World Water Congress

International Water Resources Association (IWRA)

Presentation Topic

Bias Correction of Daily Satellite-Based Precipitation Data Using Convolutional Neural Network Model

Xuan-Hien Le[1], Giha Lee[2]

[1] Ph.D., Disaster Prevention Emergency Management Institute, Kyungpook National University

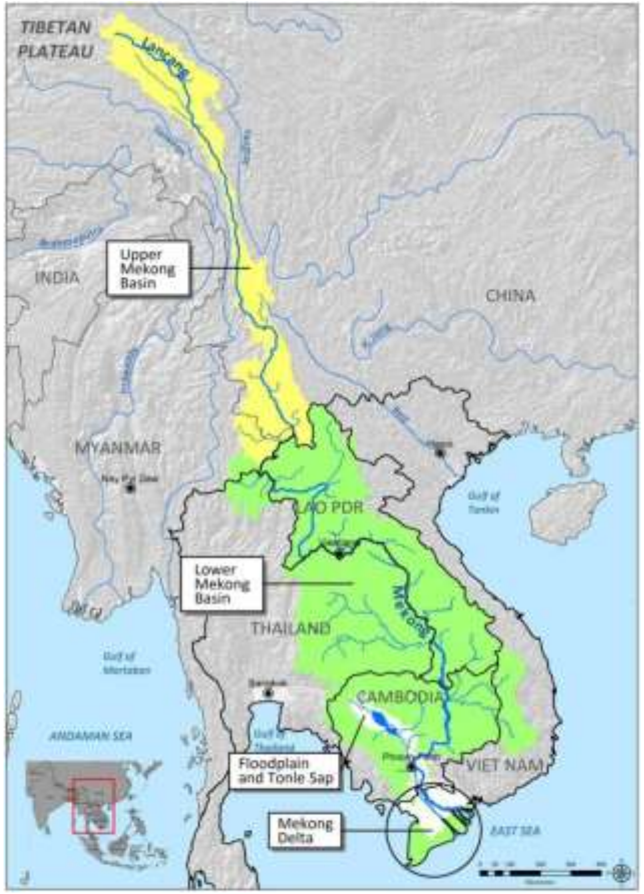
[2] Assoc. Prof., Department of Advanced Science and Technology Convergence, Kyungpook National University

Daegu, December 2021

Contents

1. Motivations

- Available rainfall data sources:
  - ✓ The ground-based rainfall gauges
  - ✓ Satellite-based precipitation products
- For river basins spanning many countries:
  - ✓ Sparse distribution of rainfall stations
  - ✓ Collecting data over a long period is a challenging task.
- **Require an up-to-date dataset for studies**

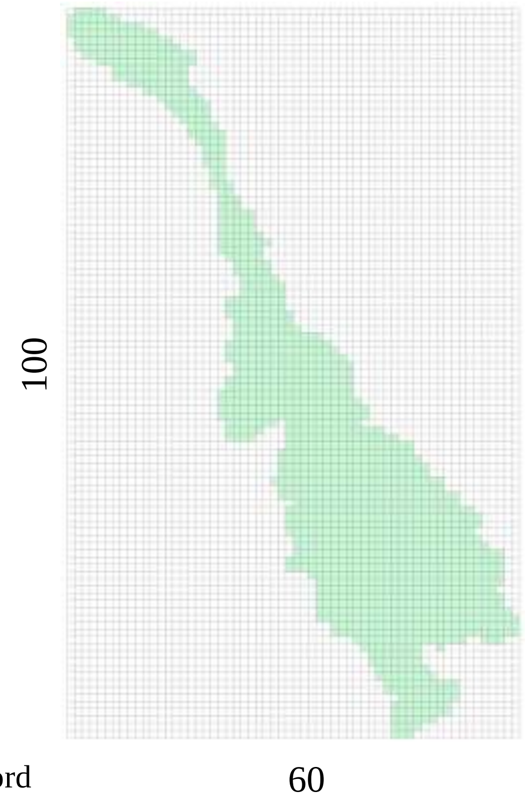


Mekong River Basin (*MRC, 2019*)

2. Study Area and Data

- **Target:** Produce a more up-to-date dataset than that of the APHRODITE product and sufficiently reliable for the Mekong basin studies.

Dataset	Version	Spatial/ Temporal resolution	Area coverage	Time coverage
APHRODITE	V1901	0.25°/daily	Monsoon Asia	1998-2015
PERSIANN	CDR	0.25°/daily	60S-60N	1983-2021
TRMM	3B42	0.25°/daily	50S-50N	1998-2020



APHRODITE: Asian Precipitation - Highly Resolved Observational Data Integration Towards Evaluation of Water Resources.

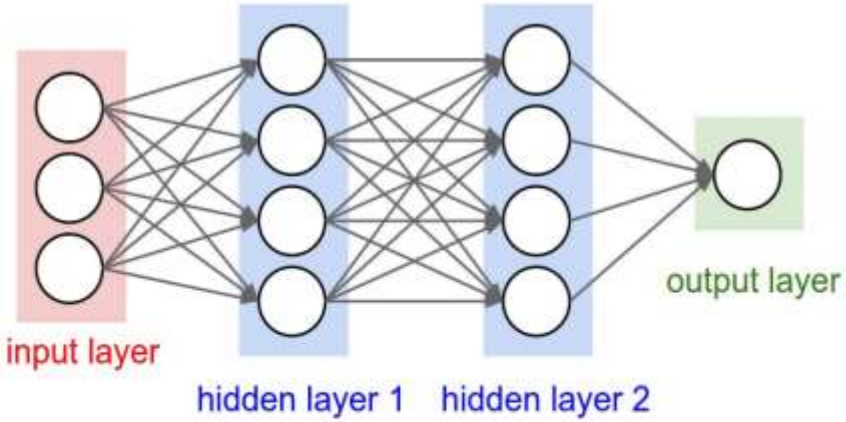
PERSIANN-CDR: Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks - Climate Data Record

TRMM: Tropical Rainfall Measuring Mission

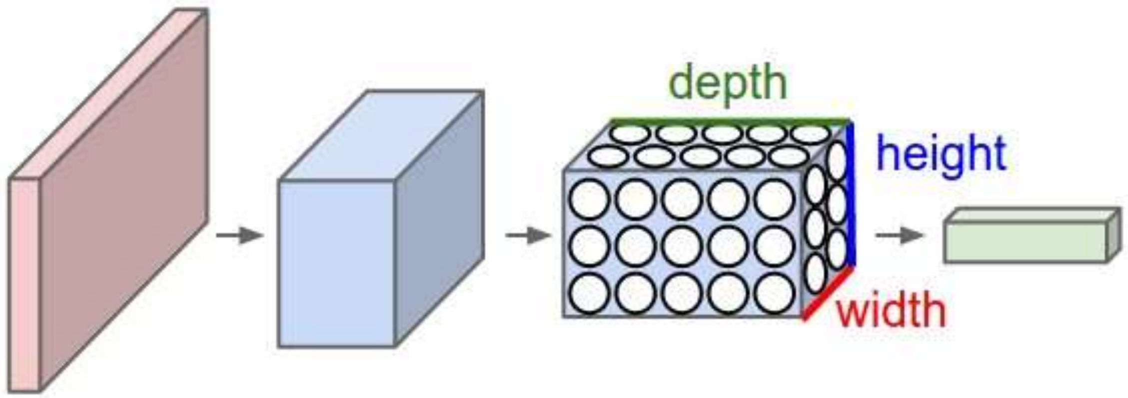
3. Methodology

➢ Convolutional Neural Network (CNN)

- ✓ CNN is very similar to ANN, consisting of neurons with learnable weights and biases.
- ✓ CNN arranges its neurons in three dimensions (**Width, Height, Depth**)
- ✓ CNN is composed of a **convolution layer** and a **pooling layer**



Artificial Neural Network - ANN

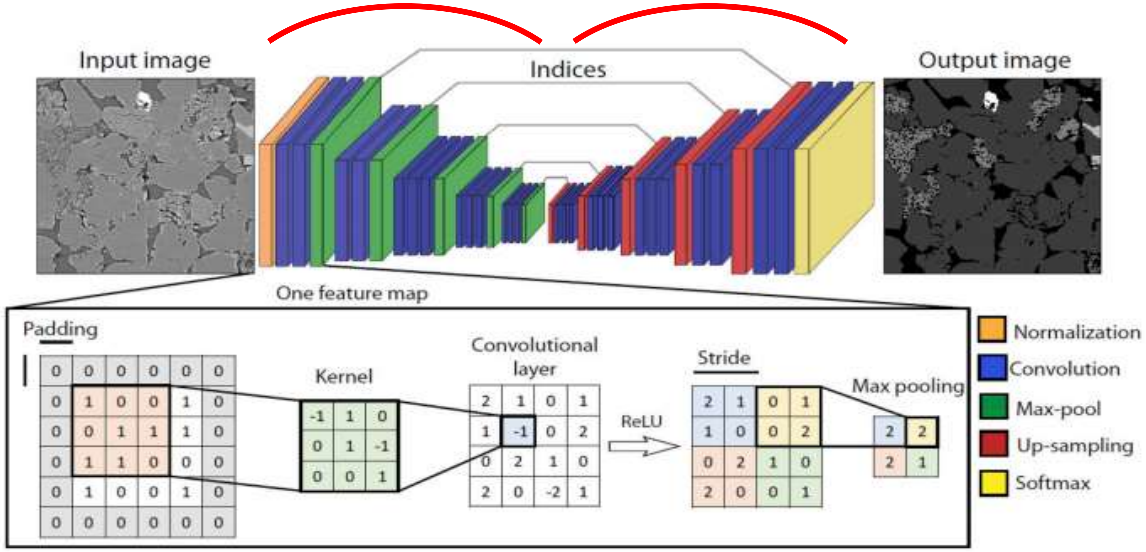


Convolutional Neural Network - CNN

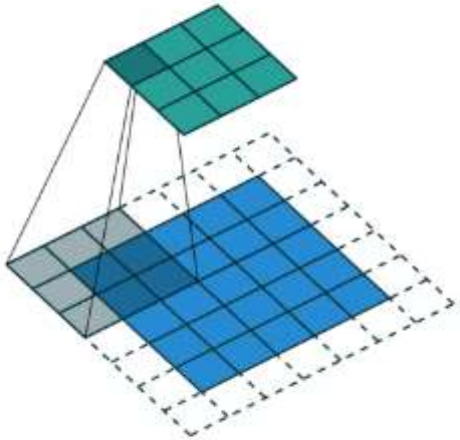
http://cs231n.github.io/convolutional-networks/

➢ Convolutional Autoencoder (CAE)

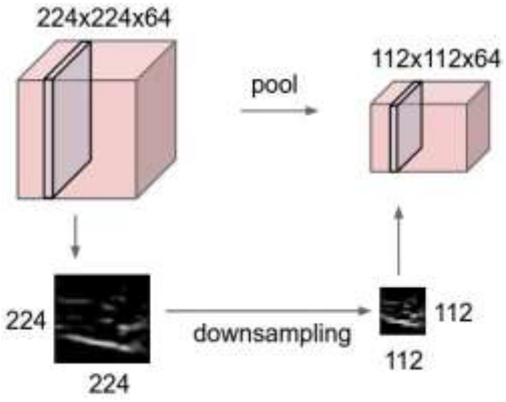
- ✓ Autoencoder structure consists of an **Encoding** process and a **Decoding** process
- ✓ ConvAE (CNN + autoencoder) receives input data in three dimensions, extracts each feature in the encoding process compresses it in a lower dimension, and reconstructs the original size through a decoding process



Convolutional Autoencoder structure



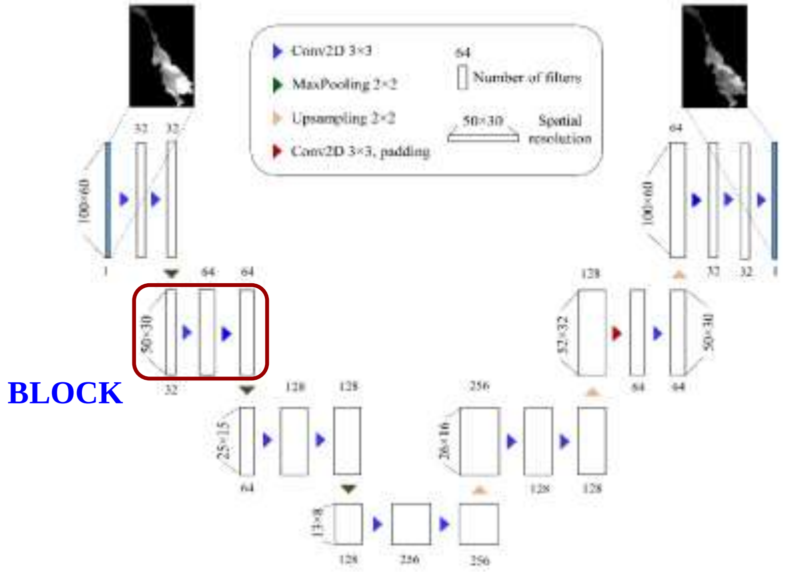
Convolution layer

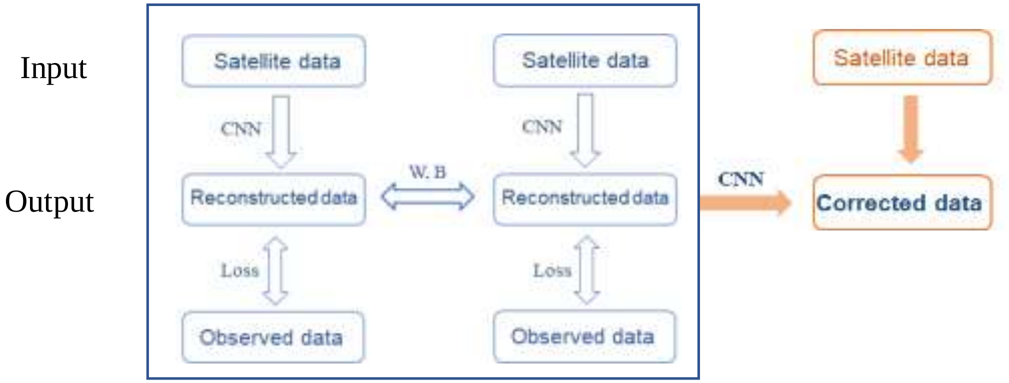


Pooling layer

(Karimpouli et al., 2019) *(Dr. Dataman, towards data science, 2019)* *http://cs231n.github.io/convolutional-networks/*

3. Proposed Model





Training + Validation (16 years)   Testing (2 years)

$$BLOCK = \begin{bmatrix} Conv2D \\ BatchNorm \\ ReLU \end{bmatrix} \times 2$$

Structure of a BLOCK

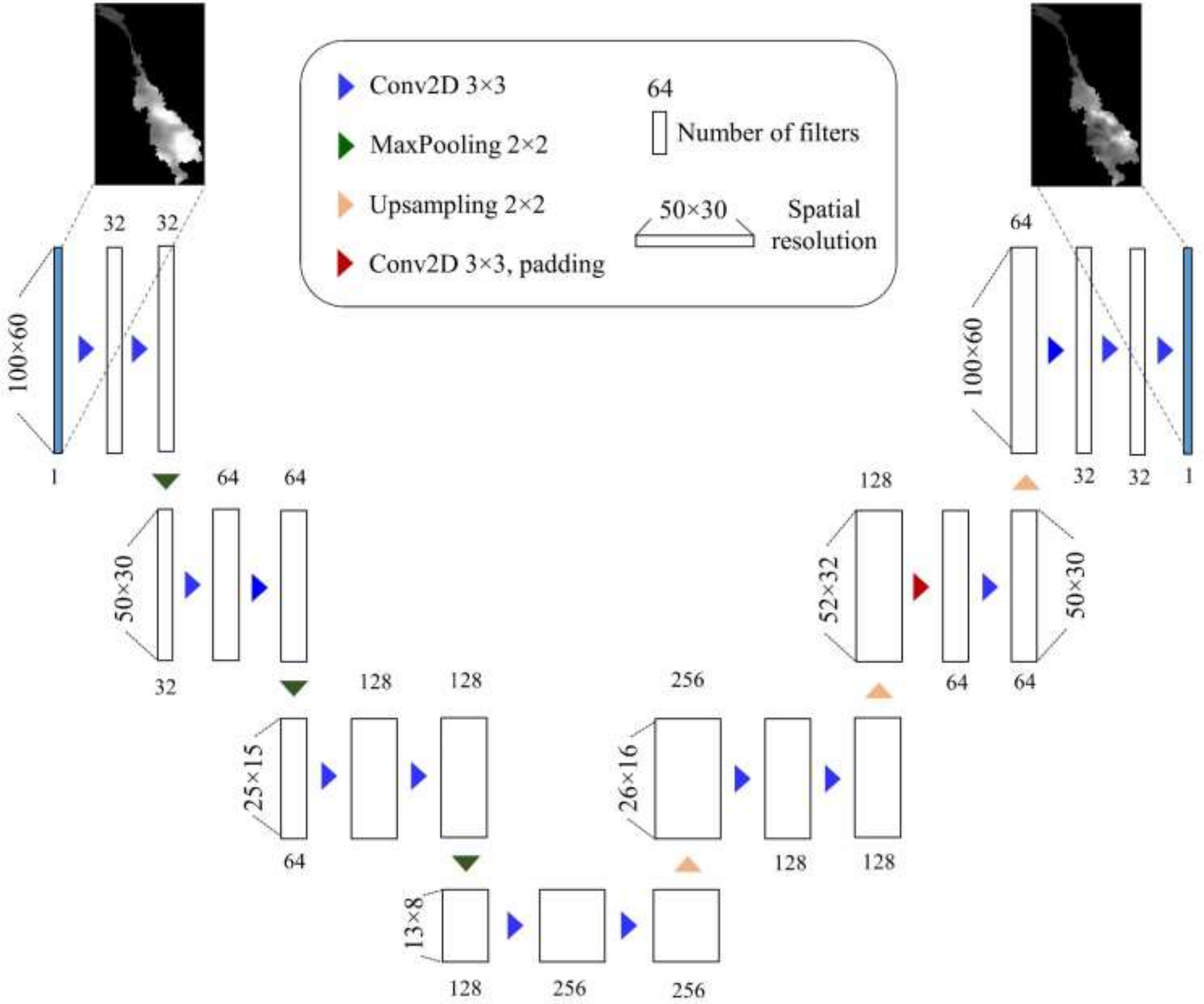
```
e11 = Conv2D(filters=n_filters, kernel_size=(3,3), padding='same',
          kernel_initializer='he_normal')(input_img_v3)
e11 = BatchNormalization()(e11)
e11 = Activation('relu')(e11)

e11 = Conv2D(filters=n_filters, kernel_size=(3,3), padding='same',
          kernel_initializer='he_normal')(e11)
e11 = BatchNormalization()(e11)
e11 = Activation('relu')(e11)

e11 = Conv2D(filters=n_filters, kernel_size=(3,3), padding='same',
          kernel_initializer='he_normal')(e11)
e11 = BatchNormalization()(e11)
e11 = Activation('relu')(e11)

e12 = MaxPooling2D(pool_size=(2, 2), padding='same')(e11)
```

CAE Model

Training parameters:

- ✓ Loss function: MSE
- ✓ Optimize function: Adam
- ✓ Learning rate: 0.001
- ✓ Batch size: 32

- ➢ Total params: 1,753,025
- ➢ Training on: **Google Colab Pro**
- ➢ Time usage: 4-5 (hour)



4. Results and Discussion

- Temporal Correlation
- Spatial Correlation
- Probability Distribution

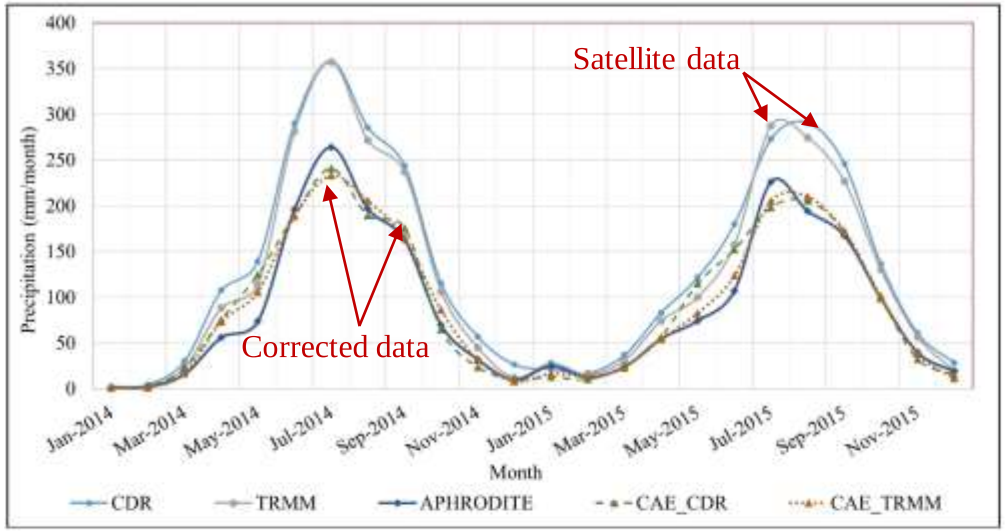
➢ Temporal Correlation

Annual Precipitation

Year	CDR	TRMM	APHRODITE	CAE_CDR	CAE_TRMM
2014	1,661	1,540	1,086	1,125	1,121
2015	1,498	1,402	1,050	1,095	1,058
Average *(mm/year)*	1,579	1,471	1,068	1,110	1,090

Monthly Precipitation

Compared with APHRODITE	MAD *(mm/year)*	RMSD *(mm/year)*	NSE
CDR	43.2	54.1	0.61
TRMM	34.0	45.6	0.74
CAE_CDR	12.4	19.0	0.97
CAE_TRMM	8.7	12.7	0.99



MAD is Mean Absolute Deviation

RMSD is Root Mean Square Deviation

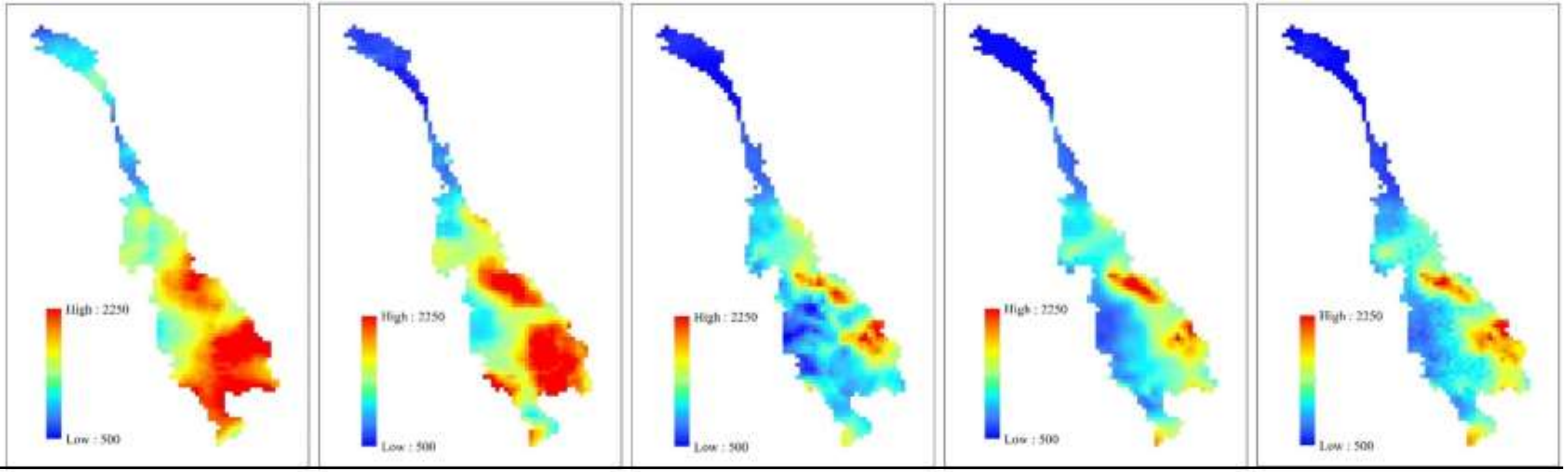
NSE is Nash-Sutcliffe Efficiency

➢ Spatial Correlation

Spatial distribution pattern of precipitation products **in 2014**

$$Bias = \frac{1}{n}\sum_{i=1}^{n}(x_i - y_i) = \bar{x} - \bar{y}$$

$$Corr = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sigma_x \sigma_y}$$

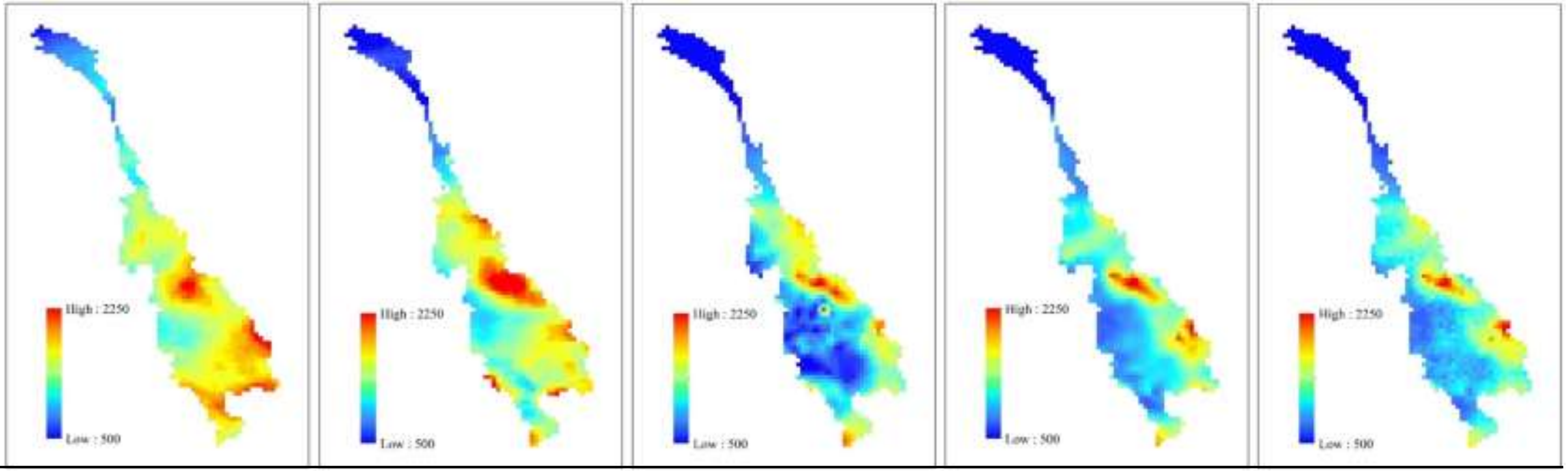


Compared with APHRODITE	**CDR**	**TRMM**	**APHRODITE**	**CAE_CDR**	**CAE_TRMM**
RMSD *(mm/year)*	690	594	-	174	177
MAD *(mm/year)*	582	461	-	134	137
Bias *(mm/year)*	574	453	-	39	35
Spatial Correlation	0.61	0.74	-	0.91	0.91

σ_x and σ_y denoting the standard deviations of x and y

➢ Spatial Correlation

Spatial distribution pattern of precipitation products **in 2015**

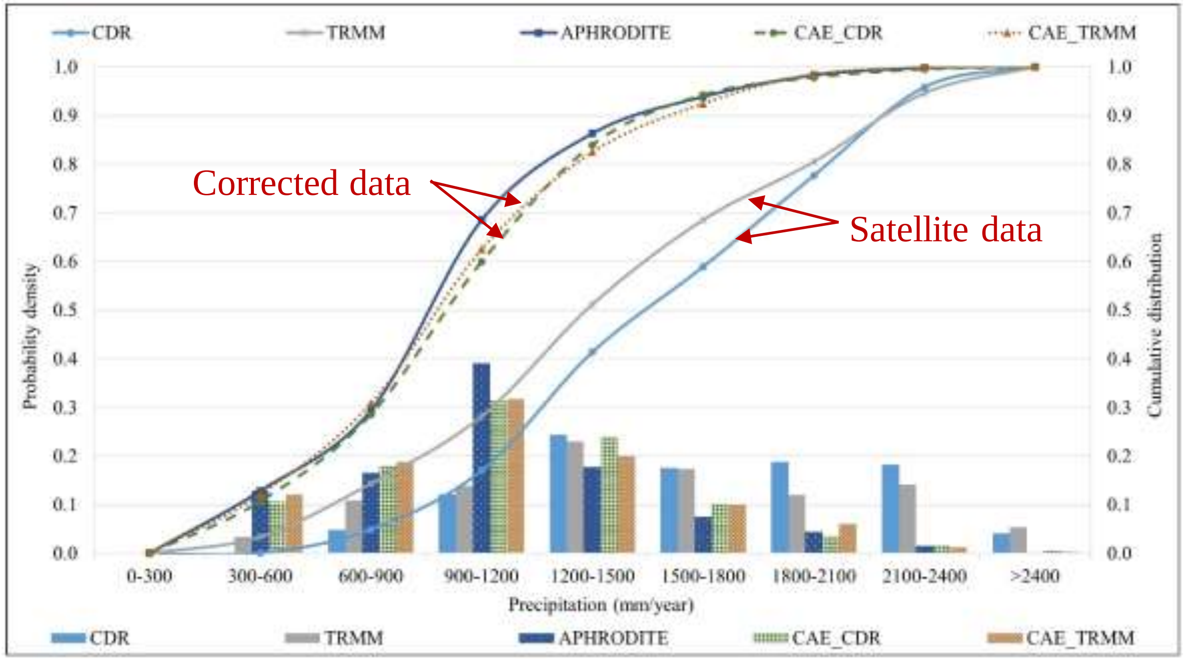


Compared with APHRODITE	CDR	TRMM	APHRODITE	CAE_CDR	CAE_TRMM
RMSD *(mm/year)*	561	450	-	236	210
MAD *(mm/year)*	480	366	-	186	166
Bias *(mm/year)*	448	352	-	46	8
Spatial Correlation	0.63	0.81	-	0.84	0.86

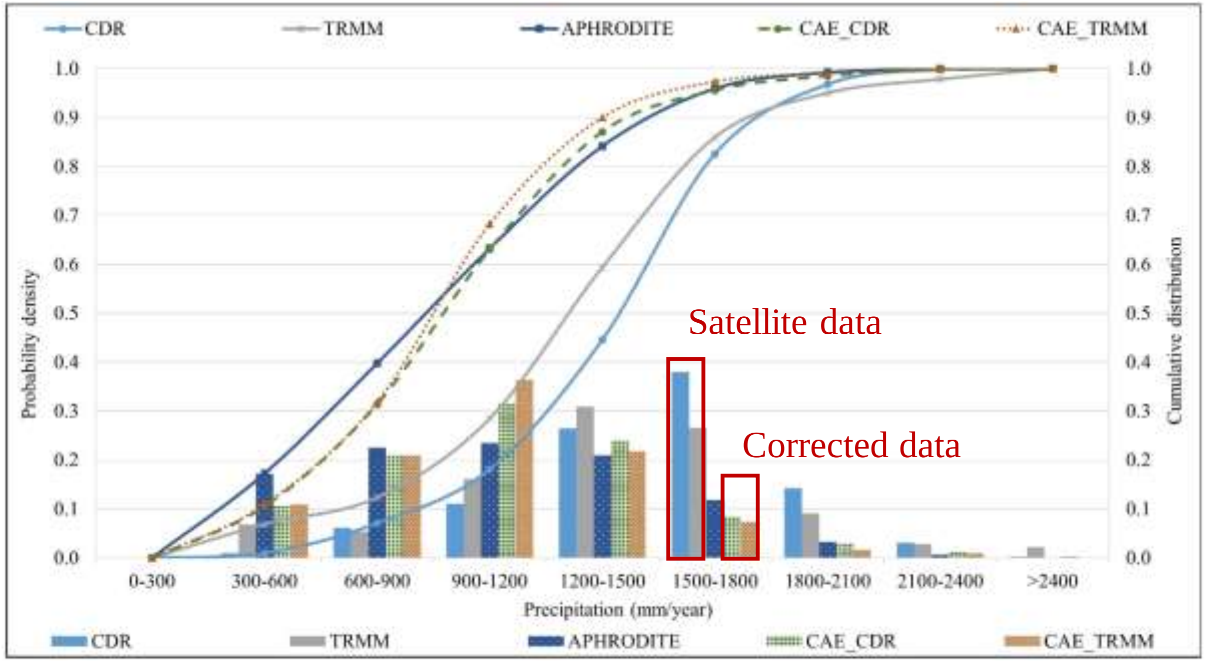
➢ Probability Distribution

Analysis of annual precipitation by pixel

- ✓ Probability Density Function - PDF
- ✓ Cumulative Distribution Function - CDF



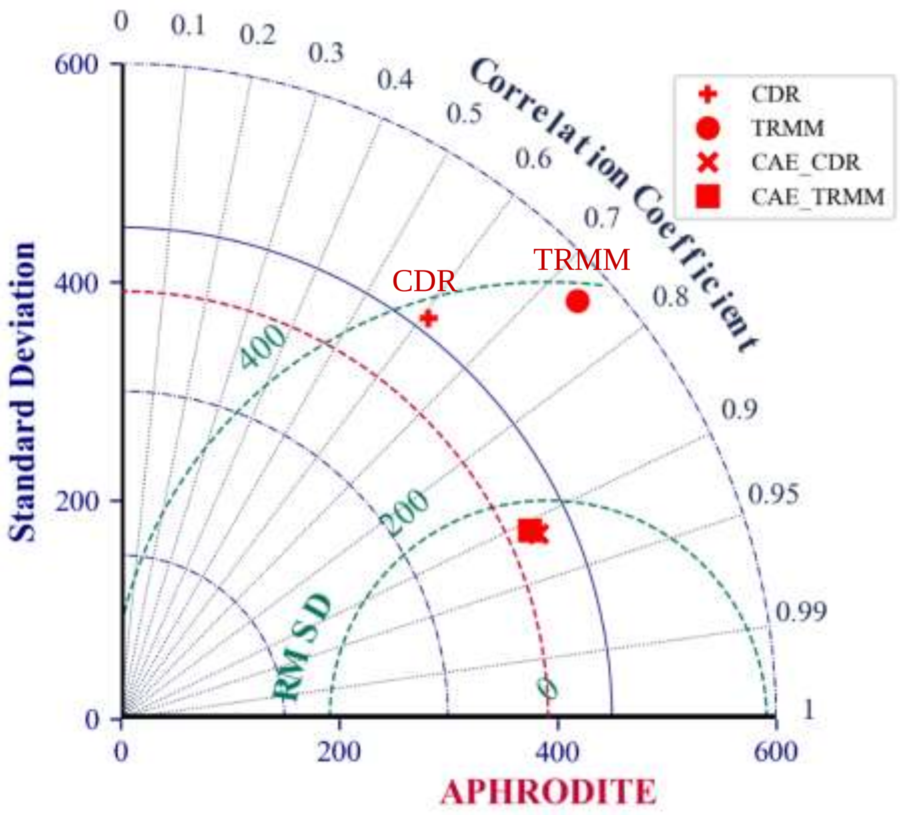
2014



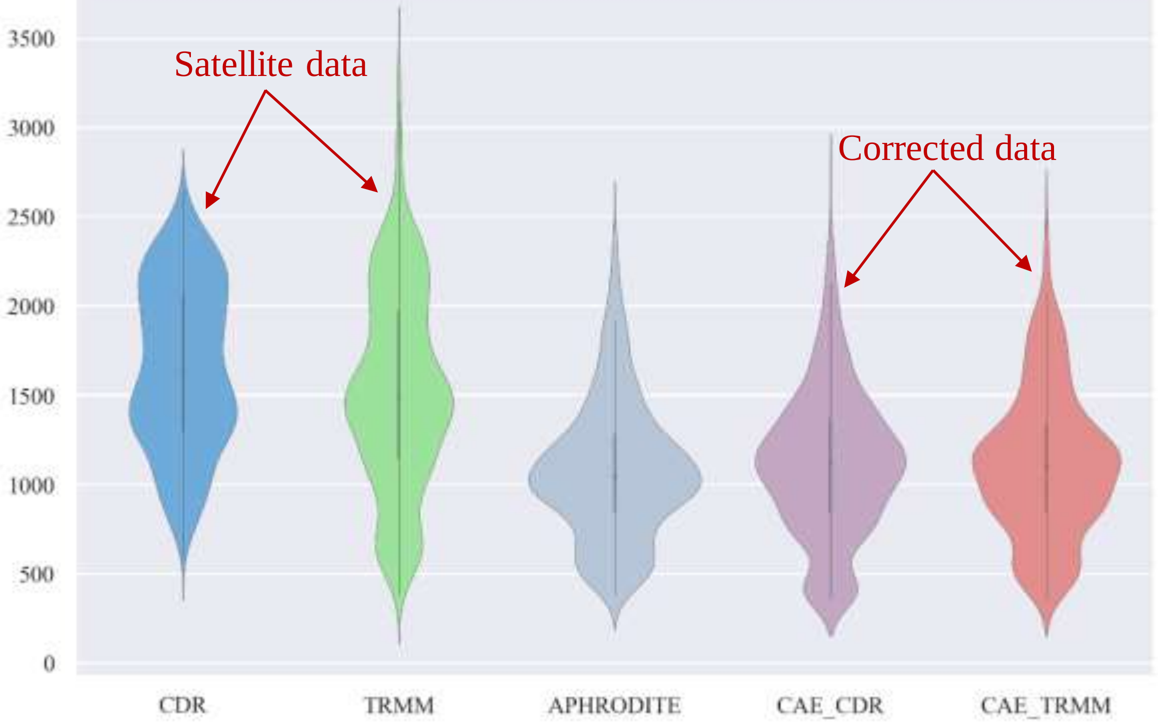
2015

➢ Probability Distribution

- These data sets represent the total annual rainfall of each grid cell



2014



Violin plot of rainfall products in **2014**

6. Concluding Remarks

- Bias-corrected rainfall products provide a higher data quality than the satellite-based products
- Despite using different data sources, the bias-adjusted precipitation products still exhibit competitively excellent performance
- CAE_TRMM is slightly better than CAE_CDR
- These products can capture the trend of rainfall distribution as well as rainfall intensity in terms of spatio-temporal
- **The effectiveness of CAE model**

Based on:

Editor’s Choice Article

Open Access | Editor’s Choice | Article

Application of Convolutional Neural Network for Spatiotemporal Bias Correction of Daily Satellite-Based Precipitation

by Xuan-Hien Le, Giha Lee, Kwansue Jung, Hyun-uk An, Seungsoo Lee and Younghun Jung

Remote Sens. **2020**, *12*(17), 2731; **https://doi.org/10.3390/rs12172731** - 24 Aug 2020

Cited by 7

Abstract Spatiotemporal precipitation data is one of the essential components in modeling hydrological problems. Although the estimation of these data has achieved remarkable accuracy owning to the recent advances in remote-sensing technology, gaps remain between satellite-based precipitation and observed data due to the dependence [...] Read more.

(This article belongs to the Special Issue **Machine and Deep Learning for Earth Observation Data Analysis**)

▶ Show Figures

THANK FOR YOUR TIME!

감사합니다!

lxhien88@gmail.com

https://www.facebook.com/lxhien88/

https://sites.google.com/view/lxhien/