Exploring the Potential of Long Short-Term Memory Networks for Improving Understanding of Continental- and Regional-Scale Snowpack Dynamics

Yuan-Heng Wang, Hoshin V. Gupta, Xubin Zeng, Guo-Yue Niu

Department of Hydrology and Atmospheric Sciences, The University of Arizona, Tucson, AZ 85721, USA

Corresponding author: Yuan-Heng Wang (yhwang0730@email.arizona.edu)

Contents of this file

Text S1
Figures S1 to S9
Table S1

Introduction

This Supporting Information provides 1 supplementary text, 9 supplementary figures and 1 supplementary table to support the discussions in the main text. The contents of these supplementary materials are as follows.

Text S1. Mathematical formulation and the introduction of Long Short-term Memory Network used in this study

Figure S1. Schematic illustration of the architecture of a standard LSTM cell as defined by supplementary materials

Figure S2. Conceptual schematic of the processes and associated parameters represented by the SNOW17 model.

Figure S3. Spatial map indicating skill of the *LSTM-A-CONUS-6M* model (trained on Pixel Set A) when tested on two of the independent testing pixels sets *from Pixel Set B*

Figure S4. Aggregate performance of the *LSTM-A-CONUS-6M* network, evaluated on the training, evaluation and testing pixels from *Pixel Set B*

Figure S5. Comparisons between UA-SWE observations and SWE predicted by the LSTM and SNOW17 models at pixels selected from each of the five regions.

Figure S6. Comparisons between UA-SWE observations and SWE predicted by the *LSTM-A-CONUS-6M* model at pixels selected from each of the five regions.

Figure S7. Aggregate performance of the Regional trained LSTM networks compared to CONUS-wide trained LSTM networks using Pixel set B, evaluated over 5,000 testing pixels from *Pixel Set B*

Figure S8. The results of transfer learning when the transferred 6ME regional LSTM networks are benchmarked against their corresponding local-regional LSTM networks, regional SNOW17 models and pixel-wise SNOW17 models.

Figure S9. The results of transfer learning when the transferred regional SNOW17 models are benchmarked against their corresponding local regional SNOW17 models, and pixel-wise SN17 models.

Table S1. The results of KGEss skill for *Figure S5*

Text S1.

An LSTM network is a type of recurrent neural network that includes memory cells that have the ability to store information over long periods of time. As shown in Figure S1, the network contains cell states and three gating operations (input, forget, output). Here, we summarize the mathematical formulation of the LSTM network.

Given an input sequence $x = [x[1], x[2] \dots \dots, x[T]]$ with T time steps, where each element $x[t]$ is a vector containing input features (model inputs) at time step t ($1 \leq t \leq T$), Equations (1) to (6) specify a single forward pass through the LSTM:

$$i[t] = \sigma(W_i x[t] + U_i h[t-1] + b_i) \tag{1}$$

$$f[t] = \sigma(W_f x[t] + U_f h[t-1] + b_f) \tag{2}$$

$$g[t] = tanh(W_g x[t] + U_g h[t-1] + b_g) \tag{3}$$

$$o[t] = \sigma(W_o x[t] + U_o h[t-1] + b_o) \tag{4}$$

$$c[t] = f[t] \odot c[t-1] + i[t] \odot g[t] \tag{5}$$

$$h[t] = o[t] \odot \tanh(c[t]) \tag{6}$$

where $i[t], f[t], o[t]$ are the input, forget and output gates respectively, $g[t]$ is the cell input, $x[t]$ is the network input at time step t ($1 \leq t \leq T$), and $h[t-1]$ is the recurrent input. The terms $c[t]$ and $c[t-1]$ indicate the cell states at the current and previous time step. At the first-time step, the hidden and cell states are initialized as vectors of zeros. The terms W, U and b are learnable parameters for each gate. The subscript refers to at which gate the particular weight matrix, or the bias vector is used. The sigmoid activation function $\sigma(\cdot)$ outputs a value between 0 and 1, while the hyperbolic tangent activation function $\tanh(\cdot)$ outputs a value between -1 and 1. The symbol $\odot$ indicates element-wise multiplication.

The values of the cell states can be modified by the forget gate $f[t]$, which can delete states. The cell update $g[t]$ can be interpreted as information that is added, while the input gate $i[t]$ controls into which cells new information is added. The output gate $o[t]$ controls which of the information stored in the cell states is output. Note that the cell states $c[t]$ characterize the memory of the system, and its characteristic of simple linear interactions with the remaining LSTM cells helps to prevent the problem of exploding or vanishing gradients during the back-propagation step of network training (Hochreiter and Schmidhuber, 1997).

The output of the final LSTM layer $h[t]$ is connected through a dense layer to a single output neuron, which computes the final output $y[t]$ prediction, as indicated by Equation 7:

$$y[t] = W_d h[t] + b_d \tag{7}$$

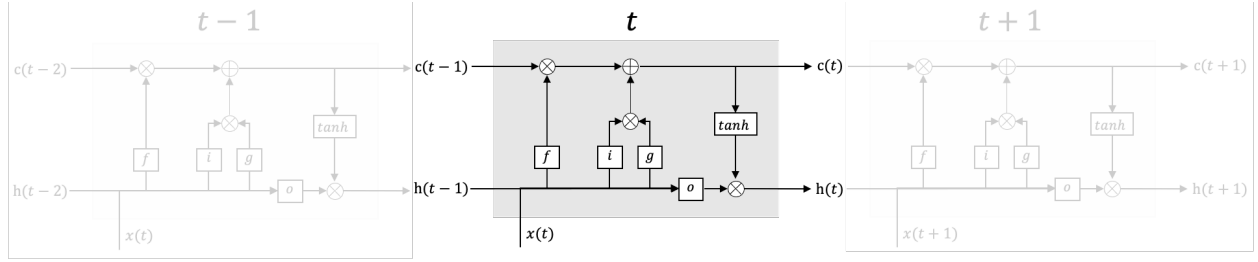where W_d and b_d are the learnable weight and bias of the output dense layer.

3

Figure S1. Schematic illustration of the architecture of a standard LSTM cell as defined by supplementary materials Eqs. (1)–(6). The symbols × and + denote element-wise multiplication and addition.
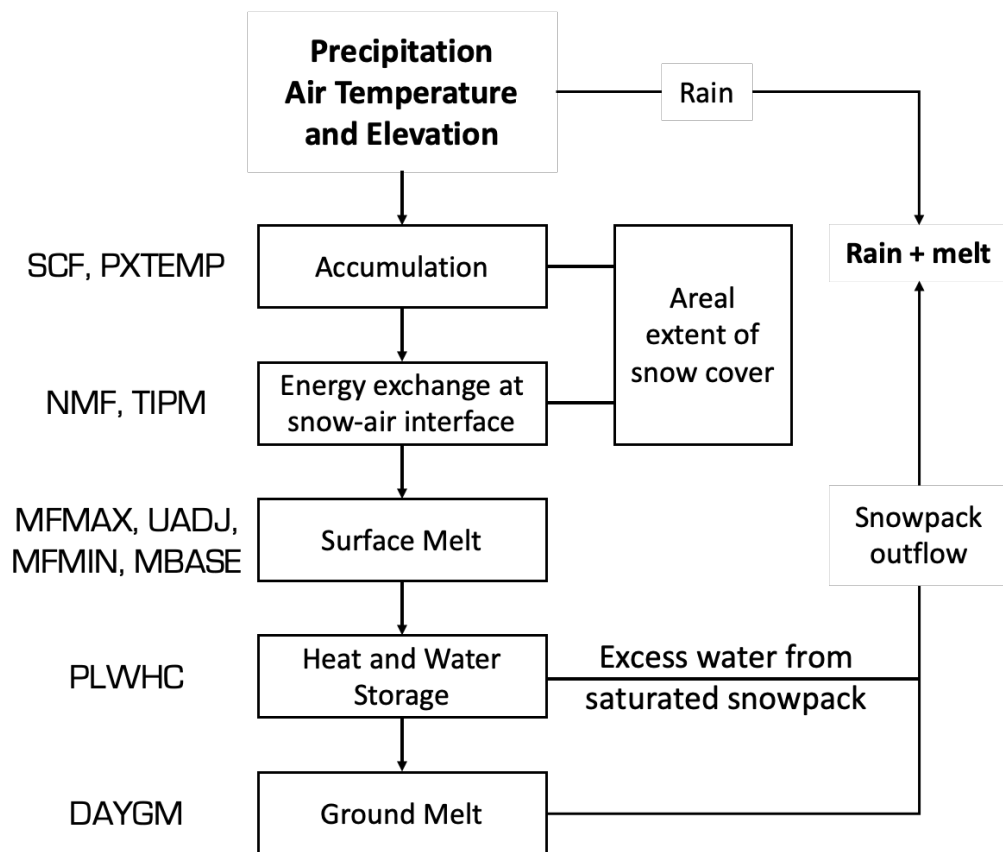
Figure S2. Conceptual schematic of the processes and associated parameters represented by the SN17 model. Inputs and outputs are highlighted in bold. Illustration derived from *He et al., (2011b)*
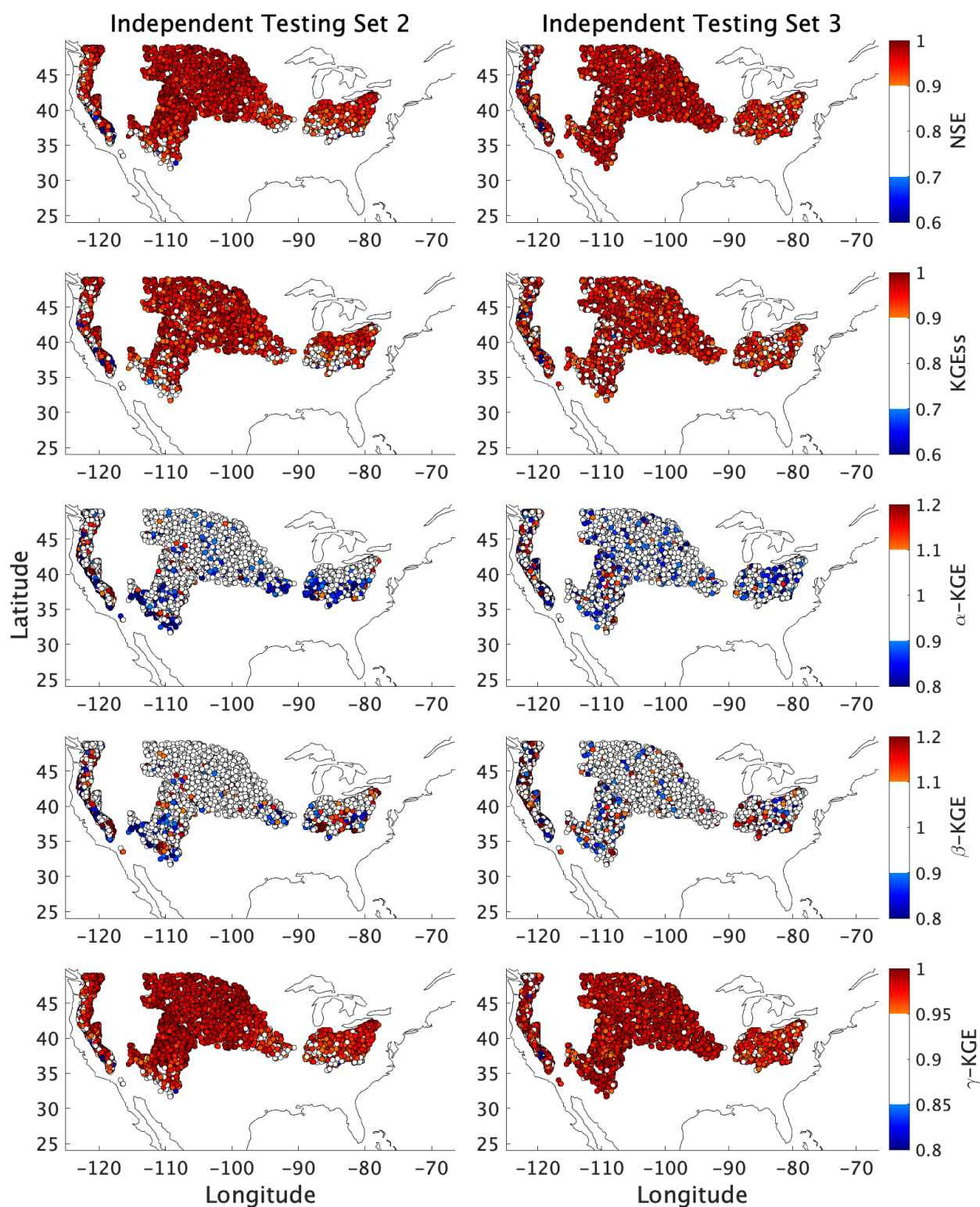
Figure S3. Spatial map indicating skill of the *LSTM-A-CONUS-6M* model (trained on *Pixel Set A*) when tested on two of the independent testing pixel sets from *Pixel Set B* (Results for independent test set 1 appear in the main text)
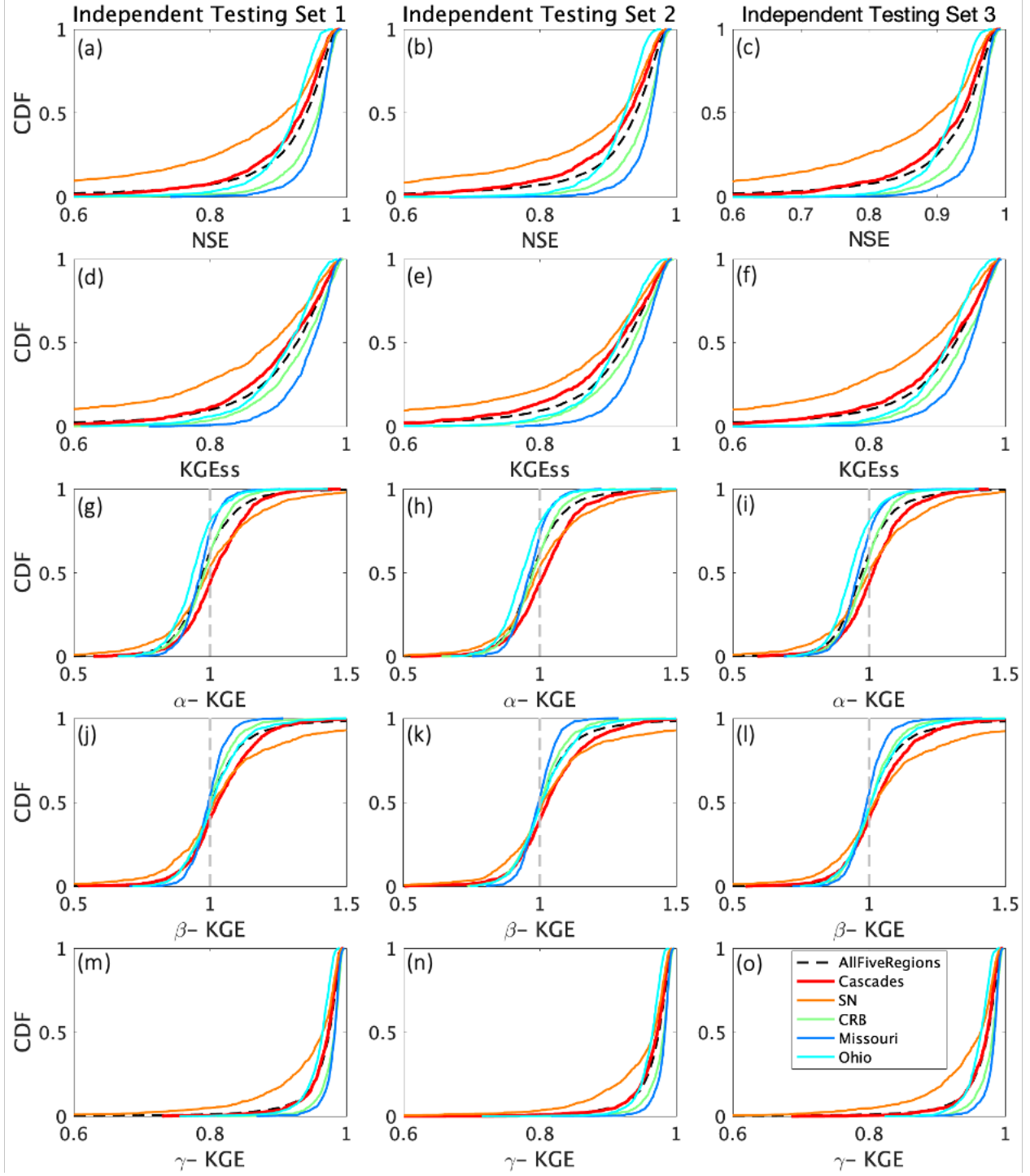
Figure S4. Aggregate performance of the *LSTM-A-CONUS-6M* network, evaluated on the training, evaluation and testing pixels from *Pixel Set B*
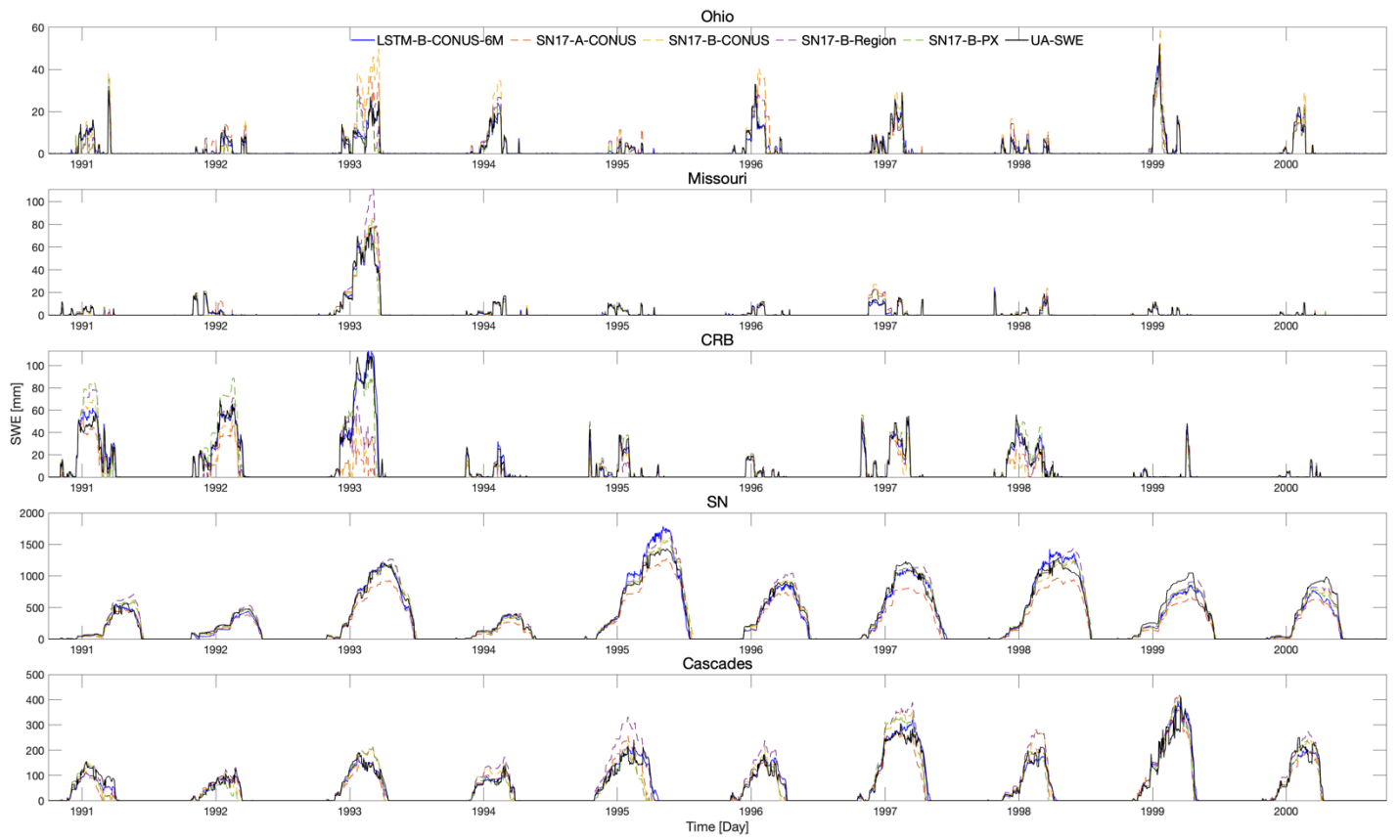
Figure S5. Comparisons between UA-SWE observations (solid black) and SWE predicted by the LSTM (solid blue) and SN17 models (dash lines) at pixels selected from each of the five regions. The pixels are from independent test set 1 (*pixel set B*) and represent locations corresponding to the 95th percentile of KGEss performance for the *LSTM-A-CONUS-6M* model. The corresponding KGEss skill for each of the models is listed in *Table S1*. The results only shown from WY1991 to WY2000.
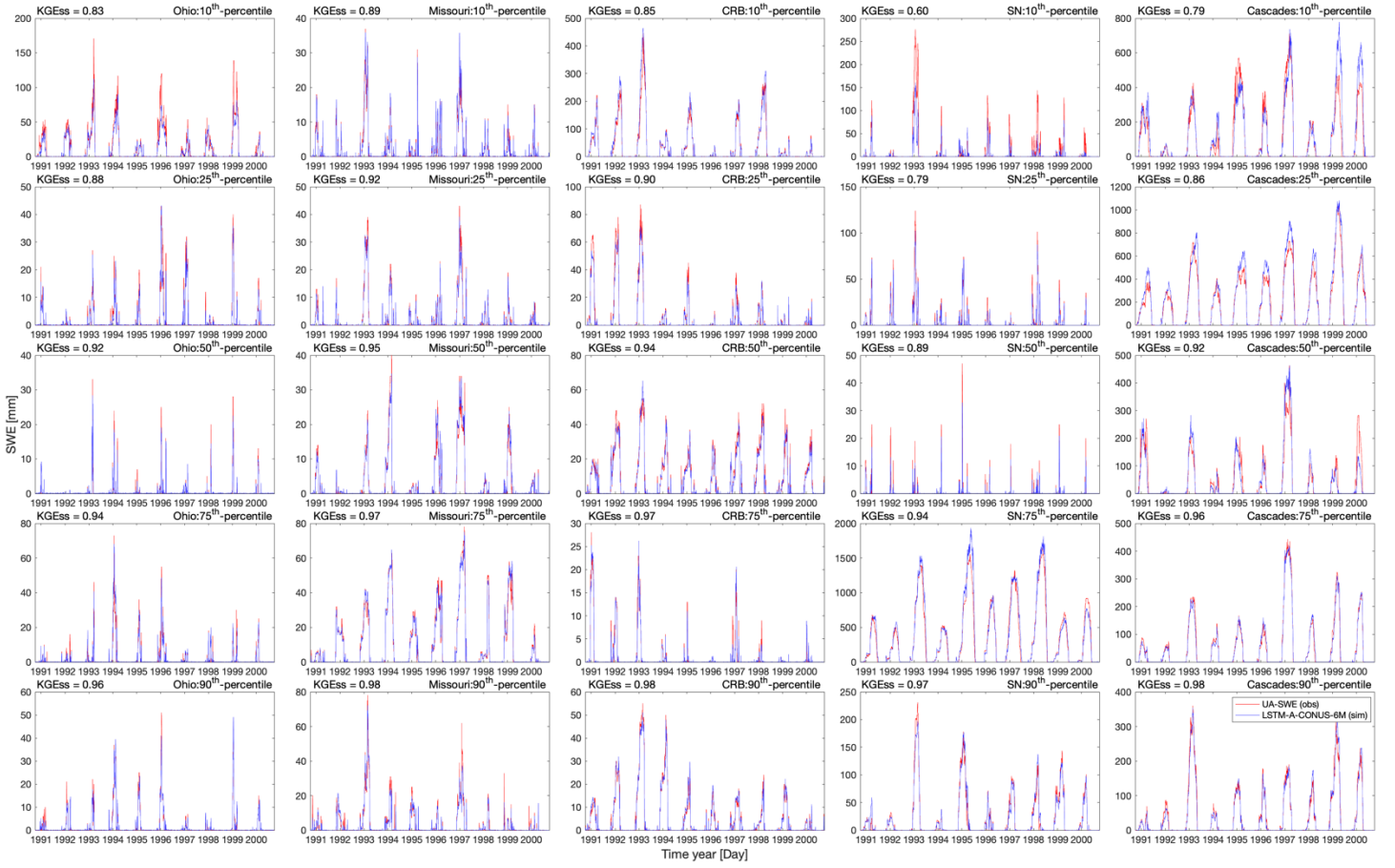
Figure S6. Comparisons between UA-SWE observations (red) and SWE predicted by the *LSTM-A-CONUS-6M* model (blue) at pixels selected from each of the five regions. The pixels are from independent test set 1 (*pixel set B*) and represent locations corresponding to the 10th, 25th, 50th, 75th, and 90th percentiles of KGEss performance for the model. The results only shown from WY1991 to WY2000.
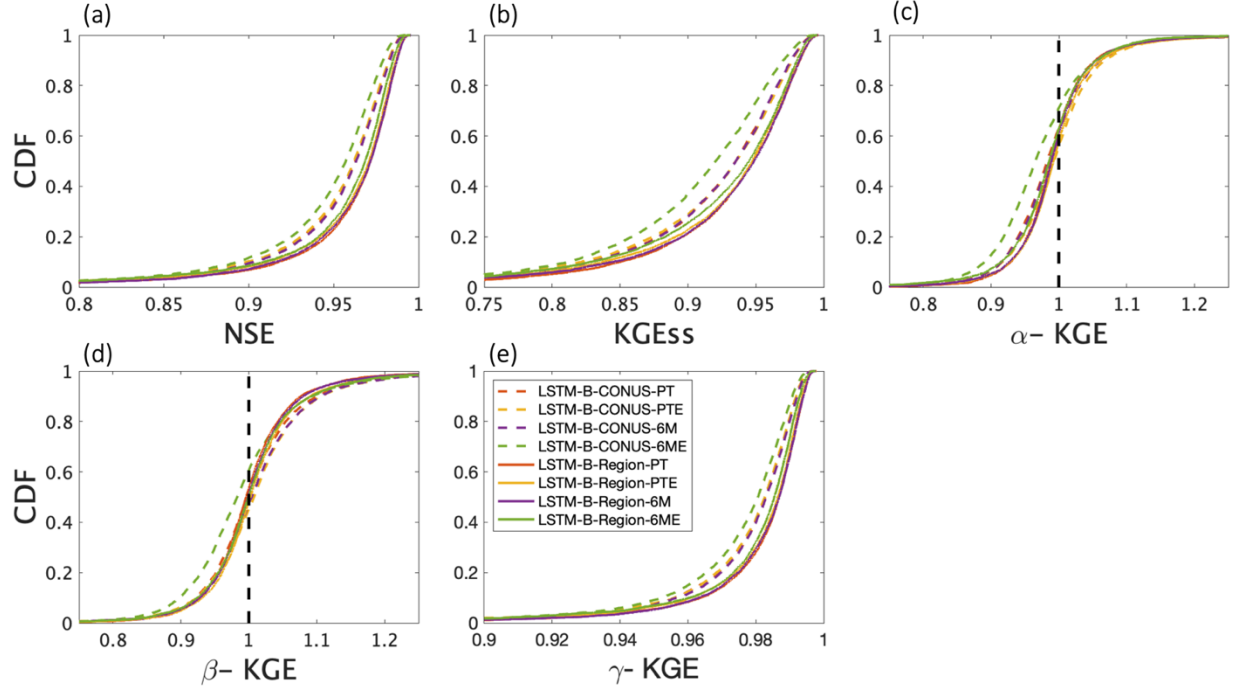
Figure S7. Aggregate performance of the Regional trained LSTM networks compared to CONUS-wide trained LSTM networks using Pixel set B, evaluated over 5,000 testing pixels from *Pixel Set B*
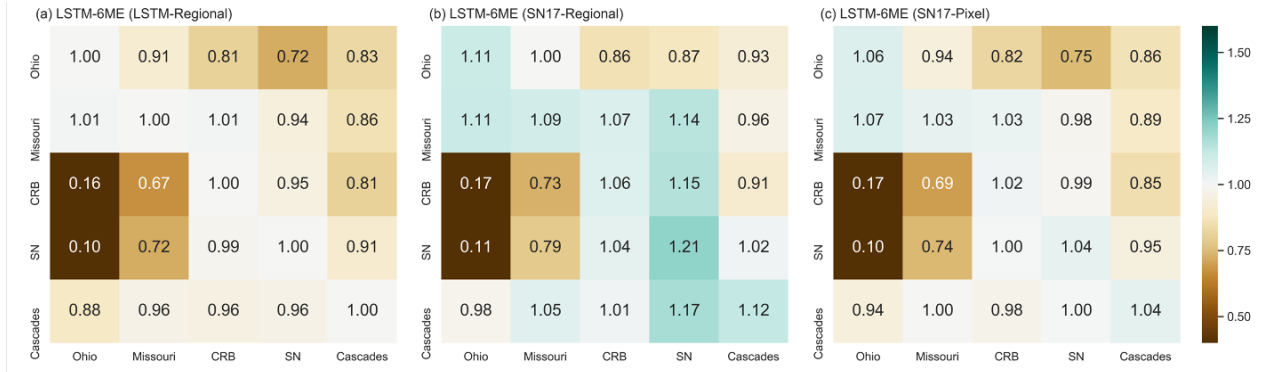
Figure S8. The results of transfer learning when the transferred 6ME regional LSTM networks are benchmarked against their corresponding local-regional LSTM networks, regional SN17 models and pixel-wise SN17 models.
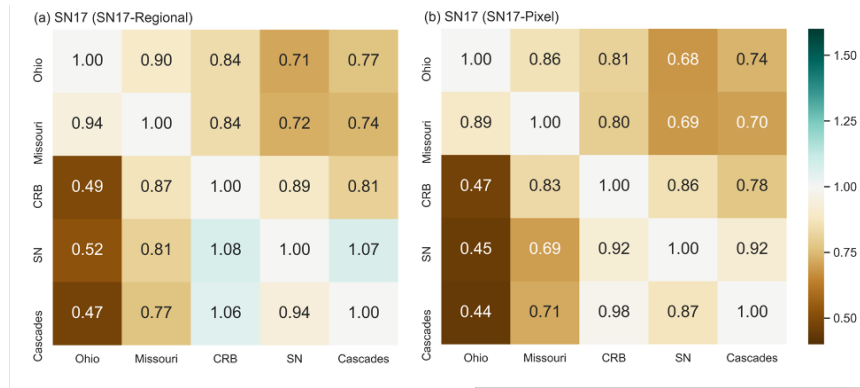
Figure S9. The results of transfer learning when the transferred regional SN17 models are benchmarked against their corresponding local regional SN17 models, and pixel-wise SN17 models.

Table S1. The results of KGEss skill for Figure S5 which Comparing the UA-SWE observations and SWE predicted by the LSTM and SNOW17 models at pixels selected from each of the five regions.

Models/Regions	Ohio	Missouri	CRB	SN	Cascades
LSTM-B-CONUS-6M	0.97	0.98	0.99	0.98	0.98
SN17-A-CONUS	0.88	0.94	0.62	0.81	0.88
SN17-B-CONUS	0.68	0.85	0.65	0.97	0.92
SN17-B-Region	0.90	0.72	0.86	0.90	0.79
SN17-B-PX	0.81	0.96	0.86	0.95	0.92

References.

Hochreiter, S. and Schmidhuber, J., 1997. Long short-term memory. *Neural computation*, 9(8), pp.1735-1780. https://doi.org/10.1162/neco.1997.9.8.1735

He, M., Hogue, T.S., Franz, K.J., Margulis, S.A. and Vrugt, J.A., 2011a. Characterizing parameter sensitivity and uncertainty for a snow model across hydroclimatic regimes. *Advances in Water Resources, 34(1), pp.114-127.* https://doi.org/10.1016/j.advwatres.2010.10.002

He, M., Hogue, T.S., Franz, K.J., Margulis, S.A. and Vrugt, J.A., 2011b. Corruption of parameter behavior and regionalization by model and forcing data errors: A Bayesian example using the SNOW17 model. *Water Resources Research, 47(7).* https://doi.org/10.1029/2010WR009753