Applications of machine learning to ARM/ASR science: The Ice Particle and Aggregate Simulator

VANESSA PRZYBYLO - ATMOSPHERIC SCIENCES RESEARCH CENTER
DR. KARA SULIA – ATMOSPHERIC SCIENCES RESEARCH CENTER
ARNOLDAS KURBANOVAS – ATMOSPHERIC SCIENCES RESEARCH CENTER - xCITE LAB
CARL SCHMITT – NCAR
What is IPAS?
The Ice Particle and Aggregate Simulator

To emulate laboratory data, IPAS acts as a ‘theoretical laboratory’ and provides likely aggregate properties given an input number of monomers and shape.

The expected falling orientations, overlap of each monomer, and any contact angle that may form through so-called constrained randomization can be recorded.
Systematic approach for calculating the aggregate aspect ratio

\[
\phi = \frac{c}{a} = \text{aspect ratio}
\]

Form aggregate

Use the peak of the distribution in bulk models

Original monomer aspect ratio

Monomer aspect ratio = 0.050 with shape = 4.948 and \( \phi_a = 0.218 \)
Objective:
• To compare and analyze the aggregate properties computed in IPAS with properties of observed aggregates

Methodology:
• Apply machine learning (ML) algorithms to classify images from a cloud particle imager (CPI) aircraft probe based on ice particle type
• Once the images are sorted and labeled, another algorithm (or statistical analysis through PDFs) will be used to determine particle dimensional characteristics such as perimeter, area, 2D length and width, etc.
  • Compare these characteristic values to IPAS to further refine simulations for the AHM
• Potentially determine trends or patterns in the particle type/size/shape as a function of height, temperature, relative humidity, etc. (e.g., by plugging in an image of a given particle, the model would be able to “predict” what environment it was likely in).
High-Res CPI dataset after preprocessing

Unsuccessful

create training dataset/test dataset with target concept

CNN training and object detection

parameter optimization

CNN model prediction

new image dataset with labels

final detection results

Concept: Successful!

unsupervised learning to classify particle type
Pre-processing

1. CPI Sheets
2. Individual CPI Images
3. Resized CPI Images
4. Hand-Labeled CPI Images
Normalized pixel dimensions ranging from 0.0-1.0

Conv Layer
Layer 1
3x32, kernel size=5, stride=1
Feature Map
ReLU activation
decide which neurons will push forward the values into the next layer
max pooling

Conv Layer
Layer 2
32x64, kernel size=2, stride=2
Feature Map
ReLU activation
max pooling

Layer 1 to Layer 2

Linear transformation
dropout

training a classification problem with multiple classes

ML hyperoptimization packages:
- Pytorch
  - Randomized Search CV
  - Comet optimizer
- Keras
  - Tensorflow backend
  - Hyperas
  - Comet Optimizer

Cross Entropy Loss
lr = .0002
Adams optimizer

a mathematical way of measuring how wrong your predictions are
typically requires little tuning and is computationally efficient
Final Model:

Accuracy of aggs : 94.444444 %
Accuracy of bullets : 54.545455 %
Accuracy of columns : 85.365854 %
Accuracy of junk : 96.000000 %
Conclusions

• Automatized classification through use of a CNN is generating realistic hydrometeor categories, which is useful to other scientists by reducing image pre-processing time
• A generous, automated, and classified dataset is useful toward better prediction of aggregate properties and segregation of the dynamical processes taking place based on environmental conditions (e.g., aggregation, riming, etc.)
• The results of the proposed work will guide future improvements of both IPAS and bulk models while identifying regions of inadequate predictability

• Next Steps:
  • Build the database of hand-labeled images
    • Apply nested classification
    • Potential for primary and secondary habits
  • Extend results to more probes
  • Perform simulations in IPAS given observed particle properties to compare and validate theoretical aggregates
  • Statistically quantify any similarities and differences
Thank You!
Questions?