Improvement in hurricane intensity forecast using Neural Networks

Tirthankar Ghosh and TN Krishnamurti
Florida State University
Tallahassee, FL-32306, USA

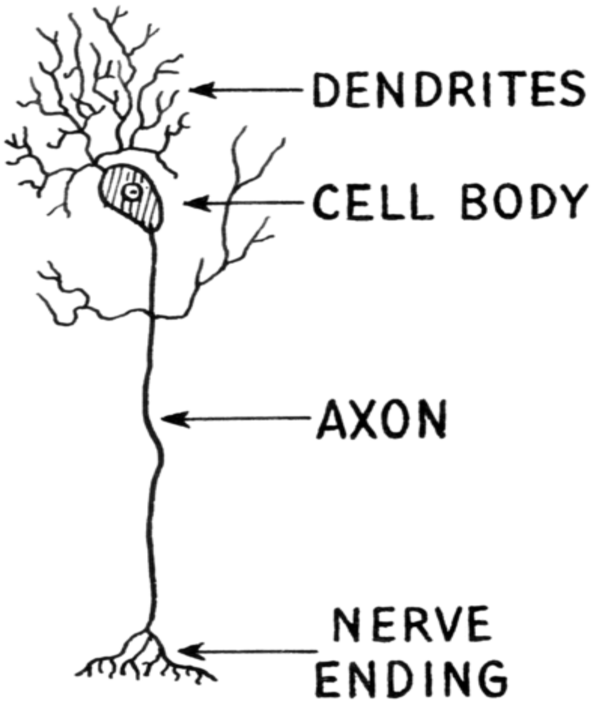
HFIP Annual Review Meeting, Jan 11-12, 2017
National Hurricane Center, Miami

Acknowledgement: HFIP, NOAA Award No. NA15OAR4320064

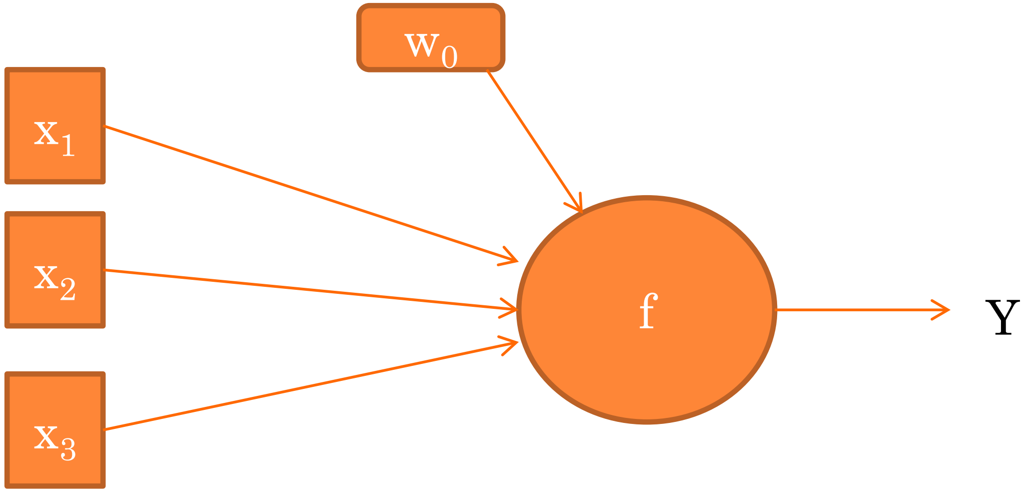
- Human brain takes best possible decision from past experience
- Information from the environment is taken by the sensory organs & passed to the brain through neurons (nerve cells)
- 10 billion nerves with 10000 synapses (meeting point of two nerves)

- Input branch (Dendrites)
- Output branch (Axon)
- Dendrites sends the received information through the cell body to the action

* Axon passes it to dendrite of the next neuron via synapse



IDEA IS FOLLOWED TO APPROXIMATE OUTPUT FOR A GIVEN SET OF INPUTS



$$Y = f\left(w_0 + \sum_{i=1}^{k} w_i x_i \right)$$

Activation function

- Linear
- Logistic
- Hyper tangent

Applications

- Classification
- Discrimination
- Estimation (time series prediction)
- Process identification
- Process control
- Etc …

Types we Consider

- Multilayer Perceptron (MLP)

- Generalized Regression Neural Network

Information flows from input to output

Learning

- Previous observations on input (s) as well as output are provided repeatedly to estimate the neuron parameters (supervised learning)

- Modification of parameters for better performance (desired output)

Choice of weights

- Let $\{x_j^k, y^k\}, k = 1, 2, \ldots, N; j = 1, 2, \ldots, p$ be a set of given observations
- Estimate y which minimizes the square error loss

$$ESS = \frac{1}{2}\sum_{k=1}^{N}\left[y^k - f\left(x^k, w\right)\right]^2$$

- The weights (here model weights) are so chosen that ESS would be minimum

$$m = \sum_{i=1}^{n} w_i m_i$$

where n = no. of input neurons, w_i = synaptic weight for the i^{th} neuron, m_i = input to the i^{th} neuron. It fits to our problem of combining some model forecasts through a linear combination. Let (o_j, M_{lj}) ,l = 1, 2, …, K; j=1,2, …, N are, respectively, given observed and corresponding input values where K is number of models and N is the number of available cases.

Hence, for a single observation the weights are adjusted for error minimization as follows.

$$\frac{\partial ESS}{\partial w_i} = -\sum_{j=1} (o_j - WM_j) m_{ji}$$

$$= -\sum_{j=1}^{N} \varepsilon_j m_{ji} \qquad W = (w_1, w_2, \ldots, w_n); M_j = (y_1, y_2, \ldots, y_n)'; \varepsilon_j = (o_j - WM_j)$$

For updating the movement, it should be in the opposite direction to the gradient.

$$w_i \leftarrow w_i - \alpha \frac{\partial ESS}{\partial w_i}$$

At each stage the error

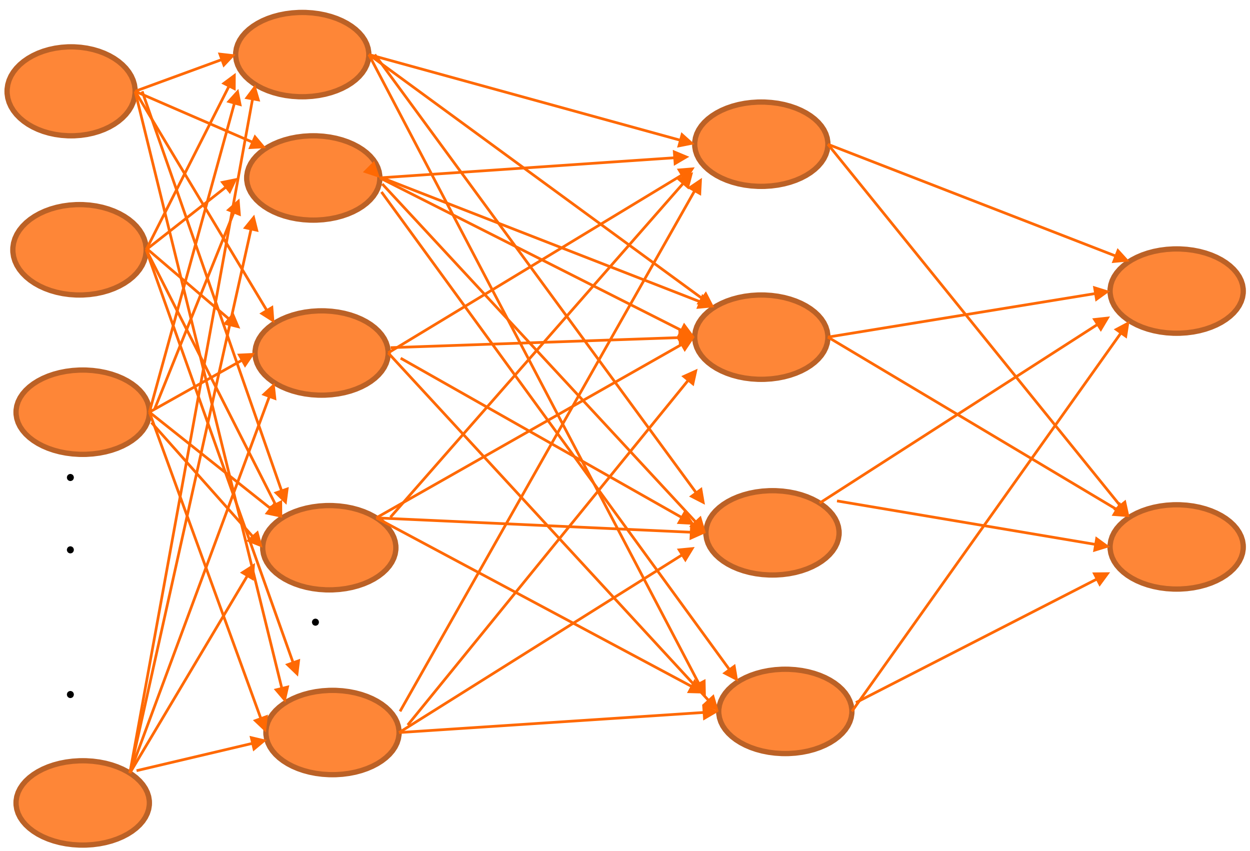
$$\varepsilon_j \leftarrow o_j - WM_j$$

is to be computed and the updated NN weights are given by

$$w_i \leftarrow w_i + \alpha \varepsilon_j m_{ji}$$

where α is known as the learning rate.

MLP



Input layer Hidden layers Output layer

Tasks

- Numbers of hidden layers (developer provided)
- Determining the learning rate (developer provided)
- Train the network
- Evaluate the performance
- Repeat the above process if not satisfied (iterative)

GRNN: Based on standard statistical theory

The conditional mean $\mathbf{E}(Y/x)$ or the regression equation, $\hat{Y}(x)$, of Y for a given value of X, **x**, is given by

$$\hat{Y}(x) = E(y/x) = \frac{\int_{-\infty}^{\infty} y f(x,y)\,dy}{\int_{-\infty}^{\infty} f(x,y)\,dy} \qquad (1)$$

where $f(x,y)$ is the joint continuous probability density function of X (vector valued) and Y. Y may also be vector valued and the corresponding estimate can be derived accordingly.

The pdf has to be estimated from sample observations (**x**, **y**) when it is unknown. If (x_i, y_i), i=1,2, …,n are the sample values of size n of the random variables (X,Y) then the estimated pdf $\hat{f}(x,y)$ is given by

$$\hat{f}(x,y) = \frac{1}{(2\pi)^{(p+1)/2}\sigma^{(p+1)}} \cdot \frac{1}{n}\sum_{i=1}^{n} \exp\left[-\frac{(x-x_i)^T(x-x_i)}{2\sigma^2}\right] . \exp\left[-\frac{(y-y_i)^T(y-y_i)}{2\sigma^2}\right] \qquad (2)$$

where p is the dimension of the vector variable X.

That is the estimate of the pdf is sum of the sample probabilities of width σ for each sample (x_i, y_i). Substitution of estimated pdf obtained in (2) to (1) provides the desired conditional mean

$\hat{Y}(x)$, which is

$$\hat{Y}(x) = \frac{\sum_{i=1}^{n} \exp\left[-\frac{(x-x_i)^T (x-x_i)}{2\sigma^2}\right] \int_{-\infty}^{\infty} y \exp\left[-\frac{(y-y_i)^T (y-y_i)}{2\sigma^2}\right] dy}{\sum_{i=1}^{n} \exp\left[-\frac{(x-x_i)^T (x-x_i)}{2\sigma^2}\right] \int_{-\infty}^{\infty} \exp\left[-\frac{(y-y_i)^T (y-y_i)}{2\sigma^2}\right] dy}$$

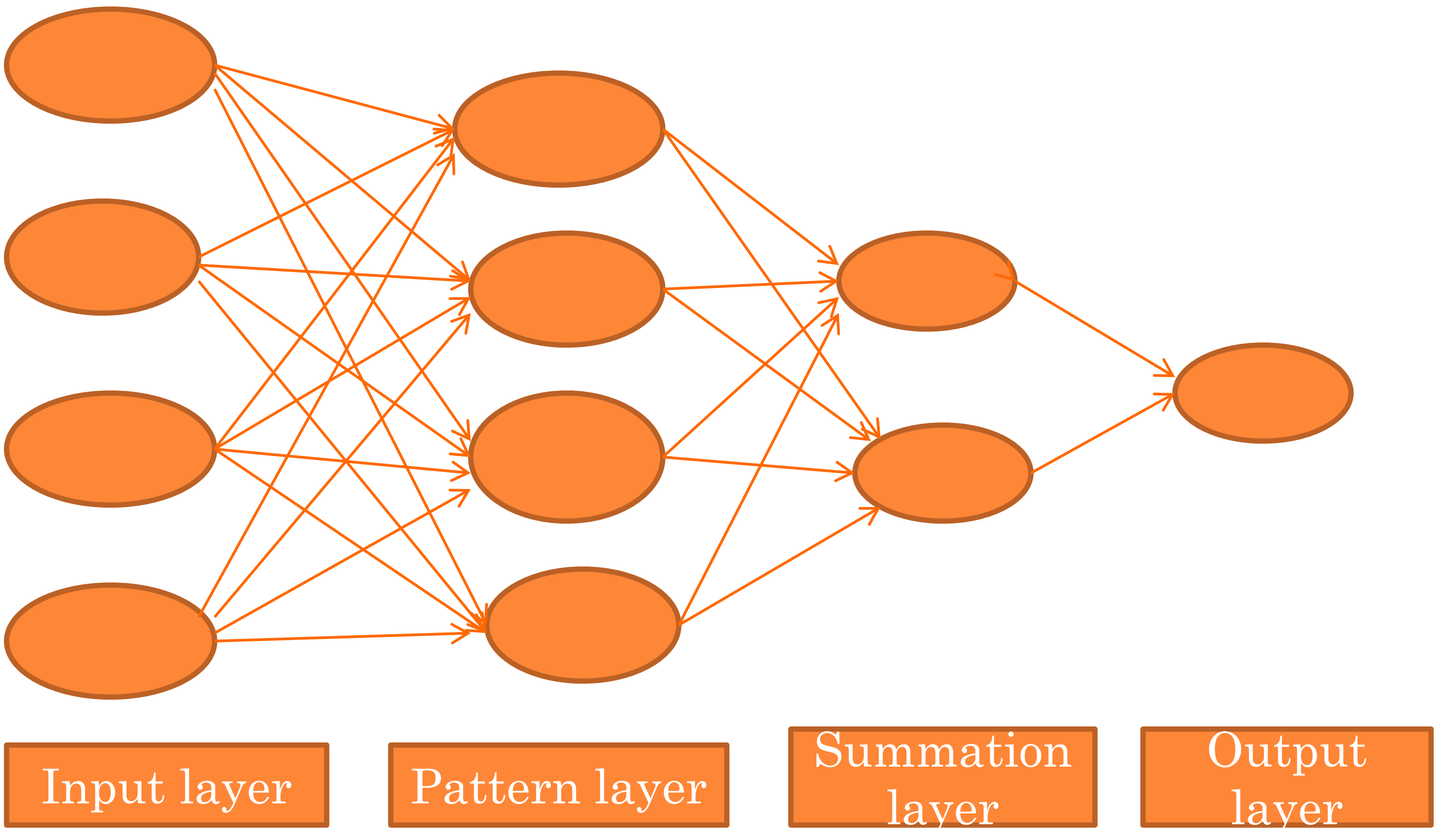
Taking $D_i^2 = (x - x_i)^T (x - x_i)$ and after the integration the expression for the conditional mean obtained, by Specht(1991), is

$$\hat{Y}(x) = \frac{\sum_{i=1}^{n} y_i \exp\left(-\frac{D_i^2}{2\sigma^2}\right)}{\sum_{i=1}^{n} \exp\left(-\frac{D_i^2}{2\sigma^2}\right)} \quad (3)$$

Parzen (1962) and Cacoullos (1966) have shown the consistency and asymptotic convergence of the estimate to the true value at all sample points where the density is continuous provided $\sigma(n) \rightarrow 0 \; as \; n \rightarrow \infty$ and $n\sigma^p(n) \rightarrow \infty \; as \; n \rightarrow \infty$. Specht (1991) had based it on the Gaussian kernel function.

Therefore, the estimated conditional mean is a weighted average of observed y_i's where each observation is weighted exponentially according to its Euclidean distance from **x**.

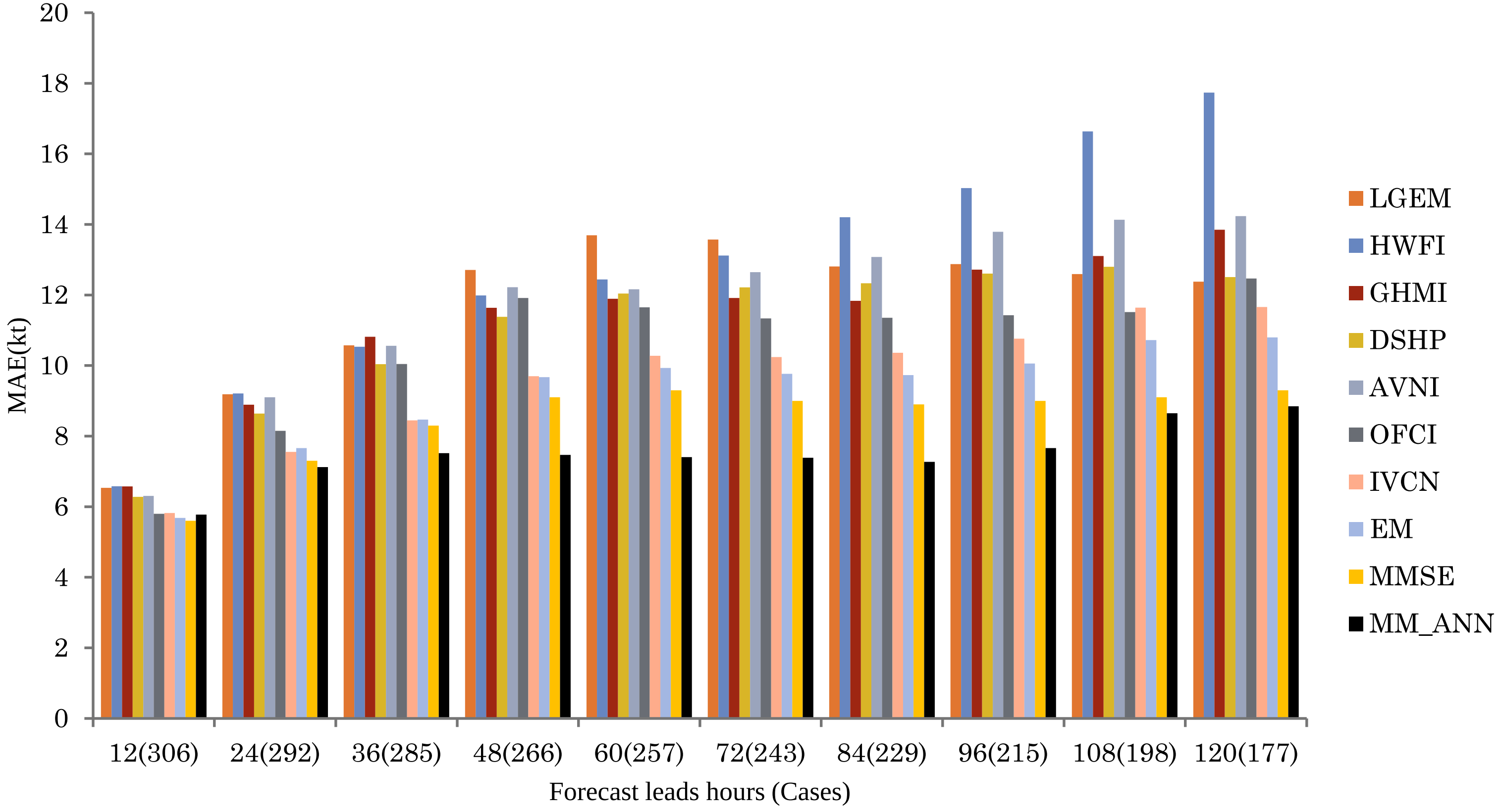
GRNN: SCHEMATIC PRESENTATION



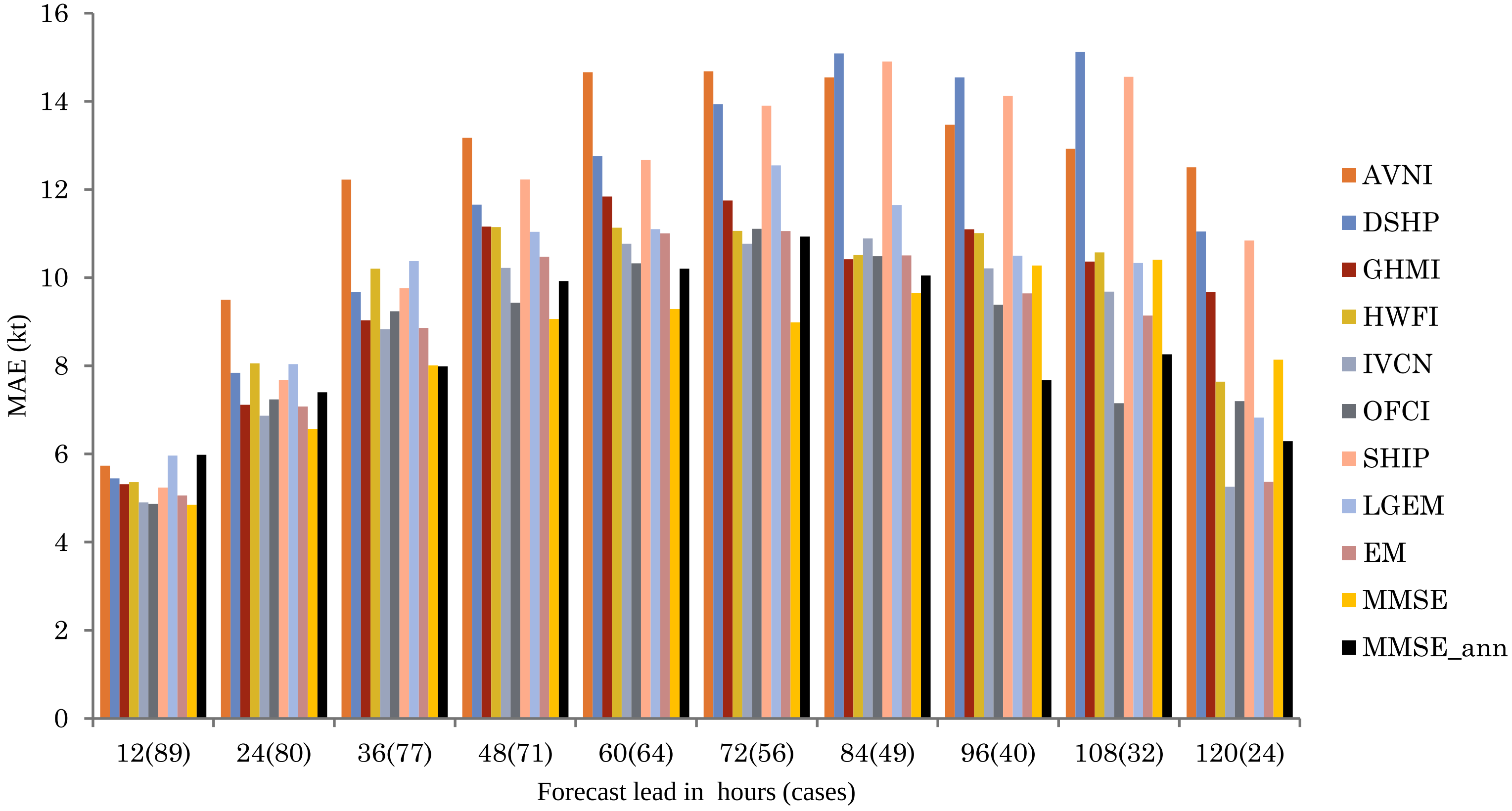
Advantages

- No user choice for the network architecture
- Only one parameter to be estimated
- Does not get trapped into the local optima
- Requires less number of data for training
- Useful for continuous data

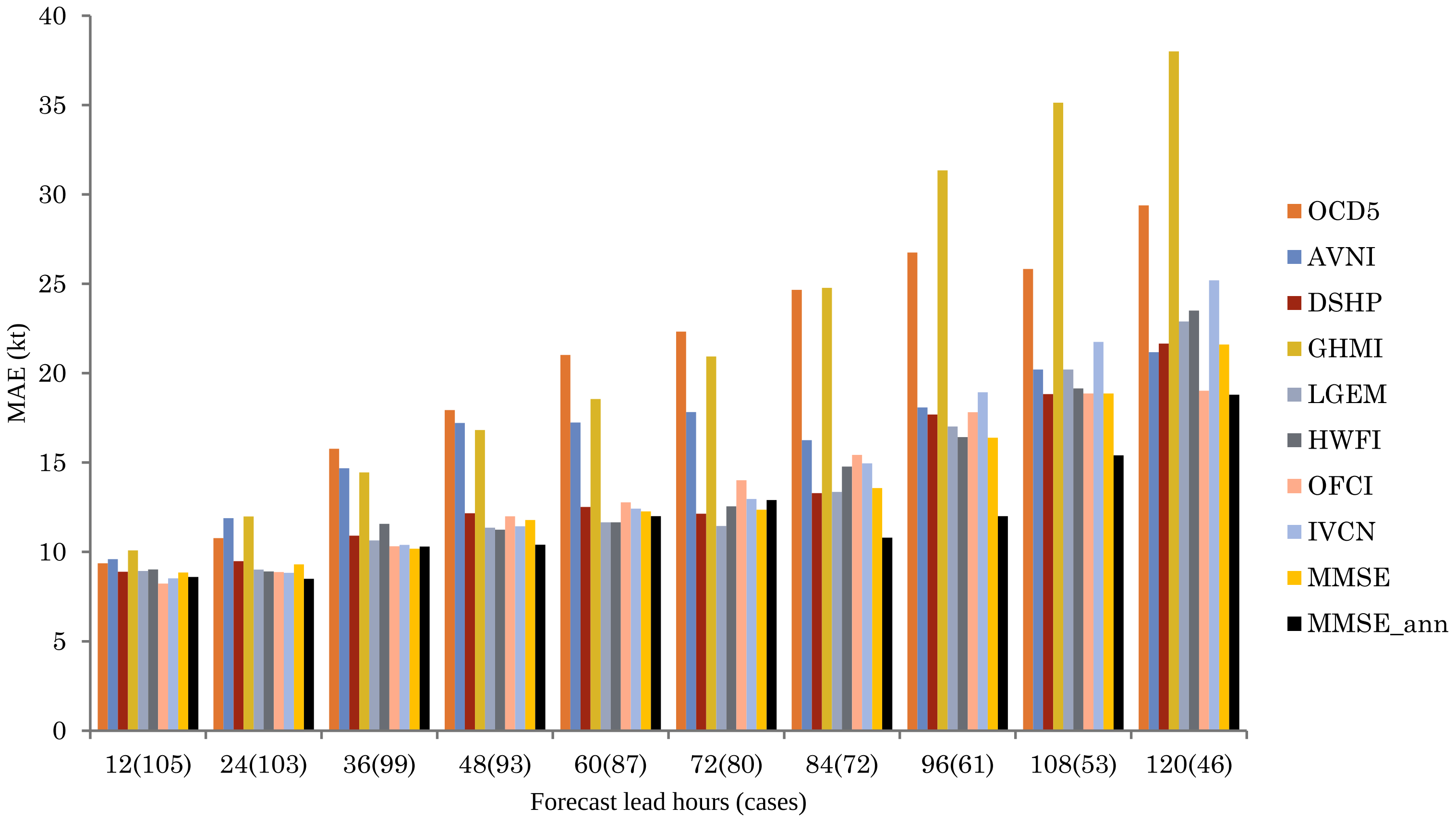
Results: Season 2012 intensity errors



Intensity errors: Season 2014

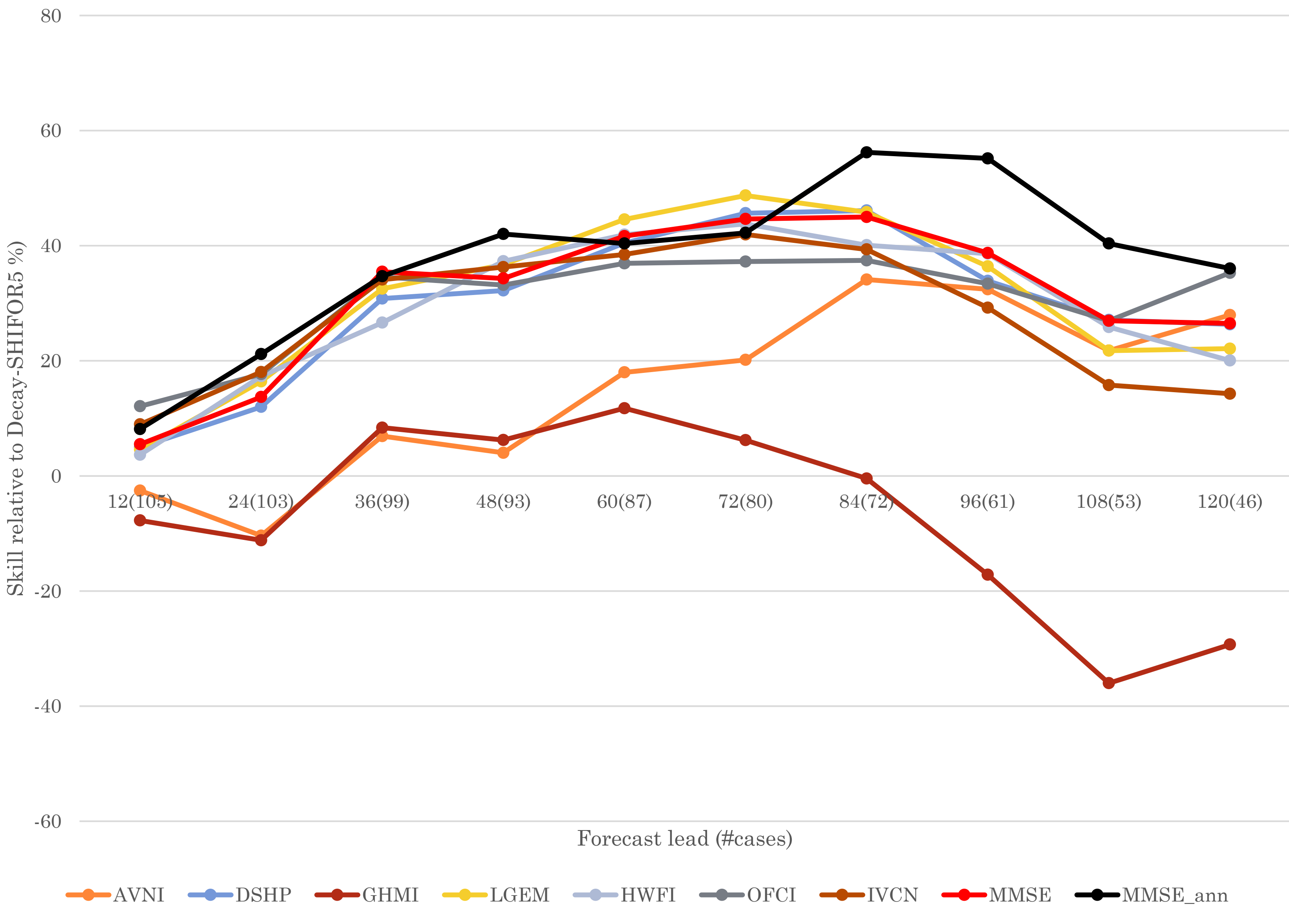


Season 2015

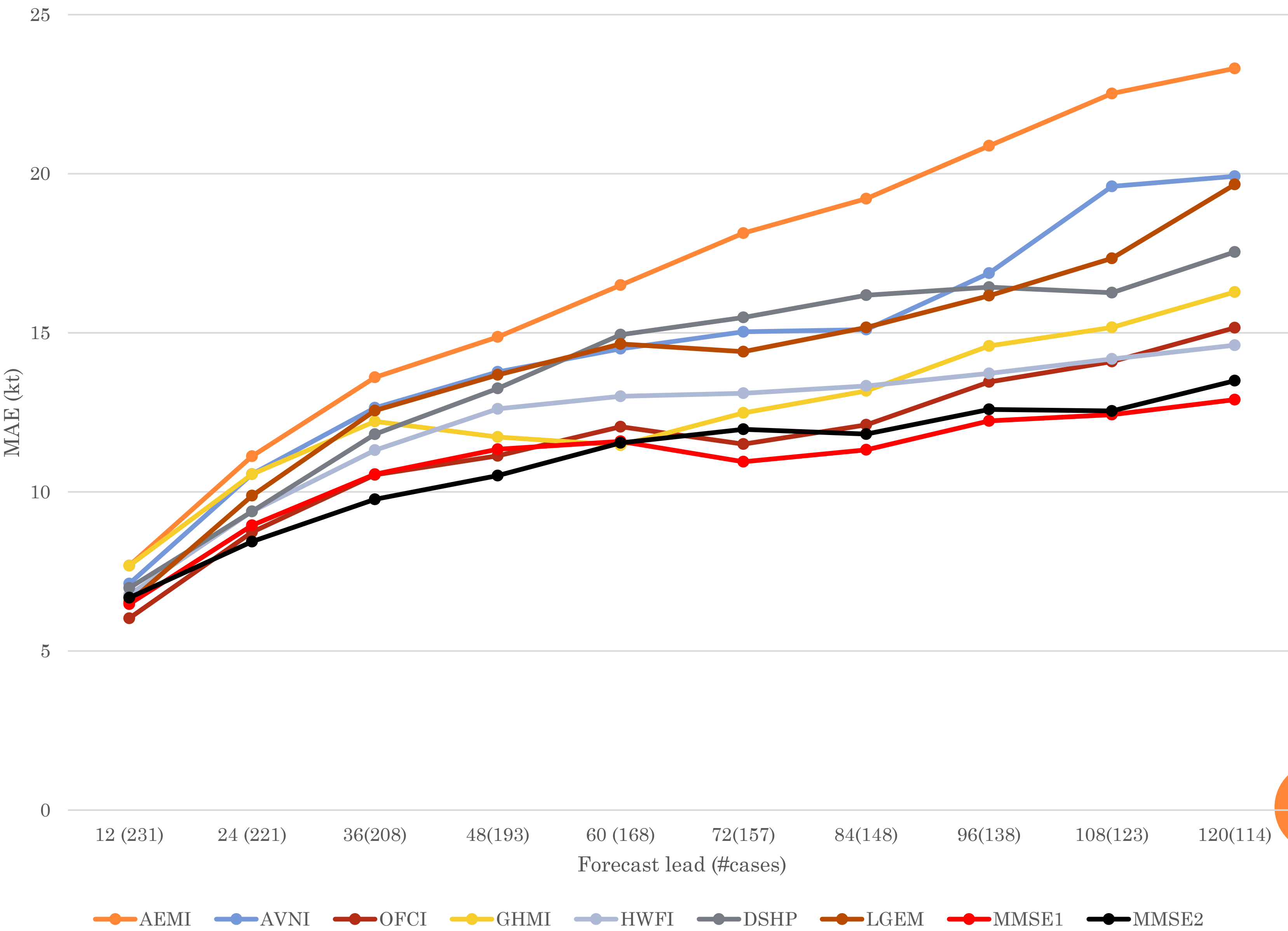


40
35
30
25
20
15
10
5
0
MAE (kt)
12(105)
24(103)
36(99)
48(93)
60(87)
72(80)
84(72)
96(61)
108(53)
120(46)
Forecast lead hours (cases)
OCD5
AVNI
DSHP
GHMI
LGEM
HWFI
OFCI
IVCN
MMSE
MMSE_ann

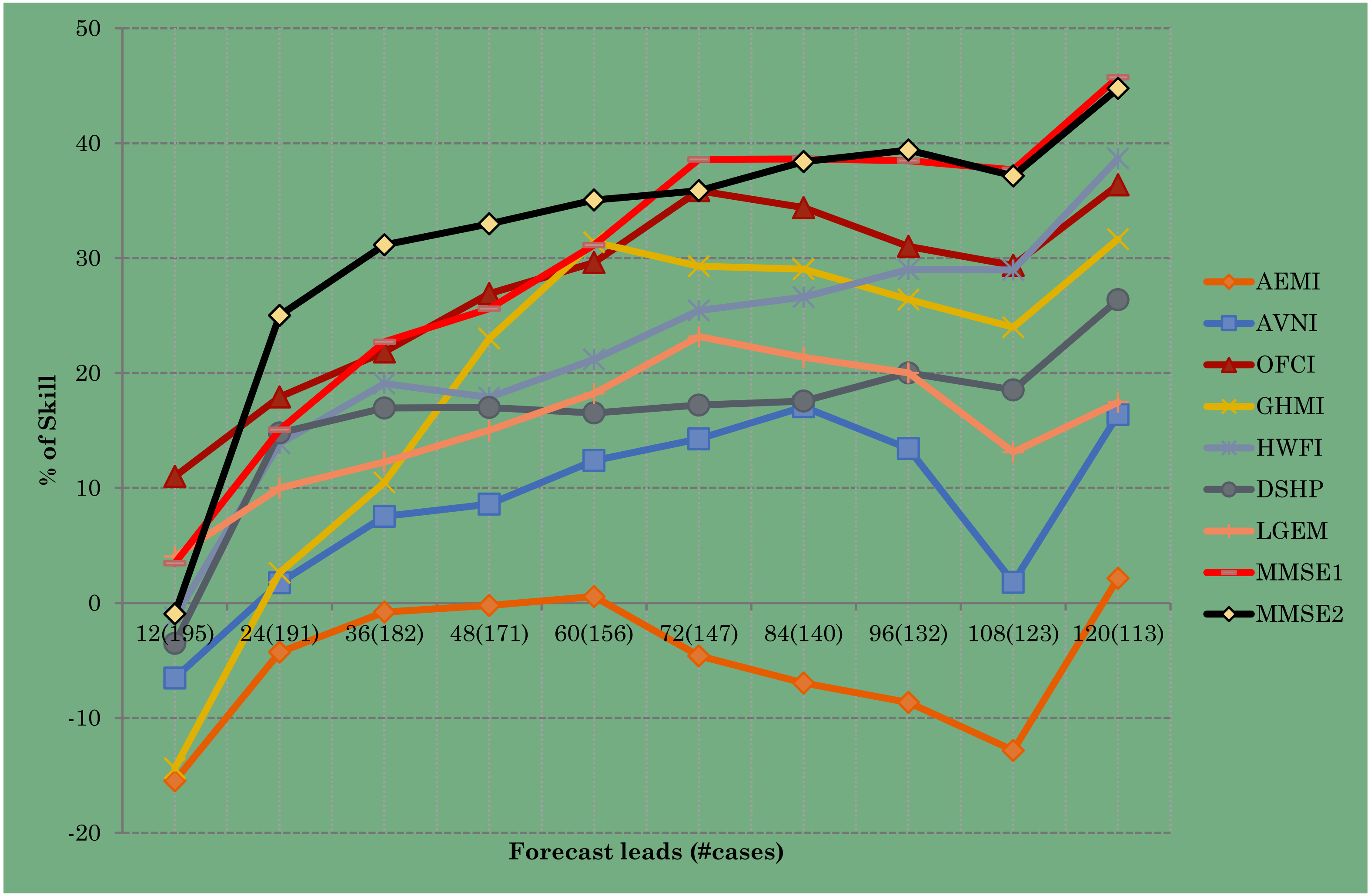
Model Skills: 2015 Season



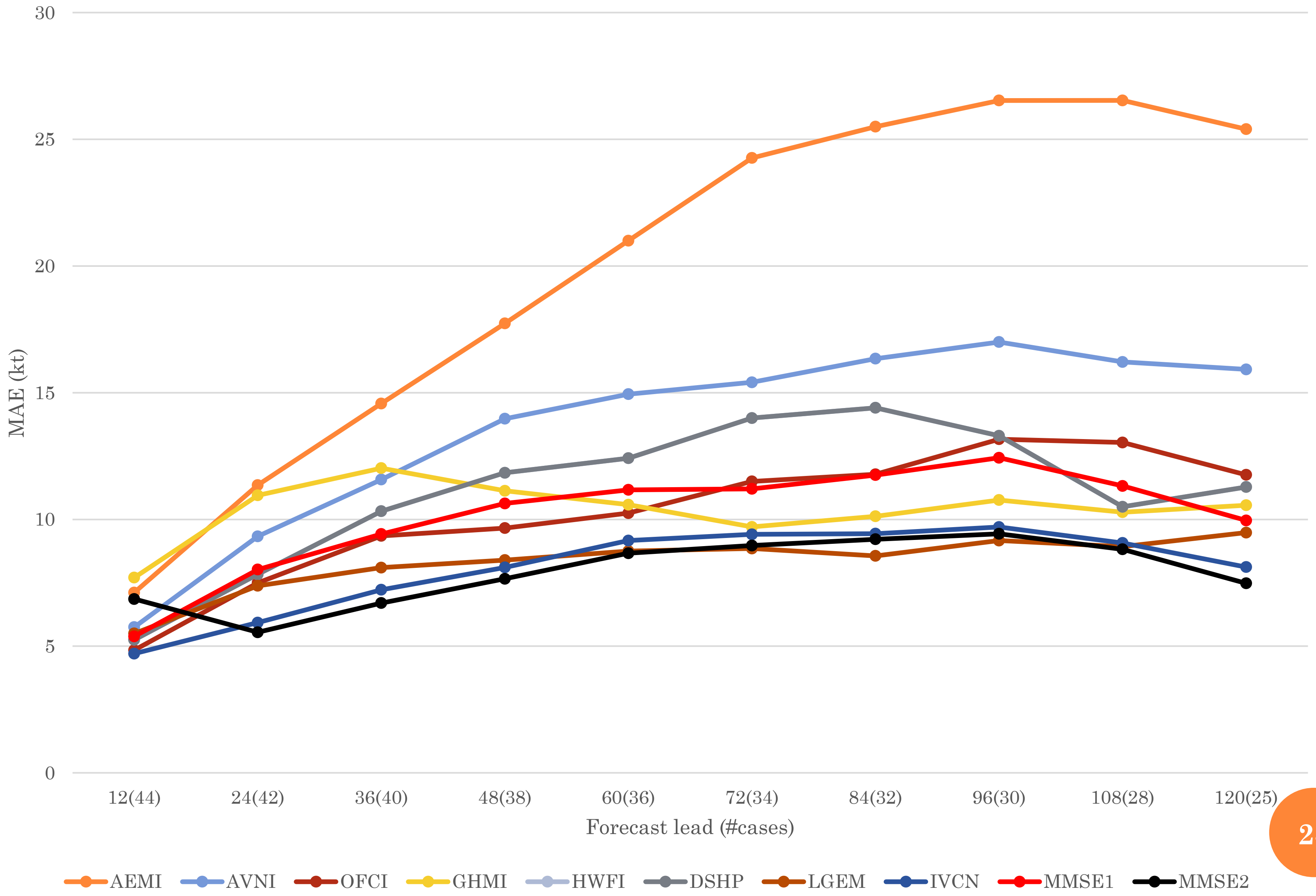
SEASON 2016: INTENSITY ERRORS



Skills: 2016 season

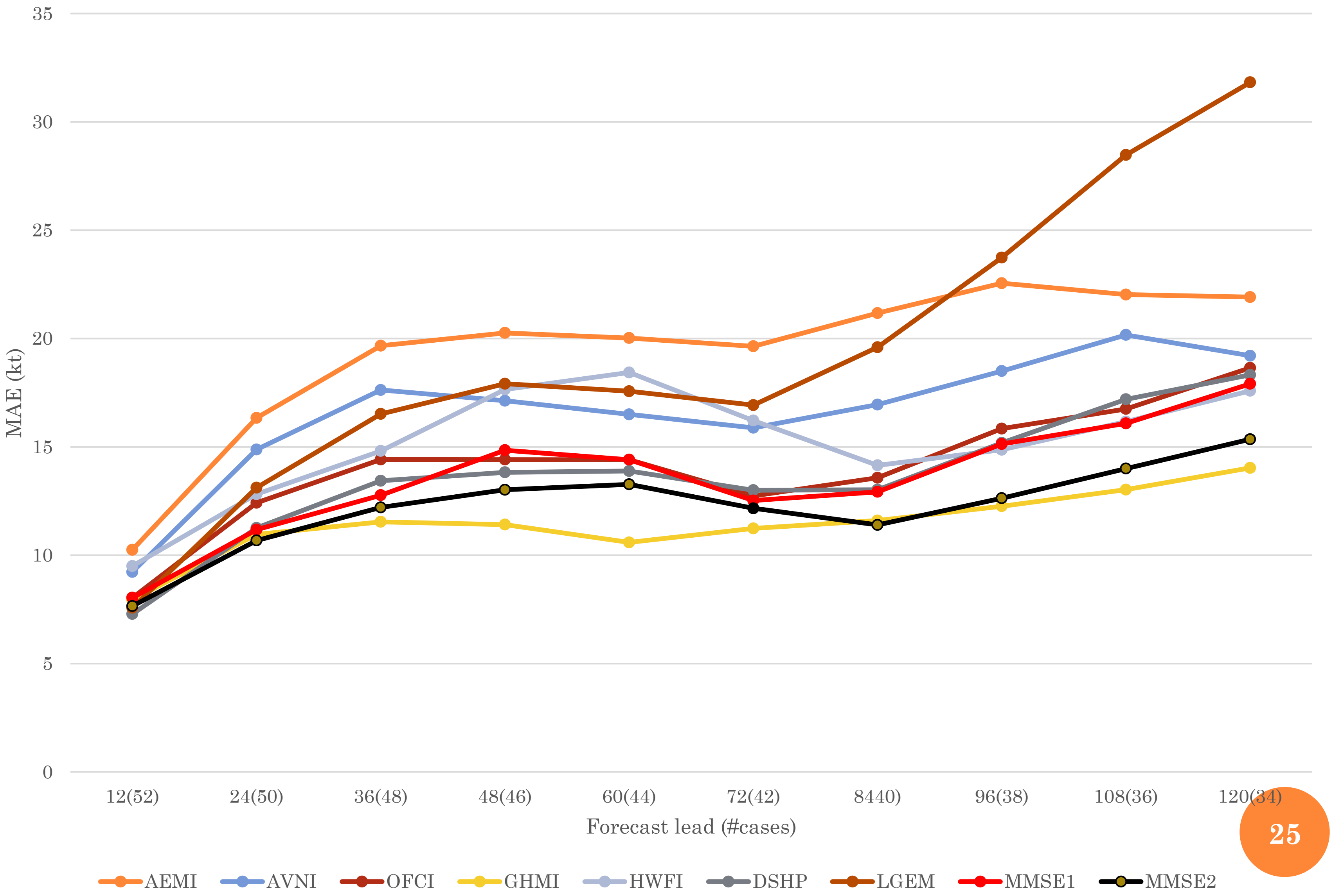


Gaston 2016



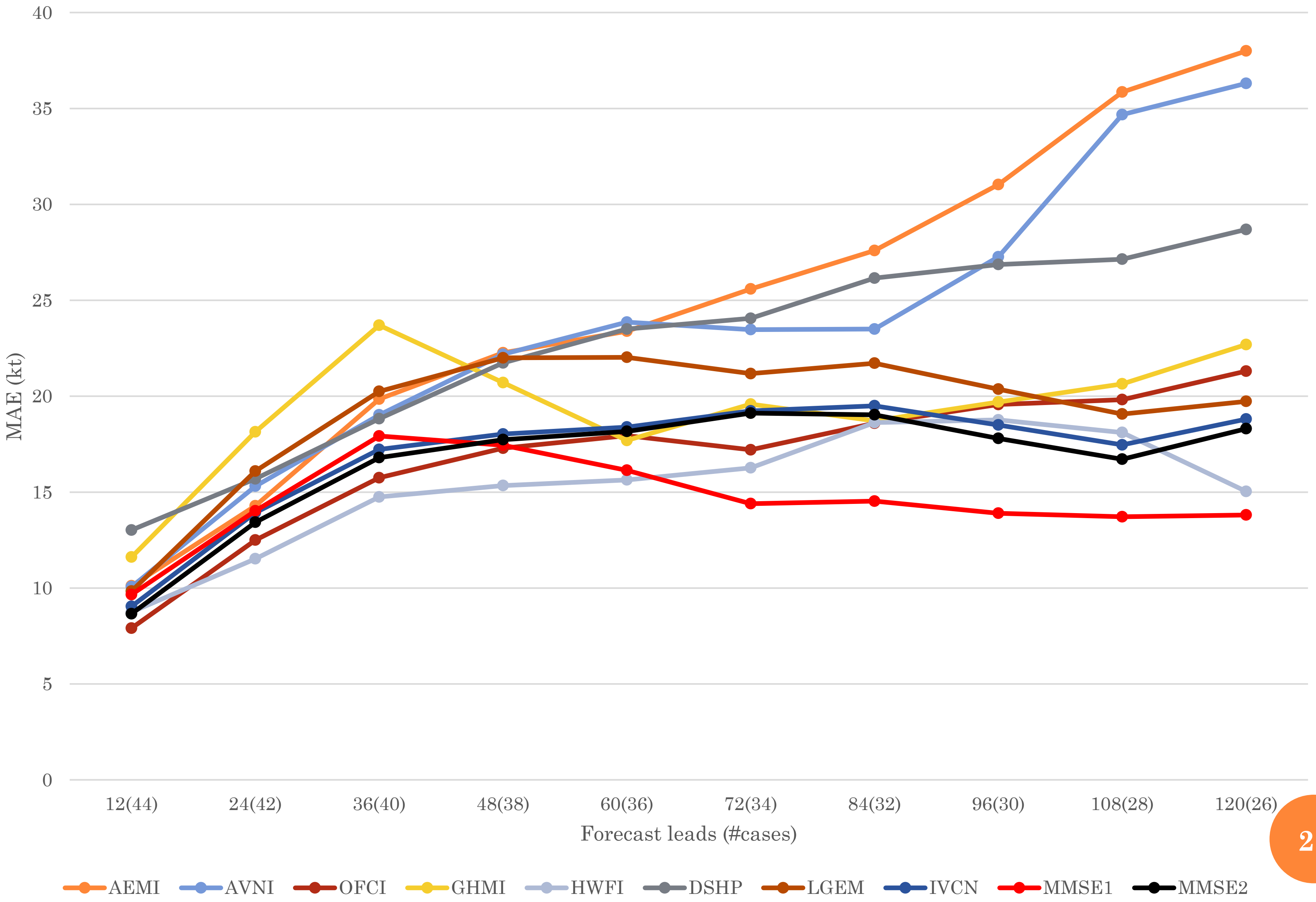
30
25
20
15
10
5
0
MAE (kt)
12(44)
24(42)
36(40)
48(38)
60(36)
72(34)
84(32)
96(30)
108(28)
120(25)
Forecast lead (#cases)
AEMI
AVNI
OFCI
GHMI
HWFI
DSHP
LGEM
IVCN
MMSE1
MMSE2

Nicole 2016



35
30
25
20
15
10
5
0
MAE (kt)
12(52)
24(50)
36(48)
48(46)
60(44)
72(42)
8440)
96(38)
108(36)
120(34)
Forecast lead (#cases)
AEMI
AVNI
OFCI
GHMI
HWFI
DSHP
LGEM
MMSE1
MMSE2

Matthew 2016



Concluding remarks

- Seasonal summaries indicate that the improved MMSE carries, **consistently**, least intensity forecast errors
- For **longer forecast leads**, beyond 60hrs, Neural Networked based MMSE performs better than the earlier forecast leads. It is very useful for government planning and evacuation, if needed
- Individual storm forecast errors show that none of the models is consistently best

Concluding remarks contd …

- Improved MMSE is the best or the second best performer for individual storms as well
- Proposed method is providing consistent consensus forecasts having least forecast errors which be depended upon
- Ensemble forecasts based on neural networks may be considered for real-time forecast guidance in case of hurricane and tropical storms
- Forecasting of tracks may also be examined

Thank you