Using Machine Learning to Assimilate Precipitating Pixel Information from GOES ABI and GLM

Kyle Hilburn [1], Imme Ebert-Uphoff [2], and Steve Miller [3]

[1] Kyle.Hilburn@colostate.edu

[1,2,3] Cooperative Institute for Research in the Atmosphere (CIRA)

[2] Department of Electrical and Computer Engineering (CSU)

Introduction

Motivation: How can we get the maximum benefit from GOES observations for forecasting?

- **Radiance assimilation** is physically-based (easy to interpret), but individual pixel information content saturates around optical depths of 160(8) during day(night) or composite reflectivity (**REFC**) of 20-25(0-5) dBZ, and does not use lightning information
- **Machine learning** is statistically-based (harder to interpret), but image gradients and spatial context provide reliable information to about 45 dBZ, and provides framework for using lightning information (*data fusion*)

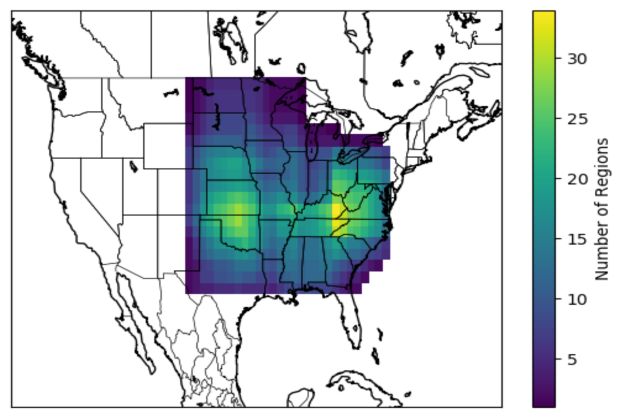
Question: what is our neural network (**NN**) learning that provides such good skill?

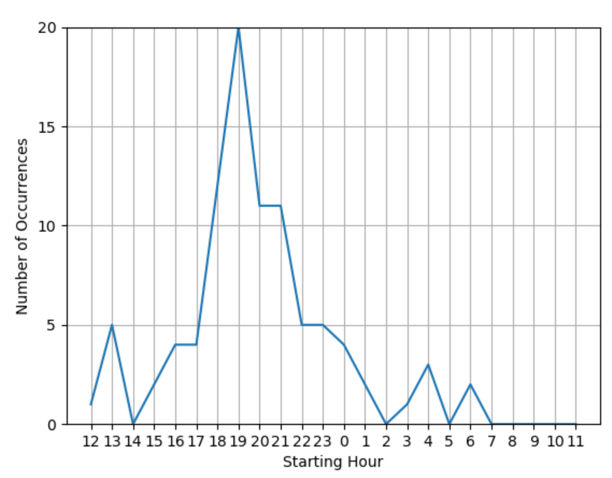
Hypothesis: the skill comes from using information in *image gradients and lightning*

Data

- GOES-16 Advanced Baseline Imager (**ABI**)
  - Channel 7, 3.9-micron, shortwave infrared window
  - Channel 9, 6.9-micron, mid-level water vapor (~442 mb)
  - Channel 13, 10.3-micron, clean longwave infrared window
- GOES-16 Geostationary Lightning Mapper (**GLM**)
  - Group extent density
- Multi-Radar Multi-Sensor (**MRMS**) Quality-Controlled Composite Reflectivity
- All fields are resampled to 3-km Lambert Conformal Conic HRRR grid

Dataset Construction





- Selected samples from the 92-day period 4/17/2019 to 7/17/2019 during which there was abundant severe weather.
- *Automatically select regions- and times-of-interest* based on maximizing the number of **SPC storm reports** (tornado, hail, wind)
  - 6-hour periods with 15-minute refresh
  - 256 x 256-pixels on 3-km HRRR grid (768 km)
- Mode of 20-50 storm reports per day
- **Top panel**: geographic preference for Southern Great Plains and Upland South
- **Bottom panel**: temporal preference for mid to late afternoon
- Split: 80% / 20% for training / validation
- An independent training dataset (*Hilburn et al., JSC 2019*) that includes nighttime and other locations produces similar results

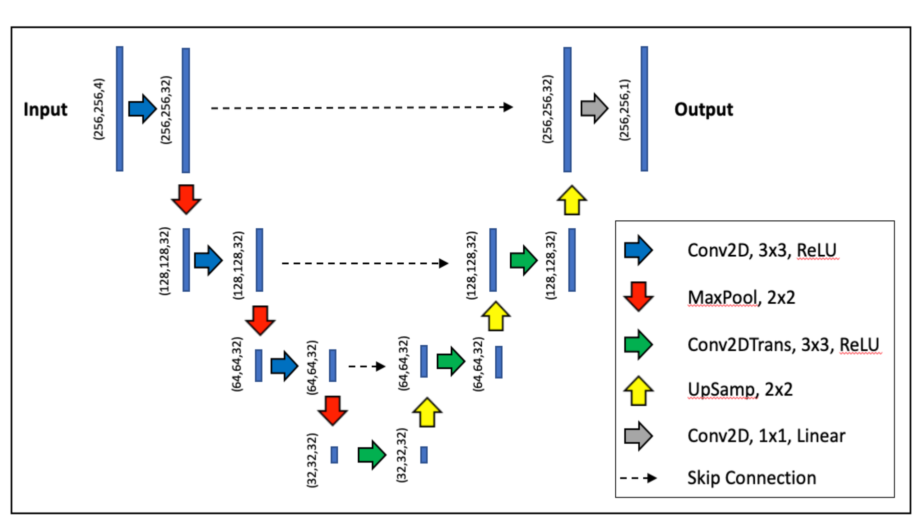
Analysis Methodology

Our approach is to produce many models and interrogate them in order to open the lid of the "black box" and identify the strategies the NN is using that produce such good skill.

- **Channel withholding experiments** to identify the information content that is most important for producing skill in certain situations
- Comparing results using standard 3x3 convolutional filters with **1x1 filters** in order to remove the spatial context and simulate an approach considering just individual pixels
- Use of **attribution methods**, such as Layer-wise Relevance Propagation, to visualize what information the NN is using to make a specific prediction
- Use of **synthetic inputs** to quantify the sensitivity of the output to variations in properties of the inputs
- Use of **metrics** that are unrelated to the loss function (MSE), such as: coefficient of determination (R^2), categorical metrics at various output threshold levels (POD, FAR, CSI, Bias), and evaluation of the MSE binned over the range of true output values

Architecture

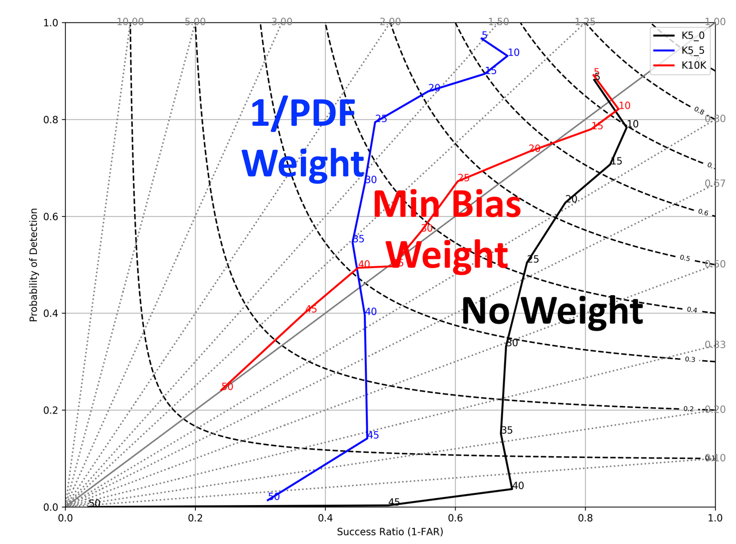
- Sequential structure based on **U-Net**
- Skip connections off: they provide very small improvements but complicate visualization
- 3 encoding and decoding layers, *deeper produces overfitting*
- 32 filters per layer, fewer do nearly as well but give blurry output
- 100 epochs validation statistics:
  - RMSD = 5.29 dBZ
  - R^2 = 0.738



This model has 47,457 trainable parameters.

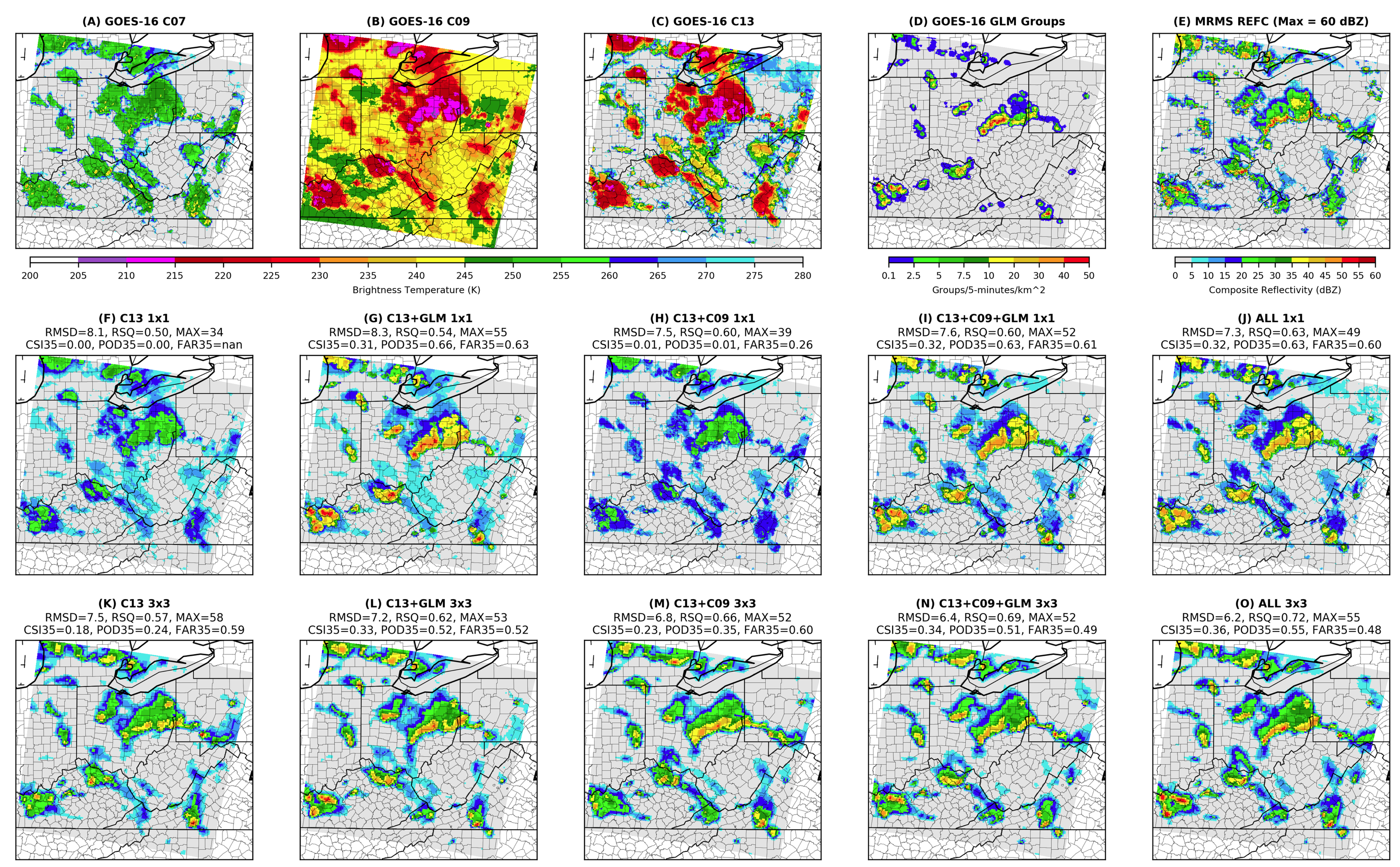
Loss Functions

- Standard unweighted MSE loss function has *sub-optimal performance at high REFC*
- REFC PDF ~ Exp(-5*y) where y=[0,1]
- Use performance diagram (**right**) to select loss function weights producing minimum bias
- Also select model (different random seeds)
- *Generalized exponential*: Wt = Exp(b*y ^{c})
- The optimal coefficients (grid search) are b=5 and c=4 (for MSE) and c=3 (for MAE)
- Connection to AUC approach but without derivative problems
- Acts as a global constraint on realism of fields



Performance for REFC: 0, 5, 10, … 50 dBZ

Results for Validation Sample 2019-07-02 23:30Z



Top row: GOES inputs (**Panels A-D**) and MRMS truth (**Panel E**)
Middle row: Predictions with 1x1 filters for various channel combinations
Bottom row: Predictions with 3x3 filters for various channel combinations

Importance of Gradients and Spatial Context

Traditional infrared imager retrievals of precipitation, which only use individual pixel information or rudimentary spatial information, have poor skill (low POD and high FAR).

- **Panel F** simulates that type of algorithm, which has poor skill at REFC > 20 dBZ
- Adding water vapor (**Panel H**) helps a little bit, but not enough at high values

Allowing the NN to use gradient information and spatial context provides tremendous improvements in skill.

- **Panel K** shows that *even with just C13, image gradients and spatial context carry a great deal of information about REFC > 35 dBZ*
- Note that RMSD and R^2 tell a limited story, and that *categorical statistics are crucial* for evaluating whether a model provides improvements
- Adding water vapor (**Panel M**) helps increase the POD in areas where the difference between C09 and C13 is small, but does this *at the expense of a high FAR*

Importance of Lightning

Given that gradients carry so much information, to isolate the importance of lightning, consider the 1x1 experiments.

- Adding lightning (**Panel G**) *provides dramatic improvements for REFC > 35 dBZ* but note that values between 20-35 dBZ are mostly absent

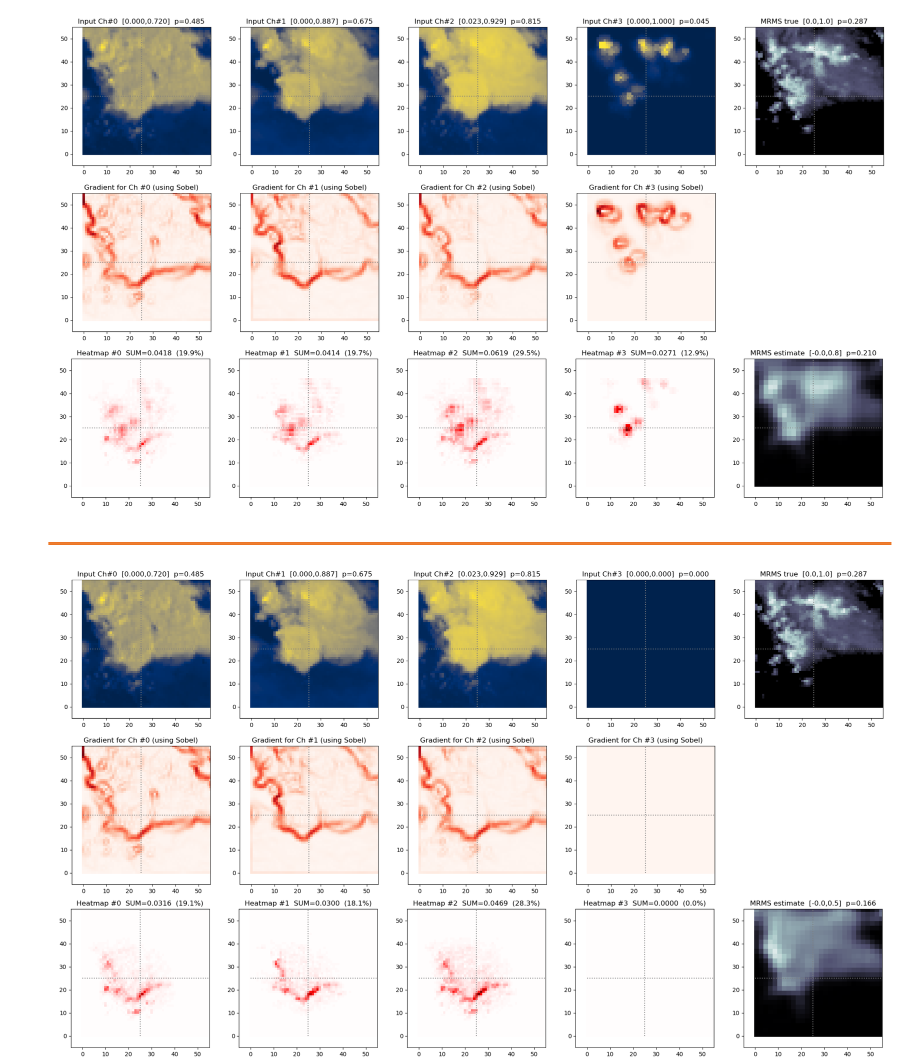
When gradients are included, lightning provides less relative value, but it still has unique characteristics.

- **Panel L** shows that combining lightning with C13 provides *dramatic improvements in POD* (0.52 vs 0.24) with reasonable FAR (categorical bias is near one)
- Unlike water vapor (**Panel M**) lightning is better able to *pinpoint locations of strong radar echoes* and provides dramatically better POD (0.52 vs 0.35)
- **Panel O** shows that *other channels work together with lighting to provide the best estimates* with sharp, well defined convective core features

The properties illustrated in this example of the skill provided by lightning and image gradients are confirmed in statistics across all validation samples.

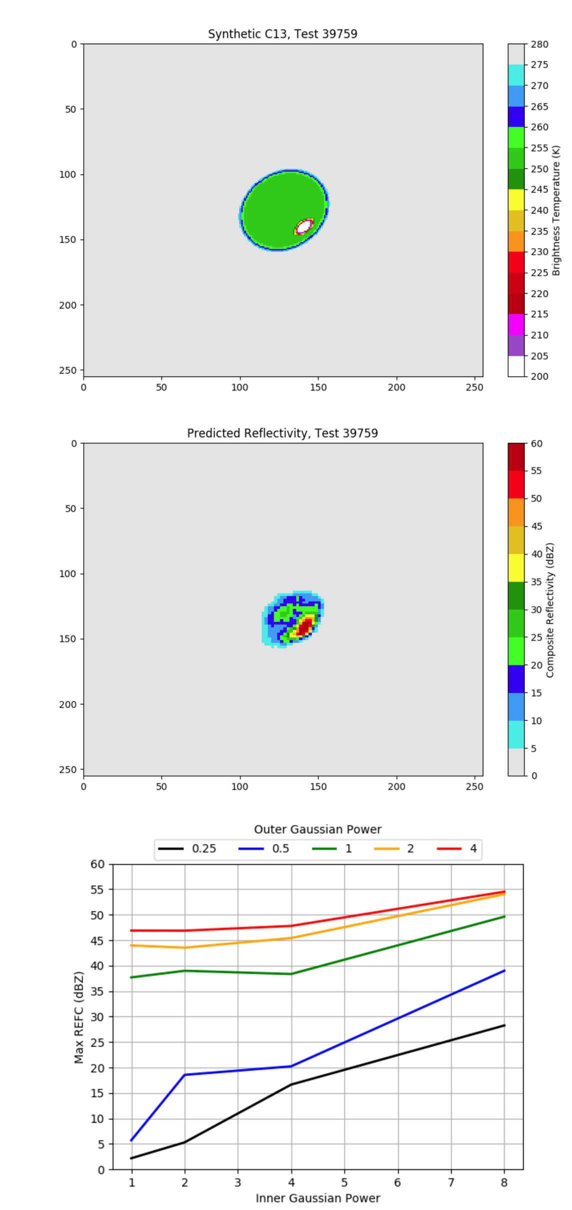
Layer-wise Relevance Propagation

- **Top set**: results with lightning
- **Bottom set**: lightning zeroed out
  - **Top rows**: input fields
  - **Middle rows**: image gradients
  - **Bottom rows**: heatmaps
- Uses LRP with alpha=1 and beta=0
- NN uses the lightning value itself more than the lightning gradient (**top set**)
- Without lightning (**bottom set**), the network uses strong infrared gradients
- With lightning, *the emphasis for the other channels is changed*, looking at gradients in locations with lightning
- While the LRP percentage of lightning is only 12.9% in this case, it impacts the interpretation of the other channels, *giving the NN additional clues of where to look*, yielding a more accurate estimate of REFC
- A *remaining question* is how does the NN learn which strong gradients to ignore and which are important?



Synthetic Inputs

- Using a *sum of Generalized Elliptical Gaussians* model that provides six parameters for the inner and outer Gaussians: 1) location, 2) amplitude, 3) size, 4) aspect, 5) orientation, and 6) sharpness (exponent)
- Evaluating 45K+ different parameter settings, the spatial patterns that most strongly activate the NN, based on maximum REFC, all resemble that shown to the **right** (**top**: synthetic input, **middle**: NN output)
- The NN has learned about *thunderstorms with overshooting tops* (OT)
- Note the *very strong gradients* along the anvil edge and along the OT edge, corresponding to large exponents
- The weakest responses have in common weak gradients and are the least physical looking
- Evaluating all the model parameters, the most influential are the inner and outer Gaussian sharpness
- An example of the sensitivity is given in **right bottom** panel, which shows the maximum REFC as functions of the inner and outer exponents
- The emergence of 35 dBZ echoes requires the outer exponent to be 1 or greater, or very large inner exponents around 8



Summary and Conclusions

- We have shown that a convolutional NN trained on *GOES ABI+GLM can accurately reproduce composite reflectivity* from MRMS over eastern CONUS warm season
- We have shown the skill comes from *gradients in infrared images and lightning* and that lightning helps the network better interpret radiance gradients
- We used *novel approaches* to derive weights for the loss function and in our analysis methodology to evaluate the importance of image gradients and lightning
- A *remaining question* is how applicable will this NN be to different meteorological regimes, such as tropical convection, and what additional meteorological information will be needed to produce robust predictions globally?
- However the *tools developed* in this work will be applied to investigate those questions
- **Goal**: GOES-derived synthetic reflectivity profiles used where ground-based radar network coverage is poor for the RAP/HRRR latent heating initialization/assimilation
- Additional details about this work will appear in *Hilburn et al. (2020, JAMC in preparation)*

Acknowledgements

- Thank you to the **GOES-R Program**; this research is support by award NA19OAR4320073
- Thank you to **NOAA RDHPCS** for access to the Fine Grain Architecture System on Hera, without which this research would not have been possible