Deep learning based precipitation retrieval algorithm using passive microwave observations

PMW-PrecipNet

Yeji Choi (yejichoi@si-analytics.ai)
AI Research team, SI Analytics

Collaborators: Seongchan Kim¹, Guohua Li¹,
Jeongmyeong Choi², Jaemin Jeong²,
Ji-Hye Kim³,
Dong-Bin Shin⁴
• **Object Detection** - Artificial intelligence to automatically detect major objects of interest in the satellite imagery - Result analysis such as object size and type

• **eXplainable AI (XAI)** - Explainable artificial intelligence for the reliability of analysis results - Reliability verification of object analysis results by explainable AI

• **Semantic Segmentation** - Semantic segmentation for analysis of topography shape, type, and features - Smart land-use analysis by terrain features

• **Super Resolution** - Increasing resolution of satellite imagery using artificial intelligence - Performance improvement of analysis technology by applying super resolution technology

• **Change Detection** - Technology for automatically detecting changed areas and objects using artificial intelligence - Detecting changes of terrain, buildings, roads, etc. and analyzing the extent and types of changes automatically
Contents

▪ Motivation: Deep learning + meteorological satellite
▪ Rain-type segmentation (with Seongchan Kim)
  ▪ RTC-U-net
  ▪ RTC-fcNN
▪ Precipitation retrieval results (with Jeongmyeong Choi, Jaemin Jeong, Ji-Hye Kim)
  ▪ PMW-PrecipNet
▪ Precipitation forecasting (on going⋯) (with Guohua Li)
Motivations
Deep learning + meteorological satellite data

NOAA Data Volume graph, Courtesy Steve Del Greco & Ken Casey, NOAA/NCEI (via Jeff de La Beaujardiere), 2016
Accurate measurements of precipitation are important not only for weather and climate scientist, but also for a wide range of decision makers, including hydrologists, agriculturalists, and emergency managers.

## Precipitation retrievals from space

From an experimental stage of remote sensing with microwave, rainfall measurements is one of the major subject to research.

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Time Period</th>
<th>Frequency</th>
<th>Channels</th>
<th>Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ESMR</strong> (electr. scan MW radiometer)</td>
<td>1972~1983</td>
<td>✓19GHz (Nimbus-5)</td>
<td>✓6.6,10.7, 18, 21, 37GHz</td>
<td>✓Resolution : 25km</td>
</tr>
<tr>
<td><strong>SSM/I→SSMIS</strong> (S SM/I, H/V)</td>
<td>1987~pres.</td>
<td>✓19GHz (Nimbus-5)</td>
<td>✓6.6,10.7, 18, 21, 37GHz</td>
<td>✓Resolution : 25km</td>
</tr>
<tr>
<td><strong>SMMR</strong> (scan multichannel MW radiometer)</td>
<td>1978~1994</td>
<td>✓6.6,10.7, 18, 21, 37GHz</td>
<td>✓Resolution : 30~155km</td>
<td></td>
</tr>
<tr>
<td><strong>TMI</strong> (TRMM microwave imager)</td>
<td>1997~2014</td>
<td>✓10, 19, 21(V), 37, 85 GHz (H/V)</td>
<td>✓Resolution : 7~50km</td>
<td></td>
</tr>
<tr>
<td><strong>AMS-R</strong> (Advanced microwave scanning radiometer)</td>
<td>2002<del>2011/2012</del></td>
<td>✓1.925, 7.3, 10.65, 18.7, 23.8, 36.5, 89.0 GHz</td>
<td>✓Resolution : 4~43km</td>
<td></td>
</tr>
<tr>
<td><strong>AMS-R-E</strong> (AMSR2) (Advanced microwave scanning radiometer)</td>
<td>2012~pres.</td>
<td>✓1.925, 7.3, 10.65, 18.7, 23.8, 36.5, 89.0 GHz</td>
<td>✓Resolution : 4~43km</td>
<td></td>
</tr>
<tr>
<td><strong>ATMS</strong> (Advanced Technology Microwave Sounder)</td>
<td>2011~pres.</td>
<td>✓23.8~183.31 GHz</td>
<td>✓Resolution : 15.8~74.8km</td>
<td></td>
</tr>
<tr>
<td><strong>GMI</strong> (Global precipitation Measurement)</td>
<td>2014~pres.</td>
<td>✓10.65, 18.70, 23.8, 36.5, 89.0, 166, 183 GHz</td>
<td>✓Resolution : 5~25km</td>
<td></td>
</tr>
<tr>
<td><strong>MADRAS</strong> (Microwave Analysis &amp; Detection of Rain &amp; Atmospheric Structures)</td>
<td>2011~pres.</td>
<td>✓18.7, 23.8(V), 36.5, 89.0, 157.0 (H/V) GHz</td>
<td>✓Resolution : 15.8~74.8km</td>
<td></td>
</tr>
</tbody>
</table>
Global precipitation measurements (GPM)

The Global Precipitation Measurement (GPM) mission is an international network of satellites that provide next-generation global observations of rain and snow. (initiated by NASA and JAXA)

GPM Constellation Status

IMERG products: every 30 min global precipitation measurements
Yonsei’s precipitation retrieval algorithm


[Bayesian Inversion Method]

\[
P(R | T_b) \propto P(R) \times P(T_b | R)
\]

[Constructing a-priori database]

\[
Z_{ku} = \frac{\lambda^4}{\pi^3 |K|^3} \int_D \sigma_b(D) N(D) dD
\]

Calculation of Z profiles with hydrometeor profiles

TRMM PR observation

Matching Z profiles over PR/DPR scan swath

3-D precipitation fields over PR scan strips

The probability of observing the TBs given the rain rate from RTM

The probability of a certain rain rate, R, from CRM
Yonsei’s precipitation retrieval algorithm


✓ TB-R relations depending on rain type

![TB-R relations depending on rain type](image)

- **TB histogram depending on rain type**

![TB histogram depending on rain type](image)

✓ Emission and scattering relations

![Emission and scattering relations](image)

Stratiform

Convective
Rain-type segmentation

Motivation

Convective and Stratiform

Separating convective and stratiform (C/S) precipitation types is very important for passive microwave rainfall retrievals.

- **Stratiform**
  - Vertical air motions are weak (mean upward air velocity: ~0.2 m/s)
  - Below the freezing level, melting occurs in an ~500m thick layer.
  - Relatively small rain drops (D~1mm)
  - Hundreds of kilometers in scale

- **Convective**
  - Strong updraft (1~10m/s)
  - Melt rapidly below the 0°C
  - Large raindrops (D>2mm)
  - A few km to about 30 km in scale

✓ C/S separation algorithm using the combination of 19, 37, 85 GHz data (Hong et al., 1999)
✓ Precipitation type classification method using 37GHz observations (Jiang et al., 2018)
✓ Using AdaBboost (Adaptive boosting algorithm) and LDA (Linear discriminant analysis)- Machine Learning Technique (T. Islam et al., 2015)
✓ Using deep learning technique—fully connected neural network (V. Petković et al., 2019)
Dataset preparation

- 2016.01~2018.12 (# of data:~17,000 (6TB))
- subset: Region(120E-175E/5N-55N), 40*40,
  # of data: 59,210

Brightness temperature (TB) fields (input)
related to the gross amount of hydrometeors in the vertical rain column

Rain type from Radar reflectivity (label)
related to the accurate vertical structures of precipitating system

GPM observation

Input(n, 40,40,10)
: TBs for 9 channels, surface type
Label(n)
: Rain rate
# Dataset preparation

## GPM GMI Specification

<table>
<thead>
<tr>
<th>Channel</th>
<th>10.65 (H/V)</th>
<th>18.7 (H/V)</th>
<th>23.8 (V)</th>
<th>36.5 (H/V)</th>
<th>89.0 (H/V)</th>
<th>166 (H/V)</th>
<th>183.31 (V)</th>
<th>183.31 ±7(V)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resolution</td>
<td>19.4</td>
<td>11.2</td>
<td>9.2</td>
<td>8.6</td>
<td>4.4</td>
<td>4.4</td>
<td>4.4</td>
<td>7.3</td>
</tr>
<tr>
<td>Sample NEDT (K)</td>
<td>32.2</td>
<td>18.3</td>
<td>15.0</td>
<td>15.0</td>
<td>7.3</td>
<td>7.3</td>
<td>7.3</td>
<td>7.3</td>
</tr>
<tr>
<td>Beam NEDT (K)</td>
<td>0.96</td>
<td>0.84</td>
<td>1.05</td>
<td>0.65</td>
<td>0.57</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td>Incidence angle</td>
<td>0.53</td>
<td>0.61</td>
<td>0.82</td>
<td>0.52</td>
<td>0.65</td>
<td>1.72</td>
<td>1.72</td>
<td>1.72</td>
</tr>
</tbody>
</table>

### Incidence angle

- Nominal Earth incidence = 52.8°
- Off-nadir angle = 48.5°

### Swath width

- 885km

### Purpose

- Liquid precip.
- Moderate to light precip.
- Water vapor emission
- Moderate to light precip.
- Detection of ice particle
- Light precip.
- Small ice particles and light rainfall and snowfall

## GPM DPR Specification

<table>
<thead>
<tr>
<th>Instrument</th>
<th>GPM Ka-PR</th>
<th>GPM Ku-PR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency (GHz)</td>
<td>35.5</td>
<td>13.6</td>
</tr>
<tr>
<td>Swath width (km)</td>
<td>120</td>
<td>245</td>
</tr>
<tr>
<td>Spatial resolution (km)</td>
<td>5.2</td>
<td>5.2</td>
</tr>
<tr>
<td>Range resolution (m)</td>
<td>250</td>
<td>250</td>
</tr>
<tr>
<td>Observation range (km)</td>
<td>18 to −3</td>
<td>18 to −5</td>
</tr>
<tr>
<td>Minimum detectable (dBZ)</td>
<td>12</td>
<td>18</td>
</tr>
<tr>
<td>(0.2 mm/h)</td>
<td>(0.5 mm/h)</td>
<td></td>
</tr>
<tr>
<td>Measurement accuracy (dBZ)</td>
<td>&lt;±1</td>
<td>&lt;±1</td>
</tr>
</tbody>
</table>
Dataset

- 2016.01~2018.12 ( # of data:~17,000 (6TB))
- subset : Region(120E-175E/5N-55N), 40*40,
  # of data: 59,210

More than 1 raining pixels over 10% of raining pixels over 50% of raining pixels

Polarization corrected temperature

\[
PCT_{10} = 2.5TB_{410V} - 1.5TB_{491H}
\]
\[
PCT_{19} = 2.4TB_{419V} - 1.4TB_{491H}
\]
\[
PCT_{37} = 2.15TB_{437V} - 1.15TB_{491H}
\]
\[
PCT_{89} = 1.7TB_{489V} - 0.7TB_{491H}
\]

TABLE I

<table>
<thead>
<tr>
<th>Dataset</th>
<th>The # of image</th>
<th>NR(%)</th>
<th>ST(%)</th>
<th>CV(%)</th>
<th>OT(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DATASET-I</td>
<td>46,239</td>
<td>90.76</td>
<td>6.25</td>
<td>2.23</td>
<td>0.76</td>
</tr>
<tr>
<td>DATASET-II</td>
<td>11,465</td>
<td>72.05</td>
<td>21.38</td>
<td>4.61</td>
<td>1.96</td>
</tr>
<tr>
<td>DATASET-III</td>
<td>1,410</td>
<td>44.65</td>
<td>48.52</td>
<td>4.51</td>
<td>2.31</td>
</tr>
</tbody>
</table>

The TBs are significantly different depending on the surface types

Fig. 2. Mean normalized TBs at each channel depending on surface type.

The TBs between rain types at channels 8~13 are more distinguishable

DATA SET-I

Fig. 1. Example of subset images of TBs (input) at 10-, 19-, 37-, and 89-GHz horizontally polarized channels from GMI observations, and rain types (label) and rain rate (referential information) from DPR observations (orbit number: 11371/2016.02.22),
Model architecture

We received the best model after 94 epoch, and it took 32.9 min for RTC-U-net. For RTC-fcNN, the best model was obtained after 84 epoch (124.7 min). We used four KIST NEURONs GPU system (NVIDIA Tesla V100 with NVLINK) for training.

- **RTC-U-net (convolutional networks)**
  - 4x4x40
  - 14 128 128 128
  - 2x2x20
  - 256 256 256
  - Conv. 3x3, ReLU
  - Max pool 2x2
  - UpSampling2D 2x2
  - Conv. 1x1, ReLU
  - Skip connection
  - 1024 → 512 → 256 → 128 → 64 → 32 → 16 → 8

- **RTC-fcNN (Fully Connected Neural Network)**
  - NR
  - ST
  - CV
  - OT
  - Loss: categorical cross-entropy loss
  - Activation: ReLU
  - Optimizer: Adam

We used four KIST NEURONs GPU system (NVIDIA Tesla V100 with NVLINK) for training.
Results

RTC-U-net, provides comparative results with the RTC-fcNN.

→ It shows the CNN technique is efficient to retrieve rain type for the entire image at once. The scores of ST and CV shows a comparative result.

→ although the training data set has a dominant number of instances for the NR class. The results with PCTs is similar despite significant PCT differences depending on rain type.

→ The PCTs are the linear combinations of TBs at two different channels, these linear relations is trained in the training process with only nine channels of observed TBs.

RTC-fcNN input: each pixel over an input image of RTC-U-net

→ The number of RTC-fcNN input data is multiplied with the number of pixels in the input image for RTC-U-net.

→ The results from RTC-fcNN with DATA SET-III are not significantly compromised with the reduced number of training data set.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>The GF of image</th>
<th>NR(%)</th>
<th>ST(%)</th>
<th>CV(%)</th>
<th>OT(%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DATA SET-I</td>
<td>46,239</td>
<td>90.76</td>
<td>6.25</td>
<td>2.23</td>
<td>0.76</td>
</tr>
<tr>
<td>DATA SET-II</td>
<td>11,465</td>
<td>72.05</td>
<td>21.38</td>
<td>4.61</td>
<td>1.96</td>
</tr>
<tr>
<td>DATA SET-III</td>
<td>1,410</td>
<td>48.65</td>
<td>48.52</td>
<td>4.51</td>
<td>2.31</td>
</tr>
</tbody>
</table>
Conclusions

- We proposed two different DNNs: RTC U-net and RTC-fcNN, for rain type classification for all surface types.
- RTC U-net based on segmentation technique with CNN showed effective results. Moreover, the CV system's small cells are well distinguishable, although the ST system surrounds it.
- The RTC-fcNN with eight hidden layers showed a comparable performance although it has simple architectures. However, the training time is four times longer than RTC-U-Net.
- Rain type has highly imbalanced distributions, and NR usually has many more instances than the other classes. The results showed the accuracy is depending on the number of instances. We also checked the results with a balanced data set. However, there was a trade-off between missed detection and false alarm showing the missed detection was reduced while the false alarm was increased.
- Although PCTs are good criteria for raining areas and scattering signatures, we confirmed that the effect of PCTs as an input feature for DL is limited.
Using same data with rain type classification but using all 13 channels of GMI

Precipitation retrievals

PMW-PrecipNET: Precipitation retrieval based on Convolutional Neural Networks from passive microwave observations
Model architecture

➢ U-Net based PMW-PrecipNET architecture

Optimizer: Adam
Activation function: Elu
Drop out: 0.5

➢ Training data re-sampling method using 5-fold cross validation
For training: Y2016-2018

➢ Evaluation metric

$MAE = |true - prediction|$ : for the precipitation rate

$F1 = 2 \cdot \frac{Recall \cdot Precision}{Recall + Precision}$ : for the precipitation location

$MOF = \frac{MAE}{F1}$

➢ Ensemble process for the inference step
Results

The performance improved the most over land among the three surface types.

PMW-PrecipNet showed more detailed feature
Channel effect on results

<table>
<thead>
<tr>
<th>Exception ch.</th>
<th>No exception</th>
<th>channel-1</th>
<th>channel-2</th>
<th>channel-3</th>
<th>channel-4</th>
<th>channel-5</th>
<th>channel-6</th>
<th>channel-7</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAE ↓</td>
<td>0.678</td>
<td>0.714</td>
<td>0.680</td>
<td>0.688</td>
<td>0.711</td>
<td>0.700</td>
<td>0.729</td>
<td>0.696</td>
</tr>
<tr>
<td>F1 ↑</td>
<td>0.799</td>
<td>0.805</td>
<td>0.808</td>
<td>0.797</td>
<td>0.796</td>
<td>0.795</td>
<td>0.782</td>
<td>0.789</td>
</tr>
<tr>
<td>MOF ↓</td>
<td>0.849</td>
<td>0.887</td>
<td>0.842</td>
<td>0.864</td>
<td>0.893</td>
<td>0.880</td>
<td>0.932</td>
<td>0.882</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Exception ch.</th>
<th>channel-8</th>
<th>channel-9</th>
<th>channel-10</th>
<th>channel-11</th>
<th>channel-12</th>
<th>channel-13</th>
<th>Longitude</th>
<th>Latitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>MAE ↓</td>
<td>0.729</td>
<td>0.716</td>
<td>0.708</td>
<td>0.708</td>
<td>0.696</td>
<td>0.709</td>
<td>0.691</td>
<td>0.705</td>
</tr>
<tr>
<td>F1 ↑</td>
<td>0.795</td>
<td>0.795</td>
<td>0.796</td>
<td>0.799</td>
<td>0.787</td>
<td>0.803</td>
<td>0.805</td>
<td>0.802</td>
</tr>
<tr>
<td>MOF ↓</td>
<td>0.916</td>
<td>0.900</td>
<td>0.889</td>
<td>0.885</td>
<td>0.883</td>
<td>0.883</td>
<td>0.859</td>
<td>0.879</td>
</tr>
</tbody>
</table>

**CH6**: the most positive impact on the results. The scattering channel is important for the rain rate accuracy.

**CH 1-2**: the negative impact on the precipitation location due to their lower resolution.
Conclusions

- We propose PMW-PrecipNET, a data-driven passive microwave retrieval algorithm.
- The results showed that the proposed algorithm performs well for both ocean and land surface types without the separate processes according to surface types.
- The results show that PMW-PrecipNet provides 19% of the improved correlation with Dual Precipitation Radar compared to the operational GPM precipitation retrieval algorithm.
- This study showed the potential of a deep learning approach to retrieve precipitation without empirical engineering for constructing an a-priori database.
- For further study, we applied this method for global precipitation retrieval and examined the regional differences.
Precipitation forecasting
HPC Cluster-based User-defined Data Integration Platform for Deep Learning in Geoscience Applications (submitted in computers and geoscience journal)

- GEO-DIP architecture

![GEO-DIP architecture diagram]

Table 4
Precipitation prediction results for after one hour of each experiment depending on the number of datasets for the training process

<table>
<thead>
<tr>
<th>Number of data source</th>
<th>Data set</th>
<th>Precision</th>
<th>Recall</th>
<th>F-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 data sources</td>
<td>GOES-17+IMERG+GFS</td>
<td>0.988</td>
<td>0.595</td>
<td>0.743</td>
</tr>
<tr>
<td>2 data sources</td>
<td>GOES-17+IMERG</td>
<td>0.989</td>
<td>0.771</td>
<td>0.740</td>
</tr>
<tr>
<td>1 data sources</td>
<td>GFS+IMERG</td>
<td>0.988</td>
<td>0.780</td>
<td>0.732</td>
</tr>
<tr>
<td></td>
<td>GOES-17+GFS</td>
<td>0.977</td>
<td>0.726</td>
<td>0.672</td>
</tr>
<tr>
<td></td>
<td>IMERG</td>
<td>0.989</td>
<td>0.769</td>
<td>0.726</td>
</tr>
<tr>
<td></td>
<td>GOES-17</td>
<td>0.982</td>
<td>0.625</td>
<td>0.563</td>
</tr>
</tbody>
</table>

Figure 7: The result scores over predicted time depending on the number of training data set

Figure 8: The scatter plots of true and predicted precipitation rate after one hour
Heavy rainfall prediction from satellite and radar observations

- The results with only radar data can not predict the heavy rainfall development and extinction and the heavy rainfall region is much wider than the true.
- The results with radar and Himawari data together show similar patterns and follow well to precipitation system development.
Take home message

- Recently, deep learning techniques are increasingly used in the weather and climate community with various applications.
- Earth system data can be considered ‘big data,’ and we should consider how to prepare ‘AI-ready data’ using ‘big earth data.’
- There is a various observed data type for the same atmospheric phenomenon from multiple sources.
  - Temporal and spatial resolution and coverage
  - The physical relations between different atmospheric variables
  - The observation characteristics (Indirect/direct, instantaneous/cumulative)
- When the supervised learning method is adopted, we should carefully consider which data can be a ground truth and the quality evaluation.
- The high-resolution weather and climate data need big memory in the training process; we need to solve GPU memory limitations. For this, domain scientists and machine learning/computer scientists should collaborate.
Thank you for attention

www.si-analytics.ai

SI Analytics Co., Ltd. (Satrec Initiative Group)
441Expo-ro, Yuseong-gu, Daejeon, 34051, Korea

yejichoi@si-analytics.ai