Deep Learning for Spatial Interpolation of Rainfall Events

Yi (Victor) Wang[1], Antonia Sebastian[2], Seung Hee Kim[1], Thomas Piechota[1], and Menas Kafatos[1]

[1]Chapman University
[2]University of North Carolina at Chapel Hill

November 24, 2022

Abstract

Traditional deterministic and geostatistical methods for rainfall interpolation usually fall short of integration of data on a variety of variables. These omitted variables include seasonal variables such as time of year, topographic variables such as elevation, and/or remote sensing variables such as radar reflectivity. Meanwhile, poor quality in data on certain variables for some data points poses challenges to modelers who are using machine learning approaches to estimate rainfall amounts for locations without gauge measurements. To overcome these limitations, this presentation introduces a novel deep learning-based approach to recreate rainfall histories for large geographic areas with a high spatio-temporal resolution. The proposed approach enables integration of data on a variety of variables by adopting a multi-layer perceptron modeling framework. The introduction of binary variables on data quality as additional input variables resolves the issue of unequal data quality for different data points. As a demonstration, historical records of rainfall at hourly and daily intervals recorded at 139 rain gauge stations in or close to Harris County, Texas, from 1986 to 2013 are used, along with other auxiliary variables, to train deep learning regression models to interpolate rainfall at surface level. Results of validation and recreated spatiotemporal distributions of rainfall indicate good performance of the proposed approach compared to both gauged and radar data. The final product of the proposed approach can be applied to other regions, with information on hindcast historical rainfall events, for pluvial flood risk analysis. The approach will assist researchers and policy specialists to validate hydrologic modeling as well as for training machine learning models to identify extreme rainfall events to facilitate early warning and emergency response.

GC42A-02: Deep Learning for Spatial Interpolation of Rainfall Events

Yi Victor Wang[1], Antonia Sebastian[2], Seung Hee Kim[1], Thomas Piechota[3], and Menas Kafatos[1]

[1]Center of Excellence in Earth Systems, Modeling and Observations, Chapman University

[2]Department of Earth, Marine and Environmental Sciences, University of North Carolina at Chapel Hill

[3]Schmid College of Science and Technology, Chapman University

2D simulation of rainfall events for hydrologic modeling and flood risk analysis requires empirical data on rainfall surfaces for model training

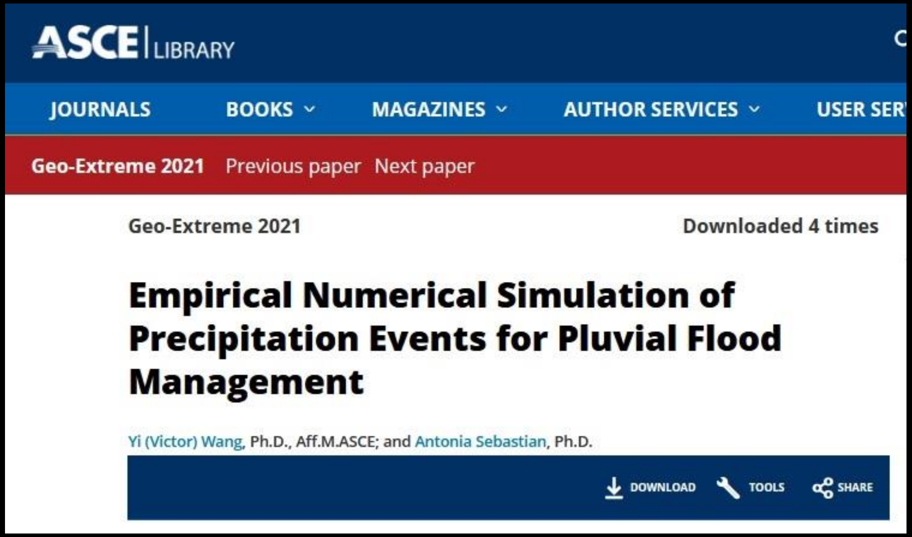
state of Texas as well as the entire United States. Future work will also extend our empirical methodology to simulate precipitation events for areas of different scales with two-dimensional modeling techniques. To achieve a two-dimensional simulation, more advanced quantitative approaches such as artificial intelligence will be necessary.

Products of radar data are good but not ground truth

Current radar data do not go far into the past

Many gauge stations have long records of rainfall

Spatial interpolation comes to the rescue!



Traditionally, we use deterministic and geostatistical methods to interpolate rainfall surfaces

Deterministic methods	Thiessen polygon	Inverse distance weighting	Spline interpolation
Geostatistical methods	Ordinary kriging	Universal kriging	Cokriging methods

Limitations

Tend to omit variables such as seasonal, topographic, and remote sensing variables

Can be affected by poor quality of data for individual timestamps

Interpolated rainfall surfaces look unnatural

To overcome the limitations of traditional methods, we propose a novel deep learning-based approach to interpolate rainfall surfaces

Output

Rainfall depth at a location in mm

Input

Rainfall depth at gauge station

Whether rainfall record is of good quality at gauge station

Latitude, longitude, and elevation of output location

Day in a year and hour in a day of rainfall record

Whether radar data is available

Image patch of radar reflectivity centered at the output location

For demonstration of proposed methodology, we use records of rainfall from gauge stations in or close to Harris County, Texas

Number of stations

139

Time period of data

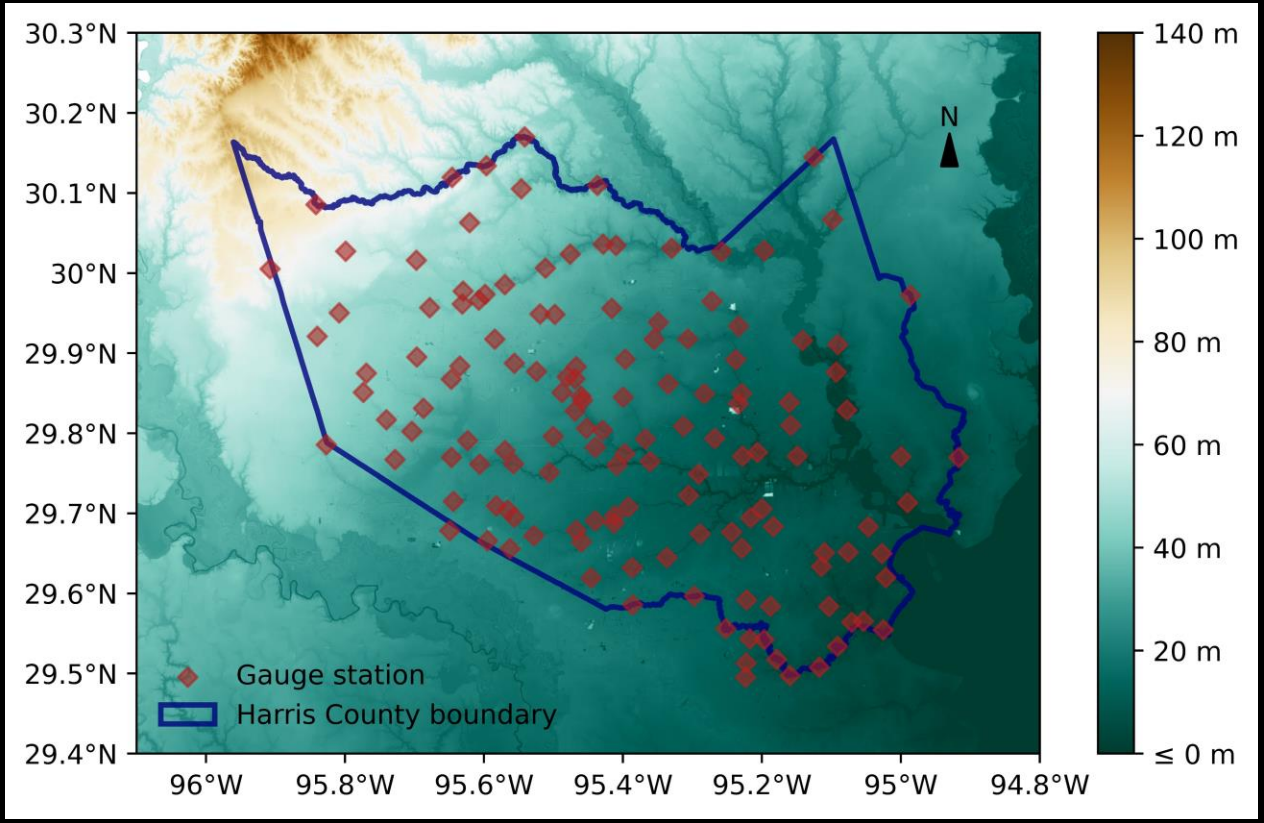
1986–2013

Temporal resolution

5 min aggregated at 1 hour

Stationarity examined by previous work

Wang and Sebastian 2021



For this presented pioneering work, we currently only use one year of radar data on reflectivity

Time period of data

January–December 1995

Temporal resolution

5 min

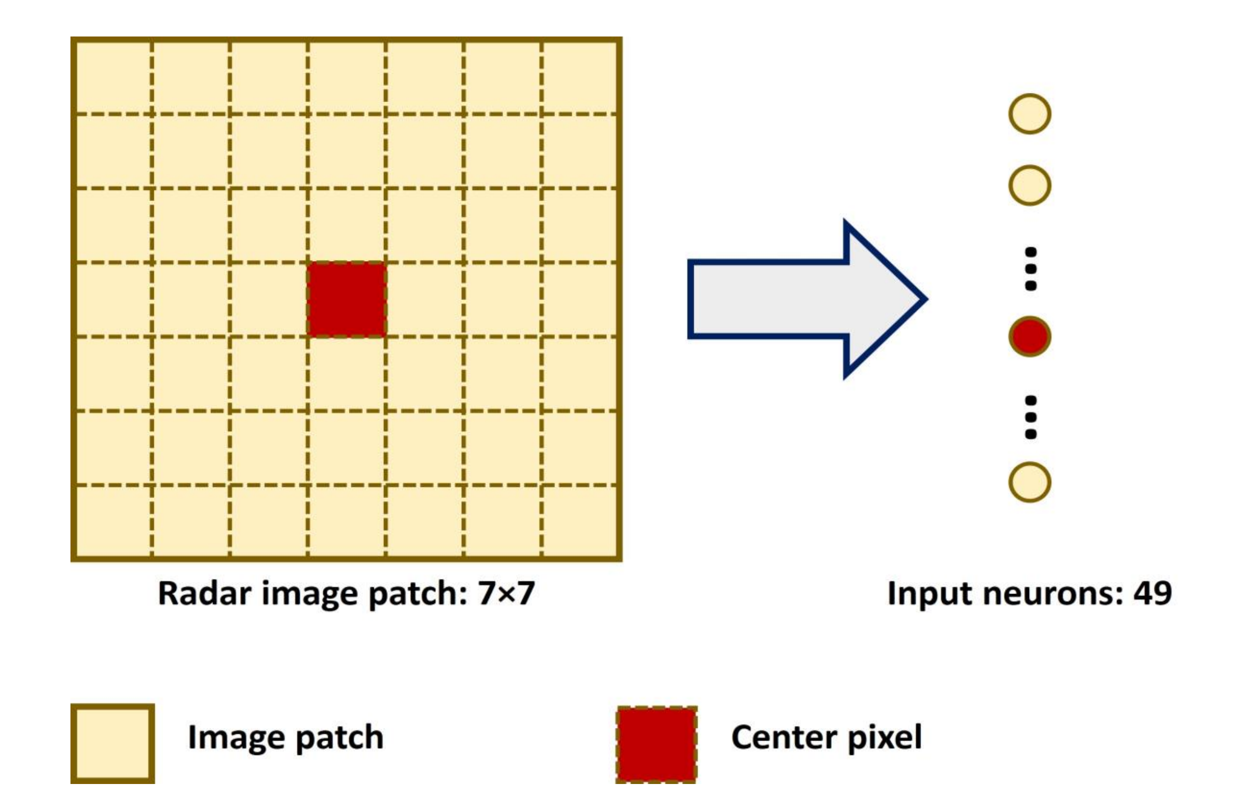
Spatial resolution

0.01° × 0.01°

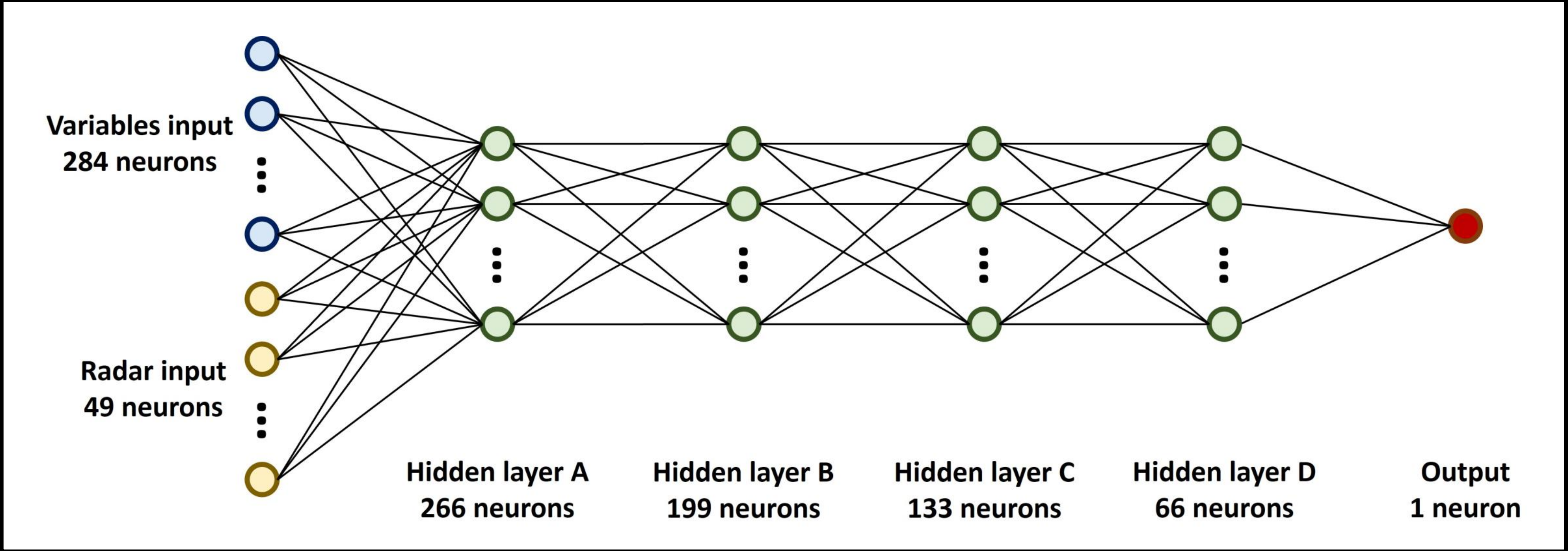
Image patch

Aggregated as pixel-wise medians

Converted into vectors



For deep learning regression, we adopt the architecture of a multi-layer perceptron (MLP) neural network with 4 hidden layers



To compare model performances, we train 10 MLPs and use the average of their predictions as the ensemble result

Loss function

Log-cosh error

Parametrization algorithm

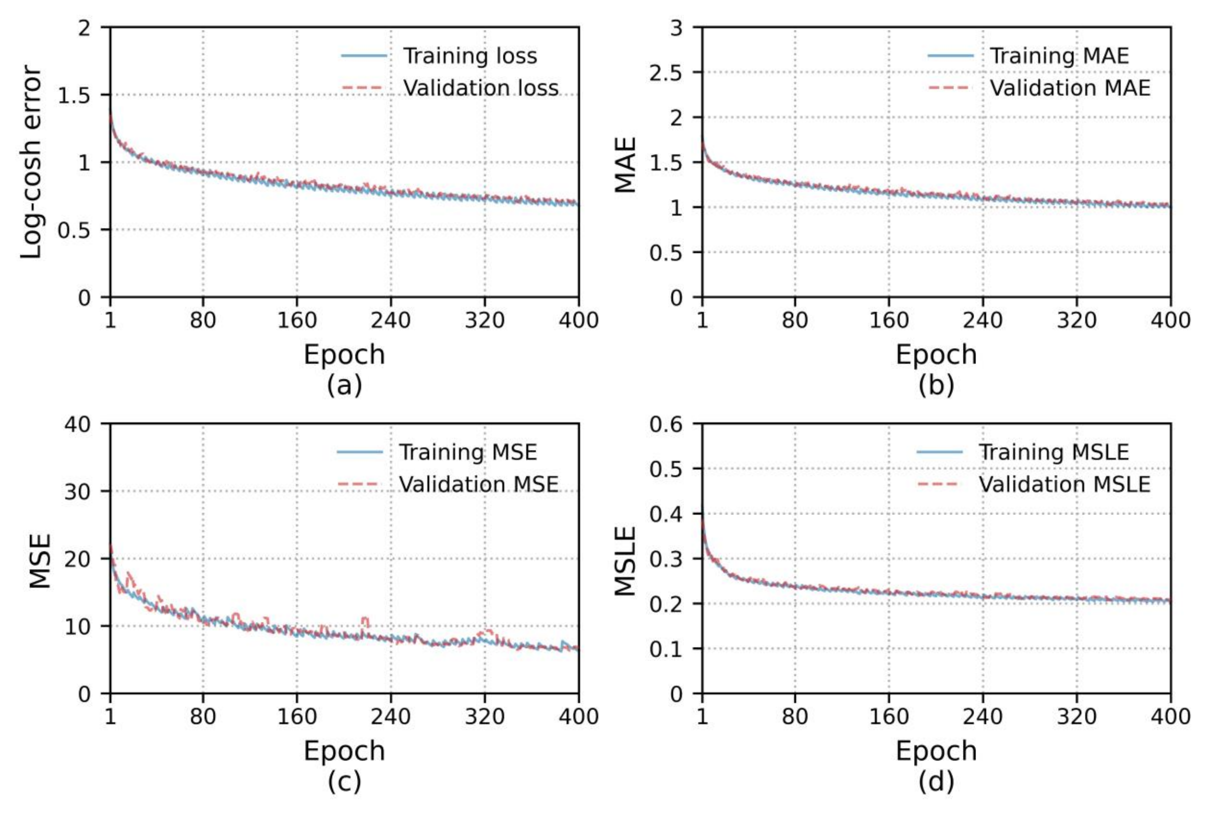
Adaptive moment estimation (Adam)

Shallow training

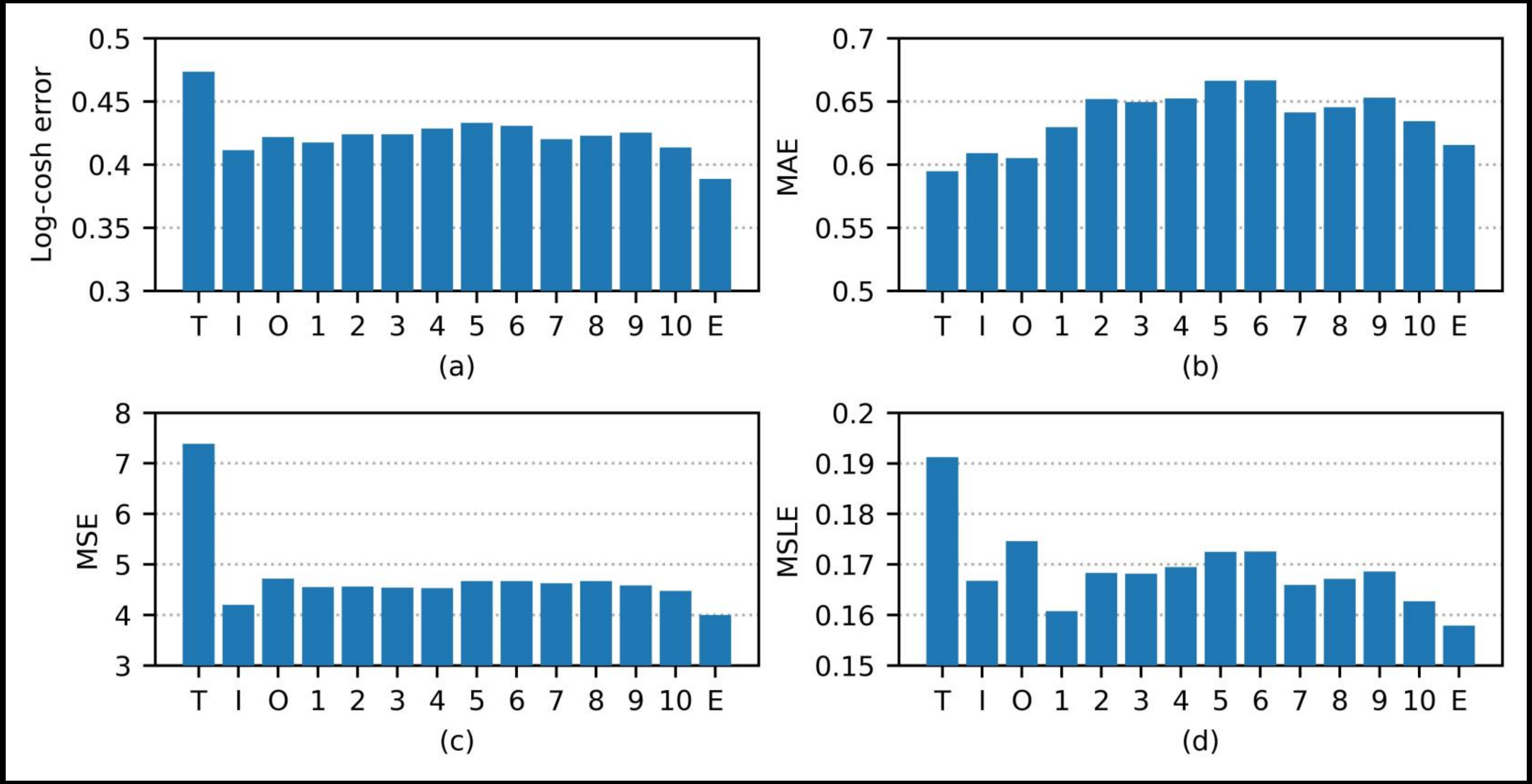
80 phases

Refresh training data for each phase

5 epochs in each phase



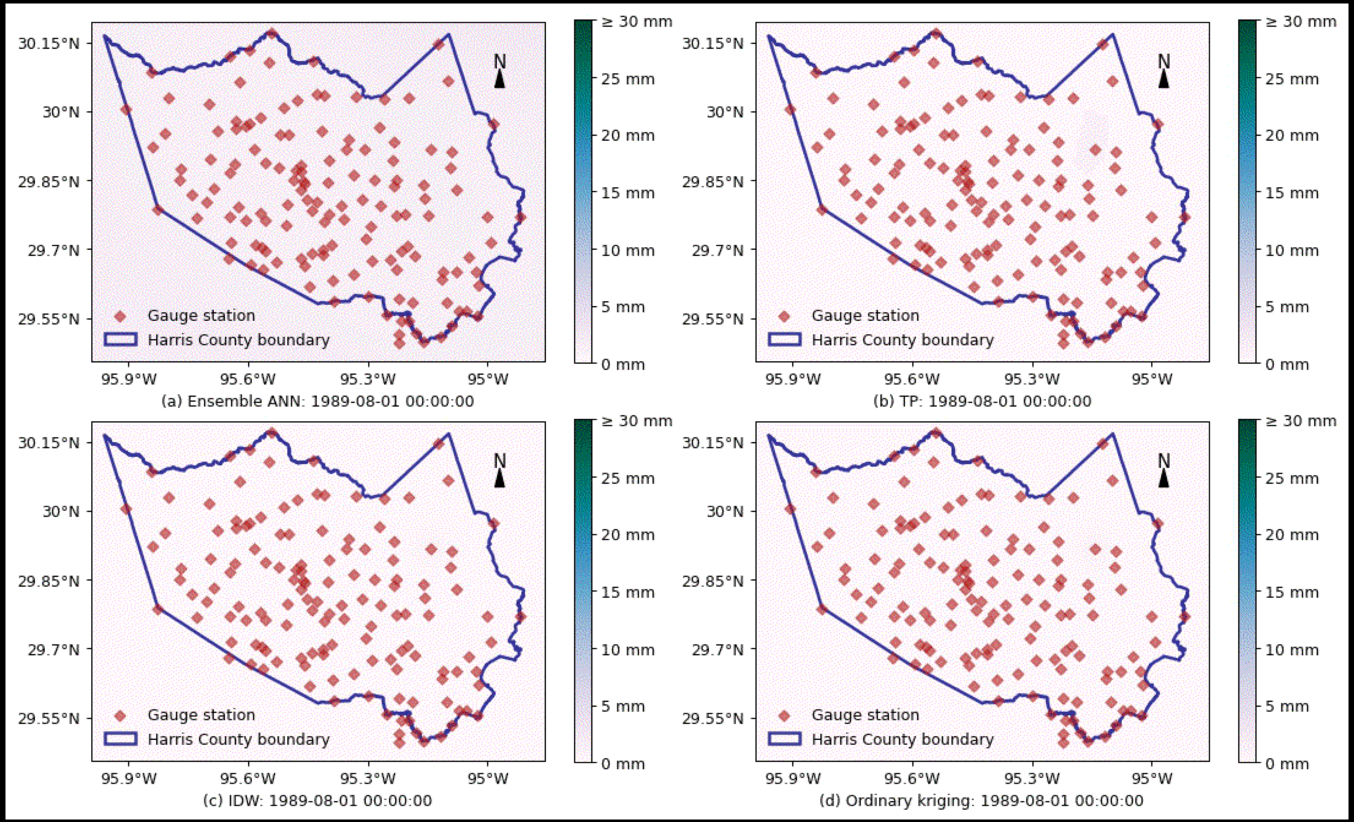
After model calibration, we can compare model performances in terms of loss metrics for model validation



We can also compare the interpolated rainfall surfaces for rainfall events with the deep learning and traditional methods

Spatial resolution

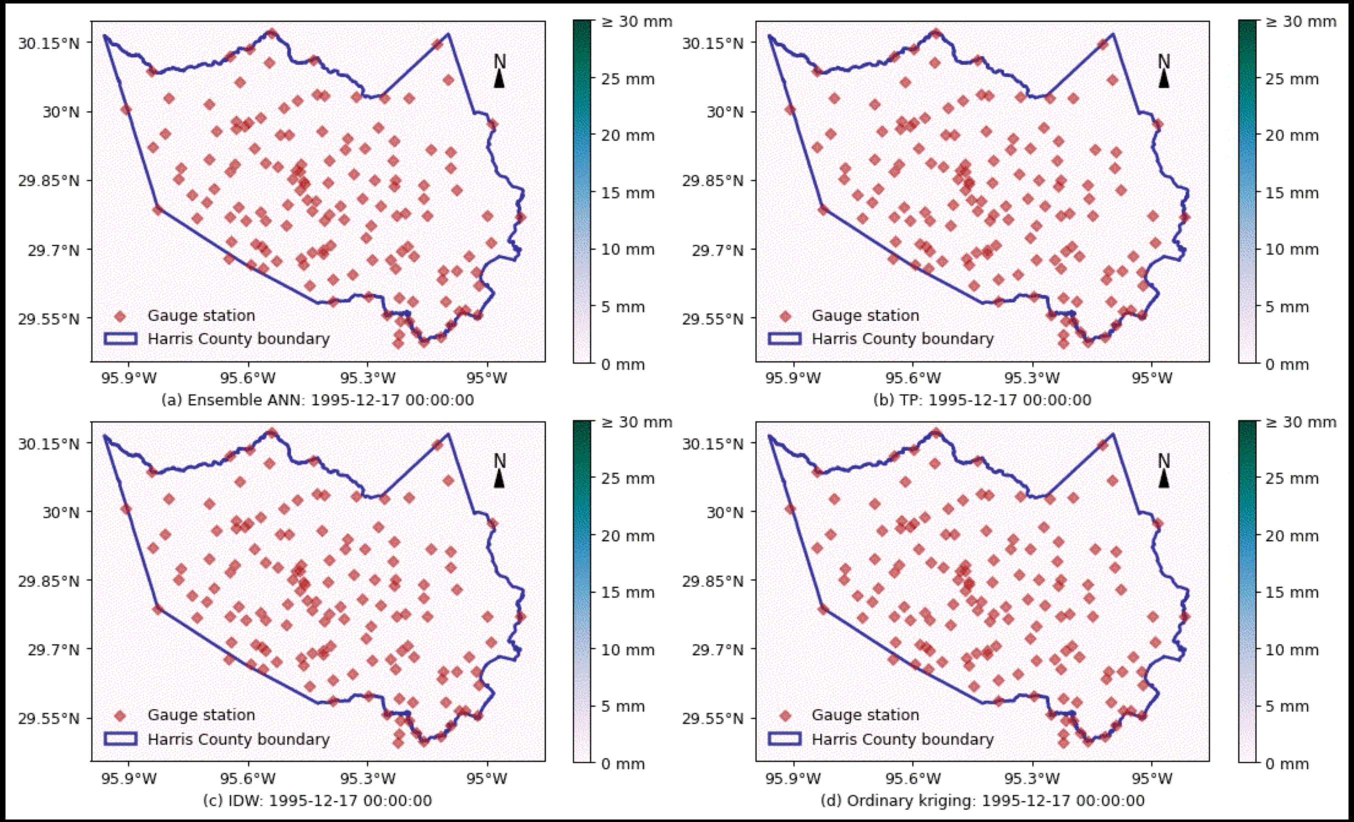
0.01° × 0.01°



In addition to events during periods without radar data, we can also have a look at the interpolated rain fall events during a period with radar data

Spatial resolution

0.01° × 0.01°



With the model results, we can conclude that the proposed deep learning-based approach has many merits

Validation shows smaller interpolation errors

Interpolated rainfall surfaces look more natural

Nicely handle the issue of missing and incorrect values

Allow inclusion of many auxiliary variables

Can be applied to other areas across the world

Provide an augmented reality of 2D rainfall history

Enhance pluvial flood risk analysis

Assist parametrization and validation of hydrologic models

Train learning models to identify extreme rainfall events

Given the encouraging results of the proposed methodology, future work needs to focus on several directions to improve the study

Examine if interpolated rainfall records underestimate or overestimate the averages compared to the records measured at gauge stations

Improve modeling to make sure that interpolated rainfall records follow the same probability distribution as the gauged records

Test if interpolated rainfall events have the same expected frequencies of exceeding key intensity measures, such as maximum rain rate and duration of event

ywang2@chapman.edu