017.122>
- Karniadakis, G., & Sherwin, S. (2005). *Spectral/hp element methods for computational fluid dynamics*. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198528692.001.0001>
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
- Kling, H., Fuchs, M., & Paulin, M. (2012). Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *Journal of Hydrology*, 424–425, 264–277. <https://doi.org/10.1016/j.jhydrol.2012.01.011>
- Kumar, B., Chattopadhyay, R., Singh, M., Chaudhari, N., Kodari, K., & Barve, A. (2021). Deep learning–based downscaling of summer monsoon rainfall data over Indian region. *Theoretical and Applied Climatology*, 143(3), 1145–1156. <https://doi.org/10.1007/s00704-020-03489-6>
- Le Roux, R., Katurji, M., Zawar-Reza, P., Quénot, H., & Sturman, A. (2018). Comparison of statistical and dynamical downscaling results from the WRF model. *Environmental Modelling & Software*, 100, 67–73. <https://doi.org/10.1016/j.envsoft.2017.11.002>
- Li, Z., Kovachki, N. B., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., & Anandkumar, A. (2021). Fourier neural operator for parametric partial differential equations. In *International conference on learning representations*. Retrieved from <https://openreview.net/forum?id=c8P9NQVtmnO>
- Liu, C., Ikeda, K., Rasmussen, R., Barlage, M., Newman, A. J., Prein, A. F., et al. (2017). Continental-scale convection-permitting modeling of the current and future climate of North America. *Climate Dynamics*, 49(1), 71–95. <https://doi.org/10.1007/s00382-016-3327-9>
- Lu, L., Pestourie, R., Johnson, S. G., & Romano, G. (2022). Multifidelity deep neural operators for efficient learning of partial differential equations with application to fast inverse design of nanoscale heat transport. *Physical Review Research*, 4(2), 023210. <https://doi.org/10.1103/PhysRevResearch.4.023210>
- Lucas-Picher, P., Argüeso, D., Brisson, E., Trambly, Y., Berg, P., Lemonsu, A., et al. (2021). Convection-permitting modeling with regional climate models: Latest developments and next steps. *WIREs Climate Change*, 12(6), e731. <https://doi.org/10.1002/wcc.731>
- Manepalli, A., Singh, A., Mudigonda, M., White, B., & Albert, A. (2020). Generalization properties of machine learning based weather model downscaling. In *Proceedings of the international conference on learning representations* (pp. 1–8).
- Meng, X., & Karniadakis, G. E. (2020). A composite neural network that learns from multi-fidelity data: Application to function approximation and inverse PDE problems. *Journal of Computational Physics*, 401, 109020. <https://doi.org/10.1016/j.jcp.2019.109020>
- Muñoz Sabater, J., Dutra, E., Agustí-Panareda, A., Albergel, C., Arduini, G., Balsamo, G., et al. (2021). Era5-land: A state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data*, 13(9), 4349–4383. <https://doi.org/10.5194/essd-13-4349-2021>
- Orr, H. G., Ekström, M., Charlton, M. B., Peat, K. L., & Fowler, H. J. (2021). Using high-resolution climate change information in water management: A decision-makers' perspective. *Philosophical Transactions of the Royal Society A: Mathematical, Physical & Engineering Sciences*, 379(2195), 20200219. <https://doi.org/10.1098/rsta.2020.0219>
- Pan, B., Hsu, K., AghaKouchak, A., & Sorooshian, S. (2019). Improving precipitation estimation using convolutional neural network. *Water Resources Research*, 55(3), 2301–2321. <https://doi.org/10.1029/2018WR024090>
- Pan, X., Zhu, X., Yang, Y., Cao, C., Zhang, X., & Shan, L. (2018). Applicability of downscaling land surface temperature by using normalized difference sand index. *Scientific Reports*, 8(1), 1–14. <https://doi.org/10.1038/s41598-018-27905-0>
- Pathak, J., Subramanian, S., Harrington, P., Raja, S., Chattopadhyay, A., Mardani, M., et al. (2022). FourCastNet: A global data-driven high-resolution weather model using adaptive Fourier neural operators. arXiv. <https://doi.org/10.48550/ARXIV.2202.11214>
- Pour, S. H., Shahid, S., Chung, E.-S., & Wang, X.-J. (2018). Model output statistics downscaling using support vector machine for the projection of spatial and temporal changes in rainfall of Bangladesh. *Atmospheric Research*, 213, 149–162. <https://doi.org/10.1016/j.atmosres.2018.06.006>
- Qian, Y., Chakraborty, T., Li, J., Li, D., He, C., Sarangi, C., et al. (2022). Urbanization impact on regional climate and extreme weather: Current understanding, uncertainties, and future research directions. *Advances in Atmospheric Sciences*, 39, 819–860.
- Rockel, B., & Woth, K. (2007). Extremes of near-surface wind speed over Europe and their future changes as estimated from an ensemble of RCM simulations. *Climatic Change*, 81(1), 267–280. <https://doi.org/10.1007/s10584-006-9227-y>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In N. Navab, J. Hornegger, W. M. Wells, & A. F. Frangi (Eds.), *Medical image computing and computer-assisted intervention – MICCAI 2015* (pp. 234–241). Springer International Publishing.
- Shainer, G., Liu, T., Michalakas, J., Liberman, J., Layton, J., Celebioglu, O., et al. (2009). Weather research and forecast (WRF) model performance and profiling analysis on advanced multi-core HPC clusters. In *10th LCI ICHPC*.
- Shi, X. (2020). Enabling smart dynamical downscaling of extreme precipitation events with machine learning. *Geophysical Research Letters*, 47(19), e2020GL090309. <https://doi.org/10.1029/2020GL090309>
- Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Liu, Z., Berner, J., et al. (2019). *A description of the advanced research WRF model version 4* (Tech. Rep.). National Center for Atmospheric Research. <https://doi.org/10.5065/1dfh-6p97>
- Song, F., Zhang, G. J., Ramanathan, V., & Leung, L. R. (2022). Trends in surface equivalent potential temperature: A more comprehensive metric for global warming and weather extremes. *Proceedings of the National Academy of Sciences*, 119(6), e2117832119. <https://doi.org/10.1073/pnas.2117832119>
- Stengel, K., Glaws, A., Hettinger, D., & King, R. N. (2020). Adversarial super-resolution of climatological wind and solar data. *Proceedings of the National Academy of Sciences*, 117(29), 16805–16815. <https://doi.org/10.1073/pnas.1918964117>
- Teutschbein, C., & Seibert, J. (2010). Regional climate models for hydrological impact studies at the catchment scale: A review of recent modeling strategies. *Geography Compass*, 4(7), 834–860. <https://doi.org/10.1111/j.1749-8198.2010.00357.x>
- Tobin, I., Jerez, S., Vautard, R., Thais, F., van Meijgaard, E., Prein, A., et al. (2016). Climate change impacts on the power generation potential of a European mid-century wind farms scenario. *Environmental Research Letters*, 11(3), 034013. <https://doi.org/10.1088/1748-9326/11/3/034013>
- Von Storch, H. (1999). On the use of “inflation” in statistical downscaling. *Journal of Climate*, 12(12), 3505–3506. [https://doi.org/10.1175/1520-0442\(1999\)012<3505:OTUOI>2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012<3505:OTUOI>2.0.CO;2)
- Wang, J., Liu, Z., Foster, I., Chang, W., Kettimuthu, R., & Kotamarthi, V. R. (2021). Fast and accurate learned multiresolution dynamical downscaling for precipitation. *Geoscientific Model Development*, 14(10), 6355–6372. <https://doi.org/10.5194/gmd-14-6355-2021>
- Wang, J., Xue, P., Pringle, W., Yang, Z., & Qian, Y. (2022). Impacts of lake surface temperature on the summer climate over the great lakes region. *Journal of Geophysical Research: Atmospheres*, 127(11), e2021JD036231. <https://doi.org/10.1029/2021JD036231>
- Wang, Z., Chen, J., & Hoi, S. C. H. (2021). Deep learning for image super-resolution: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(10), 3365–3387. <https://doi.org/10.1109/TPAMI.2020.2982166>
- Watson, C. D., Wang, C., Lynar, T., & Weldemariam, K. (2020). Investigating two super-resolution methods for downscaling precipitation: ESRGAN and CAR. arXiv. <https://doi.org/10.48550/ARXIV.2012.01233>
- Wen, G., Li, Z., Azizzadenesheli, K., Anandkumar, A., & Benson, S. M. (2022). U-FNO—An enhanced Fourier neural operator-based deep-learning model for multiphase flow. *Advances in Water Resources*, 163, 104180. <https://doi.org/10.1016/j.advwatres.2022.104180>

- Xian, Y., Lampert, C. H., Schiele, B., & Akata, Z. (2019). Zero-shot learning—A comprehensive evaluation of the good, the bad and the ugly. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *41*(9), 2251–2265. <https://doi.org/10.1109/TPAMI.2018.2857768>
- Xu, W., Zhang, Z., Long, Z., & Qin, Q. (2021). Downscaling SMAP soil moisture products with convolutional neural network. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *14*, 4051–4062. <https://doi.org/10.1109/JSTARS.2021.3069774>
- Xue, P., Malanotte-Rizzoli, P., Wei, J., & Eltahir, E. A. (2020). Coupled ocean-Atmosphere modeling over the maritime continent: A review. *Journal of Geophysical Research: Oceans*, *125*(6), e2019JC014978. <https://doi.org/10.1029/2019jc014978>
- Yang, Z., Dominguez, F., & Zeng, X. (2019). Large and local-scale features associated with heat waves in the United States in reanalysis products and the NARCCAP model ensemble. *Climate Dynamics*, *52*(3), 1883–1901. <https://doi.org/10.1007/s00382-018-4414-x>
- Zakšek, K., & Oštir, K. (2012). Downscaling land surface temperature for urban heat island diurnal cycle analysis. *Remote Sensing of Environment*, *117*, 114–124. (Remote Sensing of Urban Environments). <https://doi.org/10.1016/j.rse.2011.05.027>
- Zhang, Y., Held, I., & Fueglistaler, S. (2021). Projections of tropical heat stress constrained by atmospheric dynamics. *Nature Geoscience*, *14*(3), 133–137. <https://doi.org/10.1038/s41561-021-00695-3>