NATIONAL WEATHER SERVICE

Building a Weather-Ready Nation

Probabilistic Forecast Of Thunderstorms Using Artificial Neural Network (ANN) With Google Keras Deep Learning Libraries

Mamoudou Ba, Meteorological Development Laboratory National Weather Service, NOAA

Objectives

The overall goal is to explore the use of neural network for Post-Processing of NWP model for thunderstorm probabilistic forecasting

Forecasting thunderstorm is one of the most difficult tasks in weather prediction, due to their rather small spatial and temporal extension and the inherent non linearity of their dynamics and physics

Ingredients for thunderstorm initiation and development

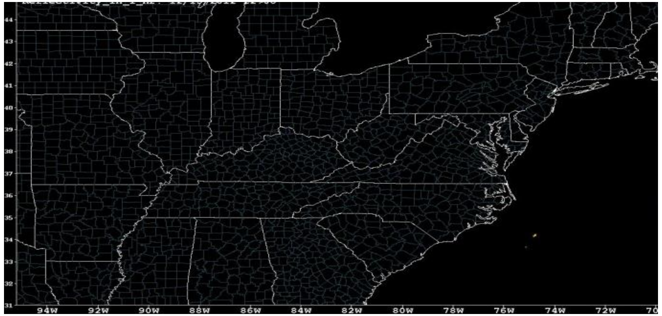
- **Three ingredients must be present for thunderstorm initiation and development:**
  - Moisture
  - Instability (parcel instability, latent instability, and convective (potential) instability
  - Lifting

Lifting mechanisms:

- Frontal boundaries, outflow boundaries
- Low level warm air advection
- Orographic flow
- Low pressure system
- Differential heating along soil, vegetation, soil moisture ➔ low level convergence
- Low level moisture advection
- Gravity wave
- Differential Positive Vorticity Advection, jet streak divergence

Data

Data for the study included RAP and surface derived fields nowcasts and radar data collected over 31-45^{o}N and 94-71^{o}W between June and September, 2012. RAP and surface derived fields used correspond to 14z, 15z, 16z, 17z, 18z, 19z, 20, 21, 22, and 23Z. Truth data are derived from the radar data nearest to the RAP forecast time. June and July data are used to train the Neural network, August and September for the validation. All data are remapped to grid of 0.02^{o} x 0.02^{o} latitude-longitude.



The study domain

DERIVATION OF THE PREDICTORS USED in the ANN model

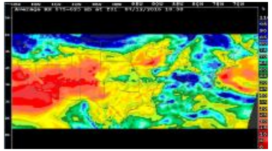
- The predictors used in the Neural Network are a subset of the fields produced by the operational NCAR AutoNowcaster system. This system analyses the characteristic features of atmosphere from NWP RAP model, surface METAR, and radiosonde to derive the predictors. The results of the analyses are 60-minute predictors which are converted into dimensionless likelihood fields.

The likelihood fields have a dynamic range from -1 to 1, where increasing positive values correspond to an increasing likelihood of storm initiation and/or sustainment, and vice-versa.

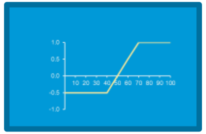
1. CAPE*
2. CIN*
3. RH*
4. W (vertical velocity)*
5. Likelihood of frontal surface*
6. Li*
7. Surface convergence
8. Vertical instability*

* Ingredients derived from Rap model

PREDICTOR

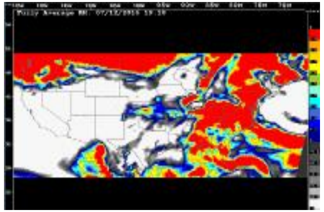


Fuzzy Function for Average RH 875-625 mb at f01



Fuzzy Function for Average RH 875-625 mb at f01

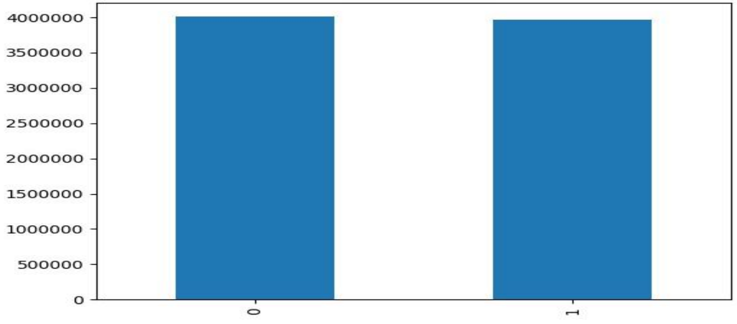
LIKELIHOOD FIELD



“Fuzzy” Average RH 875-625 mb at f01

Input Data Pre-Processing

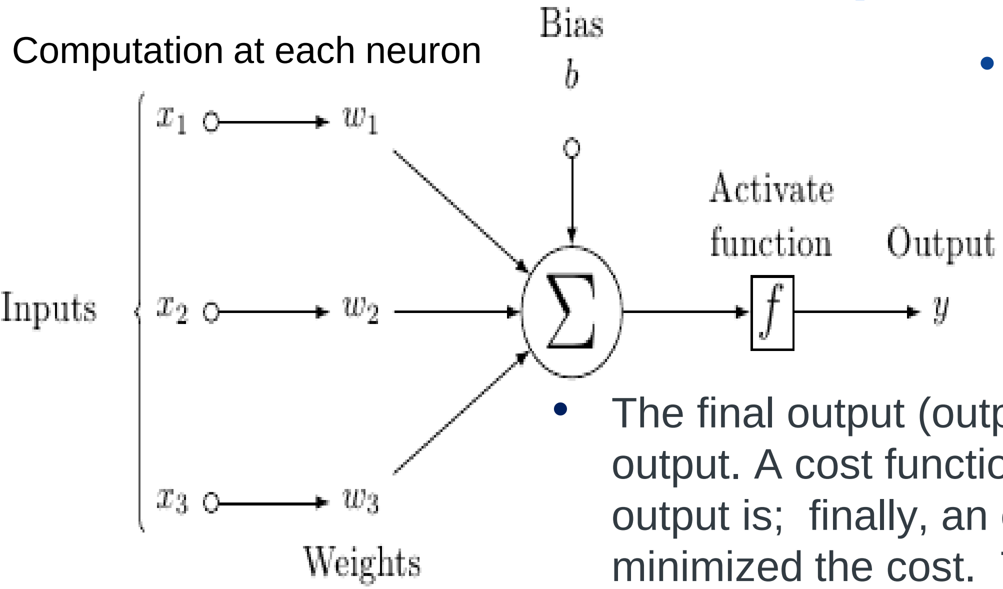
- A CSV file (of 9 columns, 8 predictors (X_i)), and the corresponding truth* (Y_i) is created from the predictor field grids and radar composite reflectivity (truth*). Each line in the CSV file represents pixel values in the input grids.
- All lines containing missing values are dropped from the CSV file.
- Because thunderstorms are very rare events, so the data are highly unbalanced. Python Pandas dataFrame API is used to balance the data.
- After shuffling the dataset, the features (X_i) and labels (Y_i) are splinted into two separate sets: train set (80%) and validation set (20%) using the sklearn library.



*Truth data consists of 3D Cartesian radar reflectivity grid composited between 2.5 and 4.5 km, inclusive. Storms are defined here with pixel values $\geq$ 35 dBZ.

A feed-forward/back-propagation ANN with two hidden layers with 500 neurons for probabilistic storm forecast.

Computation at each neuron

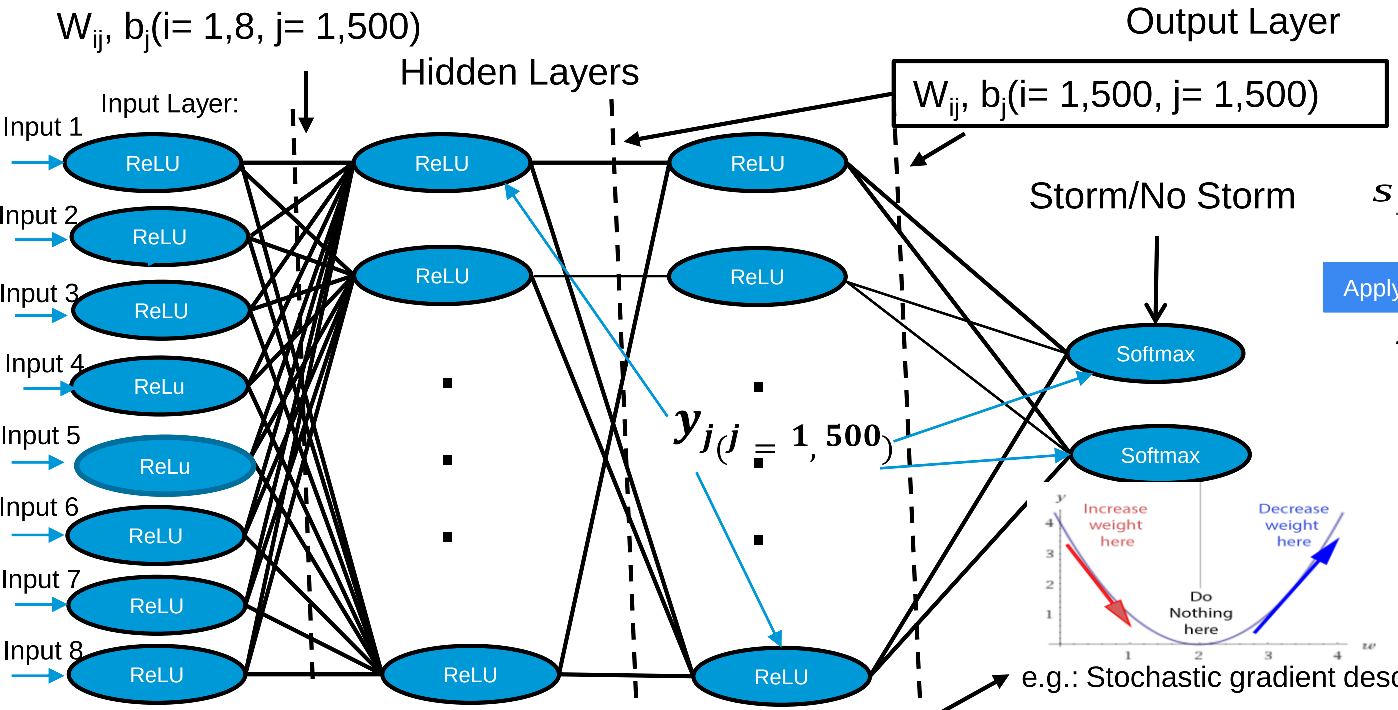


- The sum of the products of the weights and the inputs is calculated in each node, where it undergo the activation function, so the neuron can decide whether or not to fire and output some data to the following layer.

- The final output (output layer value) is then compared to the intended output. A cost function is used to determine how wrong the predicted output is; finally, an optimizer function, Adam optimizer is used to minimized the cost. The cost is minimized by tinkering with the weights (i.e. back-propagation), with the goal of lowering the cost.
- The lowering of the cost is determined by the leaning rate.

The next slide shows graphically a more complete of the ANN architecture.

A feed-forward/back-propagation ANN architecture (2 hidden layers with 500 neurons).



Computation at each neuron

$$s_j = \sum_{i=0}^{n} w_{ij} x_i + b_j$$

Apply activation functions

$$y_j = f(s_j)$$

Output number between 0 and 1

ReLU: $f(s_j) = \max(sj, 0)$

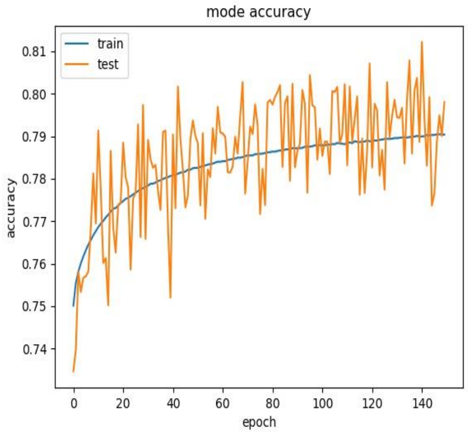
Softmax: $f(s_j) = \frac{e^{s_j}}{\sum_{k=1}^{K} e^{s_k}}$

Loss function (L) is used to minimize the error between the predicted value and the truth; this is achieved by optimizing the weights and biases (back-propagation). In this study categorical cross-entropy is used.

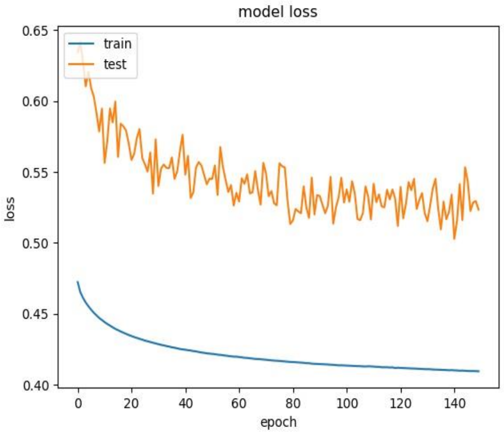
$$L(y, \hat{y}) = -\sum_{j=0}^{M} \sum_{i=0}^{N} (y_{ij} * log(\hat{y}_{ij}))$$

Plot of model accuracy, model loss, and Receiver Operating Characteristic Curve (ROC) on train and validation datasets

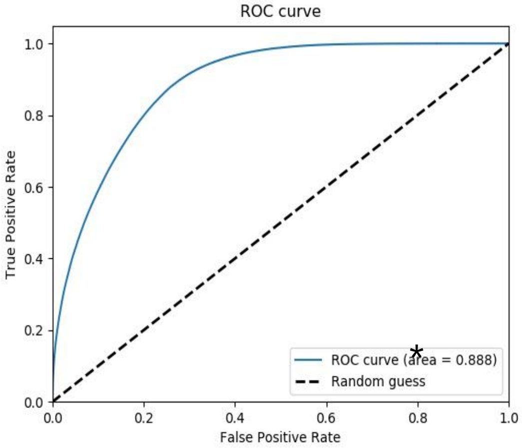
Model accuracy



Model loss



ROC: A Graphical measure of skill for binary classifiers



*The Area Under Curved (AUC) is 0.89, that means there is 89% chance that model was able to distinguish between storm class and no storm class.

Model performance scores using Confusion Matrix metrics on Validation Datasets

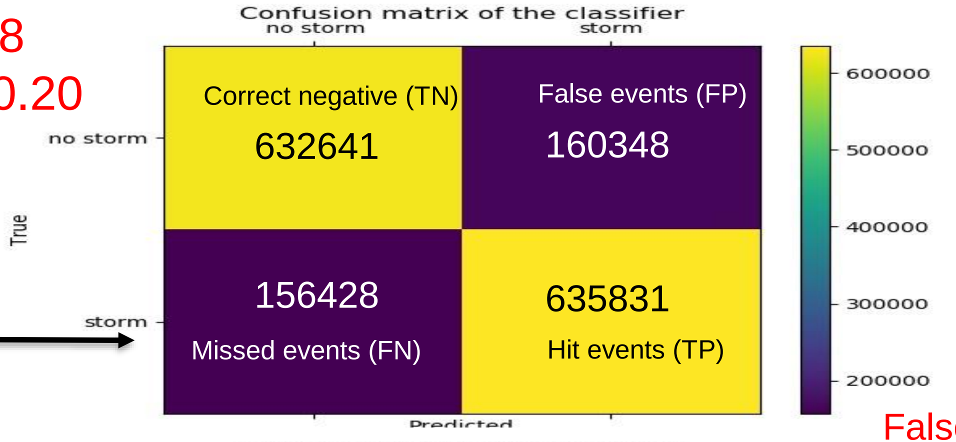
Sample size (N) = 1,585,248

Error Rate = (FP + FN)/N = 0.20

Precision = TP/(TP+FP) = 0.80

Accuracy = (TP+TN)/ N = 0.80

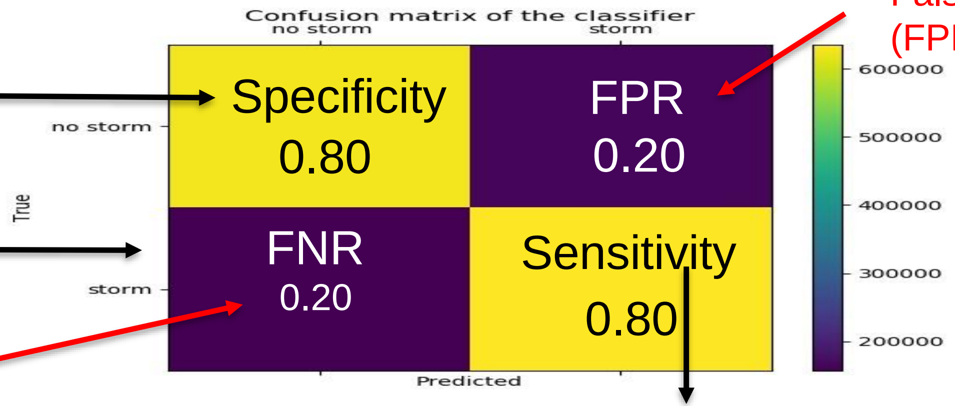
Confusion matrix, without normalization



False Positive Rate (FPR) = FP/(TN+FP)

Specificity = TN/(TN+FN)

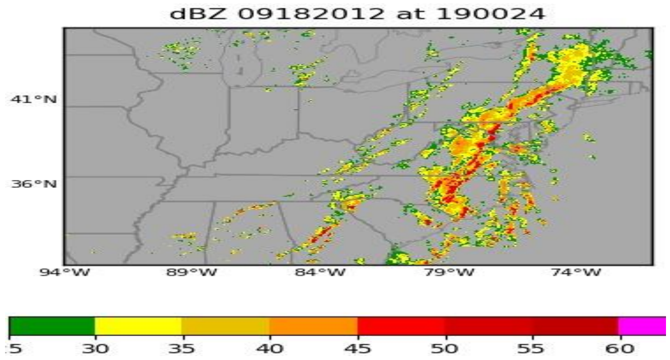
Normalized confusion matrix



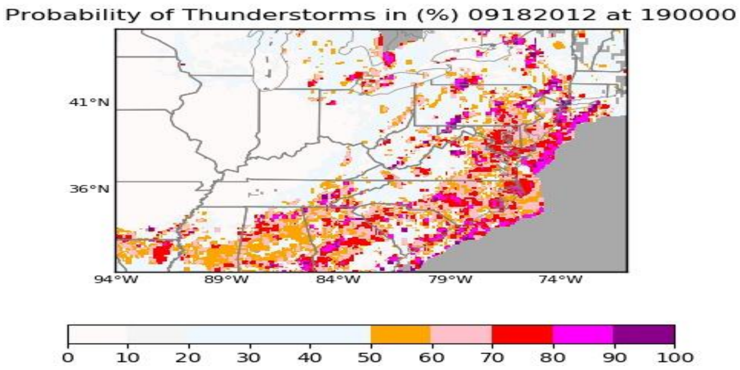
False Negative Rate (FNR) = FN / (TP + FN)

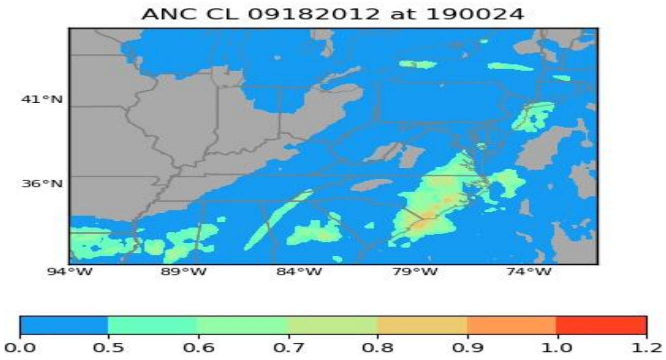
Sensitivity = TP/(TP+FN)

The performance of the Neural network is compared to that of the operational AutoNowcaster (ANC)



The saved ANN model was used to make prediction using new data (August and September 2012). These predictions were used to further assess the performance of ANN model and compare its performance to that of ANC, a deterministic model that forecasts convective likelihood.

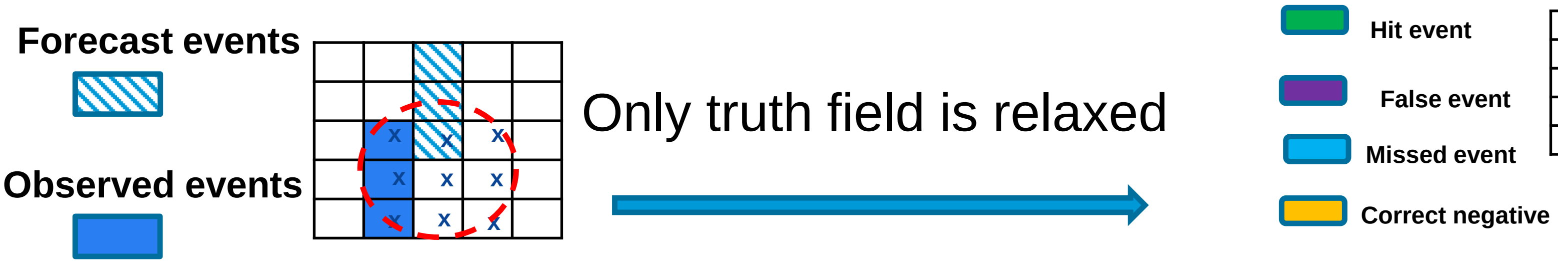




Verification Technique

- The truth field is relaxed to populate the contingency table by relaxing the traditional requirement that forecast and observations must match at the grid scale for a hit to occurs. The neighborhood relaxation window size was specified as 2*N+1 x 2*N+1.
- N=6 and N = 12 were used with an approximate spatial resolution of 25km x 25km, and 50 km x 50 km grid box.

Verification Technique (Populating the contingency table)



- Hit → an event is forecasted at grid point and the event is observed at any of the grid points within the neighborhood
- Miss → an event is observed at any of the grid points within the neighborhood and the event is not forecasted at the grid point
- False alarm → an event is forecasted at a grid point and not observed at any of the grid points within the neighborhood
- Negative correct → no event is forecast at a grid point and no event is observed at any of the grid points within the neighborhood

Statistics Presentation

- Box-and-whisker plots:

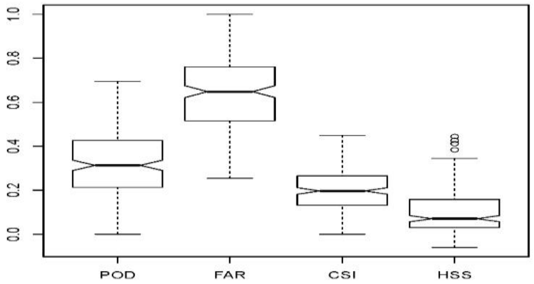
A box-and-whisker plot represents a distribution of values: the box denotes the interquartile range (IQR; i.e., 25th to 75th percentiles), while the notches represent roughly a 95% confidence interval for differences between the medians of the two distributions. The whiskers extend 1.5 times the length of the box, with all data points further than this from the median drawn as circles.

- Performance diagram:

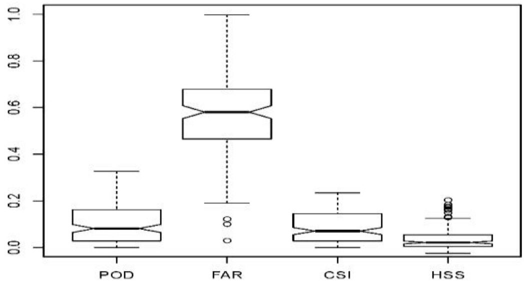
A performance diagram summarizes multiple measures of forecast quality SR, POD, bias, and CSI. Curved lines are CSI and straight lines represent bias. For good forecasts, POD, SR, bias and CSI approach unity, such that a perfect forecast lies in the upper right of the diagram.

Box-and-whisker plots of verification statistics calculated for ANN vs. ANC's Convective Likelihood (CL) forecast over the study domain (August and September 2012); Only truth field is relaxed.

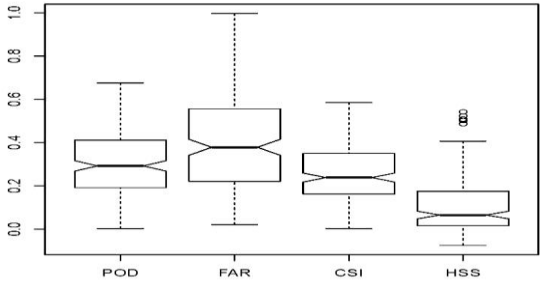
ANN 25 km



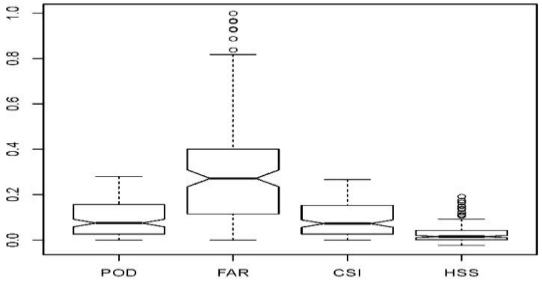
CL 25 km



ANN 50 km



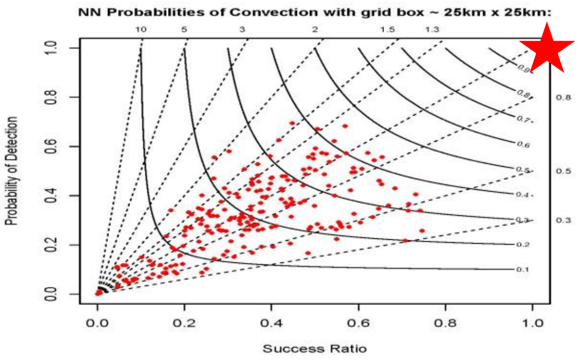
CL 50 km



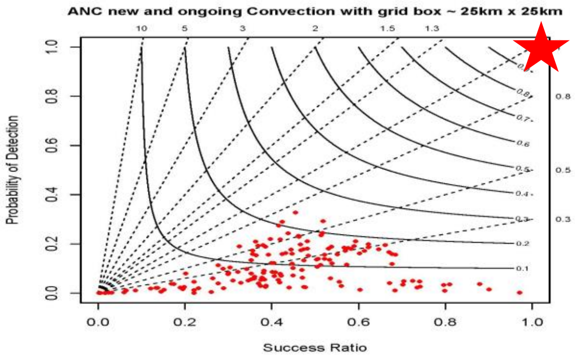
Performance diagram plots of verification statistics calculated for ANN vs. ANC's Convective Likelihood (CL) forecast over the entire study domain (August and September 2012); Only truth field is relaxed.

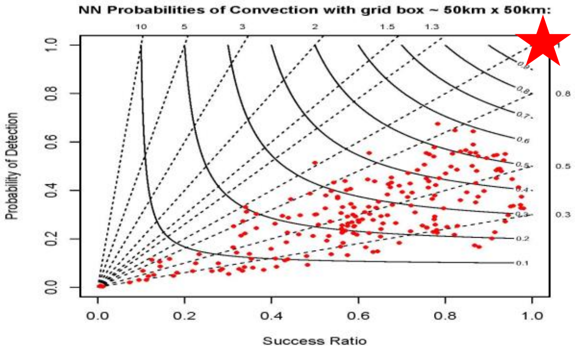
★ Best scores

NN 25 km

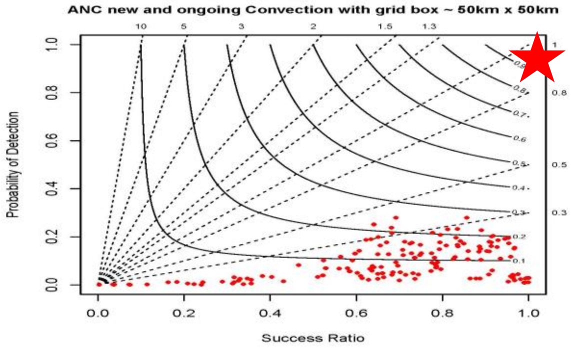


CL 25 km





NN 50 km



CL 50 km

Curved lines are CSI and straight lines represent bias. For good forecasts, POD, SR, bias and CSI approach unity, such that a perfect forecast lies in the upper right of the diagram ★

CONCLUSIONS AND FUTURE DEVELOPMENT

Conclusion:

The Neural Network was found to have significant skill as measured by standard statistics metrics to classify storm vs. no storm using RAP derived instability fields. The results of this exploratory study shows that the ANN outperforms the current ANC based forecast.

Future development:

Expand input data to include more thunderstorm ingredients derived from RAP, e.g. Collins and Tissot 2015 suggested that the following features as relevant predictors of thunderstorm development:

- Wind at surface, and at 850 hPa
- Convective precipitation
- Vertical wind shear at 800-600 hPa, and vertical wind shear at 10m – 800 hPa.

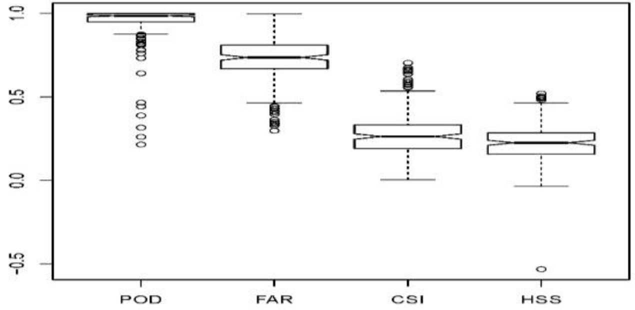
Backup Slides

Both forecast and truth fields are relaxed

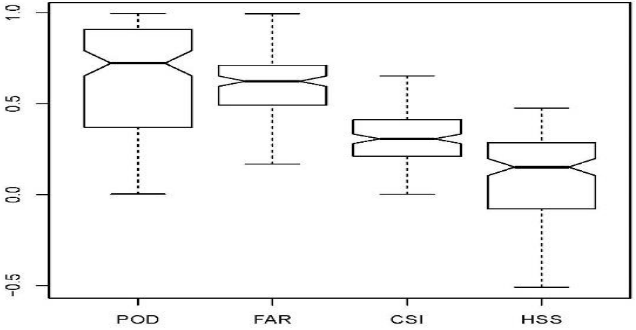
•Hit → an event is observed at grid point and the event is forecasted at the grid point or at any grid point within the neighborhood, or an event is forecast at a grid point and the event is observed at the grid point or at any grid point within the neighborhood
•Miss → an event is observed at a grid point and none of the grid points within the neighborhood forecast the event.
•False alarm → an event is forecasted at a grid point and not observed at any of the grid points within the neighborhood
•Negative correct → No event is observed at a grid point and no event is forecasted at that grid point

Box-and-whisker plots of verification statistics calculated for NN vs. ANC's Convective Likelihood (CL) forecast over the entire Golden Triangle domain (August and September 2012); both fields are relaxed.

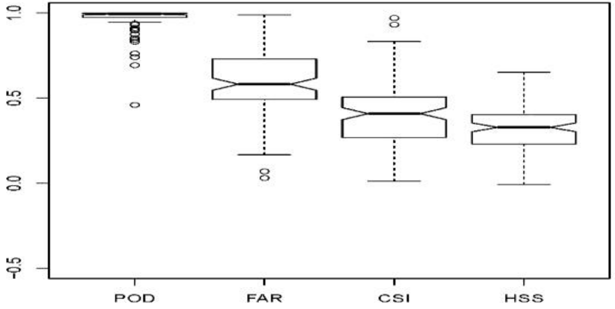
NN 25 km



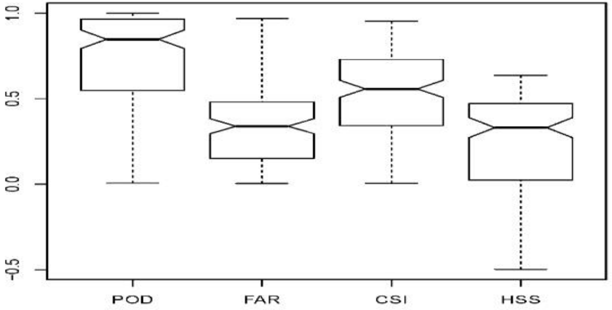
CL 25 km



NN 50 km

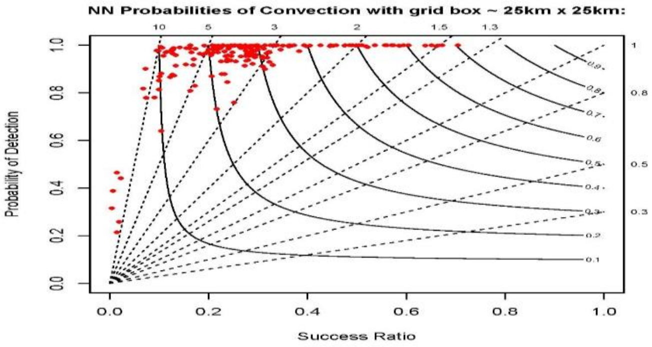


CL 50 km

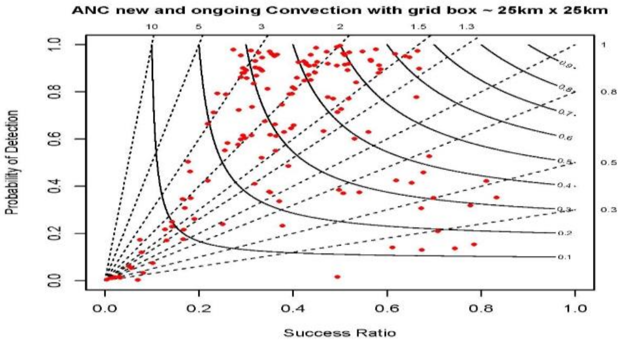


Box-and-whisker plots of verification statistics calculated for NN vs. ANC's Convective Likelihood (CL) forecast over the entire Golden Triangle domain (August and September 2012); both fields are relaxed.

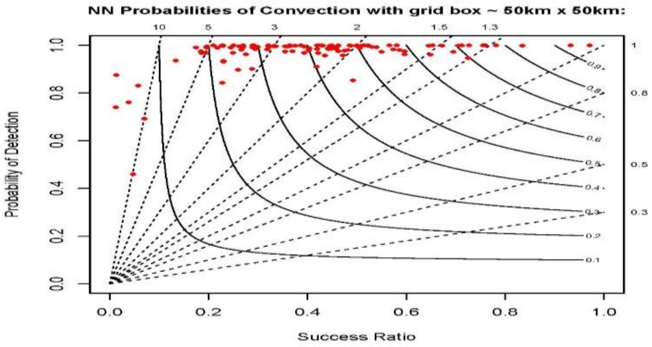
NN 25 km



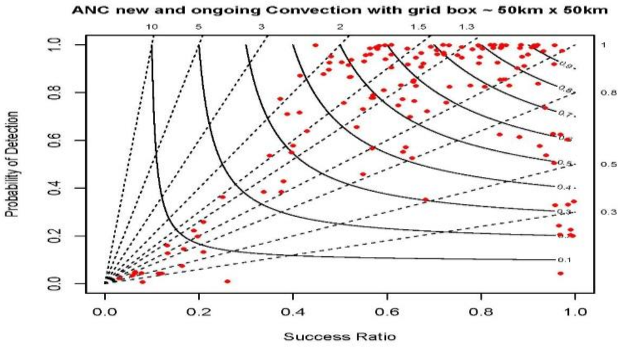
CL 25 km



NN 50 km



CL 50 km



Ingredients used in the Neural Network of this study are a subset of convective fields used in the Operational AutoNowcaster.

1. CAPE*
2. CIN*
3. RH*
4. W (vertical velocity)*
5. Likelihood of frontal surface*
6. Li*
7. Surface convergence
8. Vertical instability*

* Ingredients derived from Rap model

Surface convergence is computed from surface observations