Probabilistic Solar Proxy Forecasting with Neural Network Ensembles

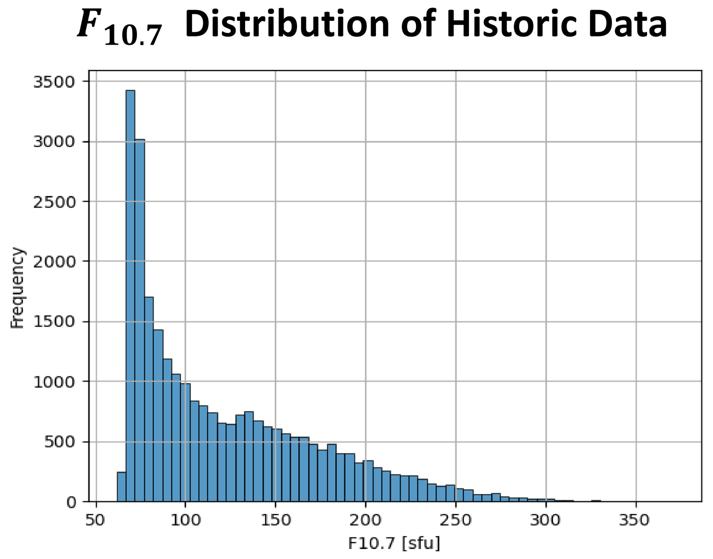
Joshua D. Daniell and Piyush M. Mehta**

Background & Our Work

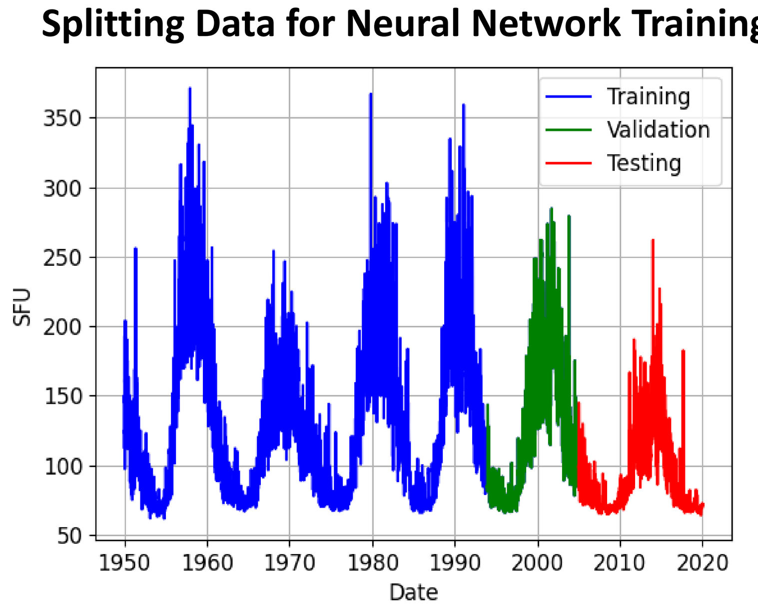
$F_{10.7\,cm}$ is an input (driver) for almost all thermosphere density models, including the operational HASDM. The thermosphere density models majorly impact satellite drag calculations. The USAF contracts a prediction of the proxy using a linear algorithm.[1] Previous non-linear *AR* methods are successful at forecasting $F_{10.7}$ but are aimed at longer term forecasting.[2] This work develops an **improved short term forecasting method of $F_{10.7}$ using neural network ensembles,** we find that the new probabilistic approach provides **significantly better relative metrics** compared against the current linear approach. Our novel work provides improvements in forecasting and includes **robust and reliable uncertainty estimation for the predicted proxy.**

Data

$F_{10.7}$ Distribution of Historic Data



Splitting Data for Neural Network Training



- †Data: (daily $F_{10.7\,cm}$ observed values) 1948-2020.
- For machine learning data splitting must be done carefully!
  - ➢ Want to maximize the size of the training set.
  - ➢ Want to include all activity levels in split sets.
- Auto-Regressive models forecast $F_{10.7}$ using only past values.

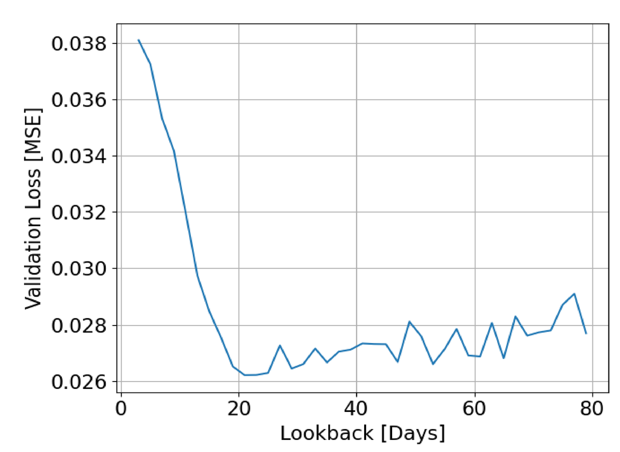
Sliding Window Approach

Forecast Epoch	Day									
Day 4	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Day 10
Day 5	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Day 10
Day 6	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Day 10
Day 7	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Day 10
Day 8	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Day 8	Day 9	Day 10

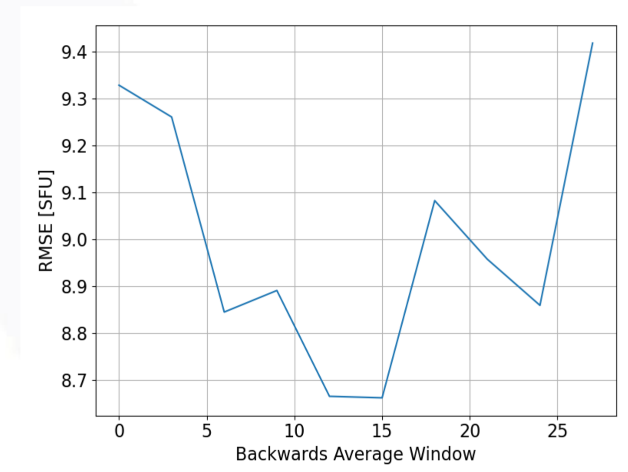
Current Day | Lookback | Horizon Values | Unused for Step

Methods: Input Investigation

Variation of Lookback



Variation of Backwards Average



- How can we extract more information, while only using historical $F_{10.7\,cm}$ data?
- *Varied Lookback*: Include variety of previous daily values as inputs to identify more trends.
- *Backwards Average*: Allows short term trend to be input to models directly.
- *AR* methods are limited by lack of varied information. We lack "the whole picture".

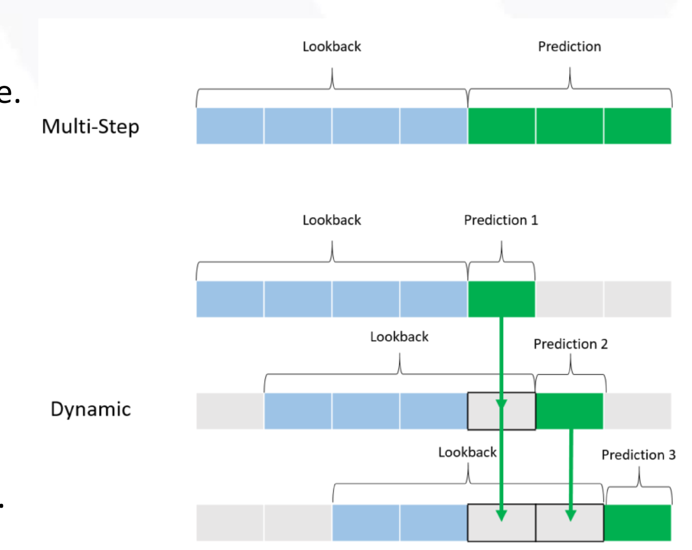
Methods: Forecasting: Multi-Step vs. Dynamic

Multi-Step

- Trained to predict multiple time steps at once.
- Prediction step occurs 1 time per day.
- Trained & limited to a fixed forecast horizon.

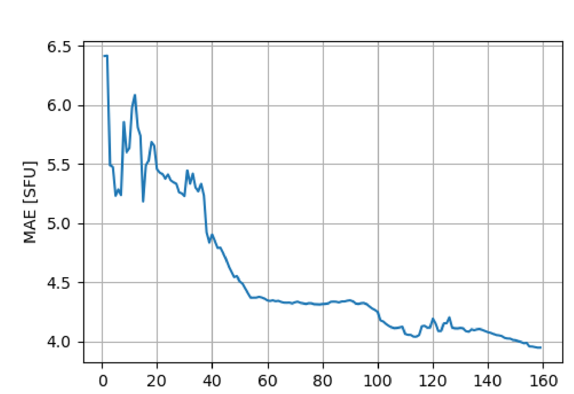
Dynamic

- Trained to predict single time step values.
- Prediction occurs recursively 6 times per day.
- Horizon can be changed during forecasting.

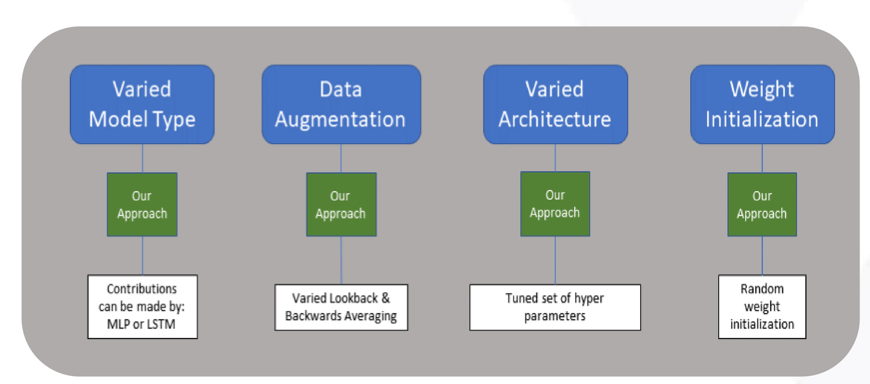


Methods: Ensembles & Diversity

Model Ensembles Decrease Error



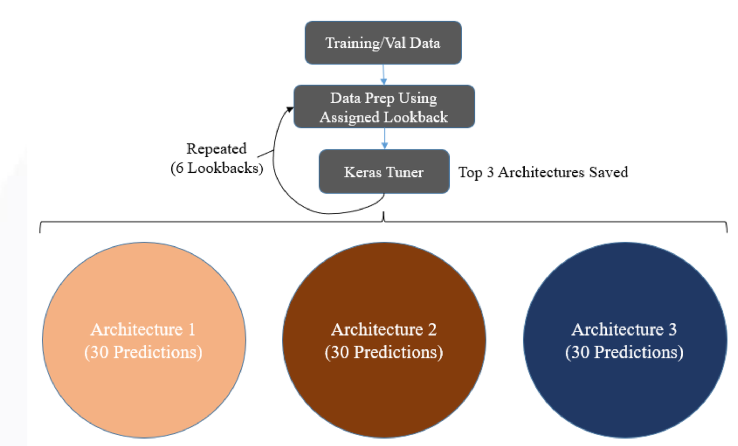
Diversity Strategies for Ensembles



- Individual model forecasts are deterministic, *model ensemble forecasts are probabilistic*.
- Ensemble members can be skilled in different areas which provides a diverse set of predictions.
- Multiple predictions are equally weighted to form an ensemble prediction, in this case 180 models.
- *A 48% decrease in MAE between 1 model and 180 models, demonstrates the skill of an ensemble.*

Methods: Neural Network Models (Ensemble Members)

Generating Ensemble Members



Tunable Hyper Parameters

MLP (Multi Layer Perceptron)

Tuner Option	Choice	Parameter	Value/Range
Scheme	Bayesian Optimization	Number of Dense Layers	[1-8]
Total Trials	100	Dense Neurons	[4-256]
Initial Points	5	Activation	[relu,tanh,sigmoid]
Repeats per Trial	3	Optimizer	[adam, SGD]
Minimization Parameter	MSE	Learning Rate	[.01, .001, .0001]
Epochs	100	Early Stopping Criteria	validation loss (MSE)
Early Stopping Patience	30		

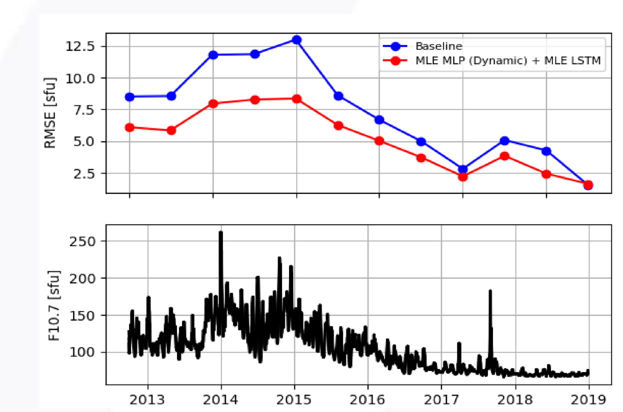
LSTM (Long-Short Term Memory)

Tuner Option	Choice	Parameter	Value/Range
Scheme	Bayesian Optimization	Number of LSTM Layers	[1-3]
Total Trials	75	LSTM Neurons	[4-128]
Initial Points	5	LSTM Activation	tanh
Repeats per Trial	3	Number of Dense Layers	[1-4]
Minimization Parameter	MSE	Dense Neurons	[4-256]
Epochs	50	Dense Activations	[relu, tanh, sigmoid]
Early Stopping Criteria	validation loss (MSE)	Learning Rate	[.01, .001, .0001]
Early Stopping Patience	10	Optimizer	[adam, SGD]

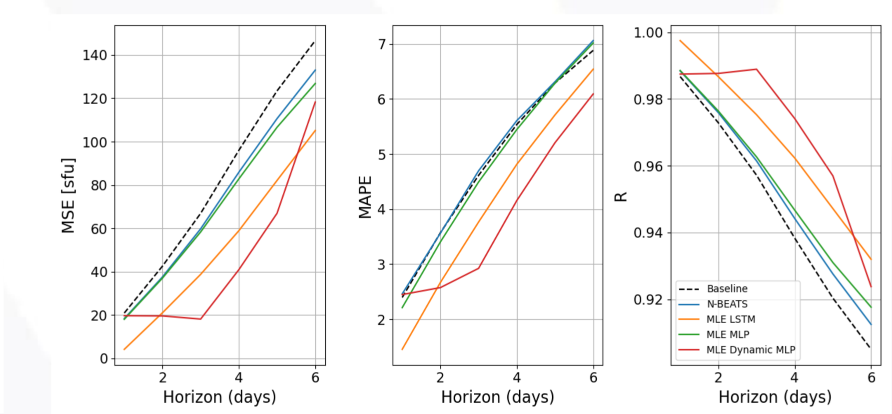
- Mean Squared Error (MSE) was used as loss function during model selection and training.
- Set of 6 lookbacks were used [7, 10, 13, 16, 19, 22 days] to promote model diversity (differing skill areas).
- The top 3 hyper parameter sets (*architectures*) are then chosen for each lookback for diversity.
  - ➢ Top 3 architecture choice made based on minimizing validation loss after hyper parameter tuning.
- 10 models of a given architecture with randomly initialized weights are trained and saved.

Results: Proxy Forecasting

Comparison of 180 Day Averaged Error



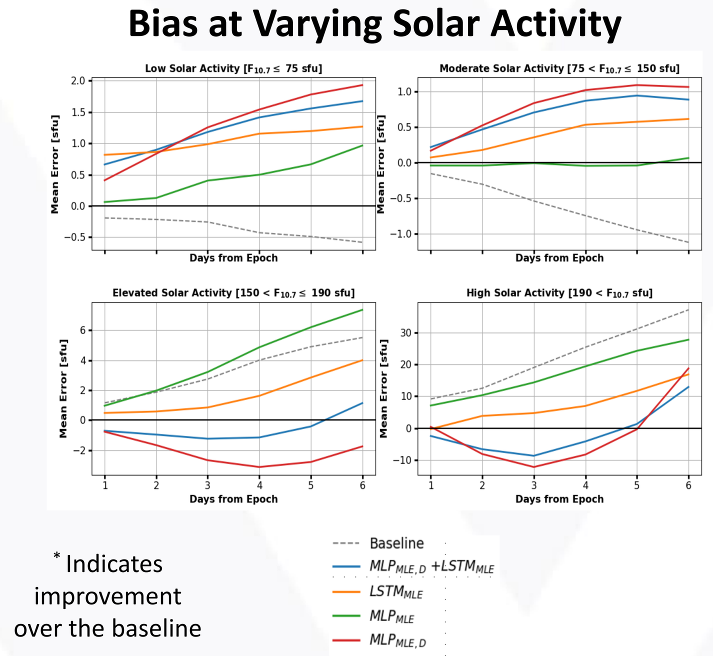
Performance Metrics Over Horizon



- Varied model types (MLP & LSTM) for predictions.
- 180 models used to form ensemble forecast.
- Ensemble methods improved performance over the baseline linear method.
- As expected, forecast error changes over time.

Model	* Relative Mean Squared Error (MSE)		
	Test (2006-2020)	*Validation (1994-2004)*	*Training Sample (1964-1974)*
$MLP_{MLE\ D} + LSTM_{MLE}$	44.2%	46.2%	55.5%

Bias at Varying Solar Activity



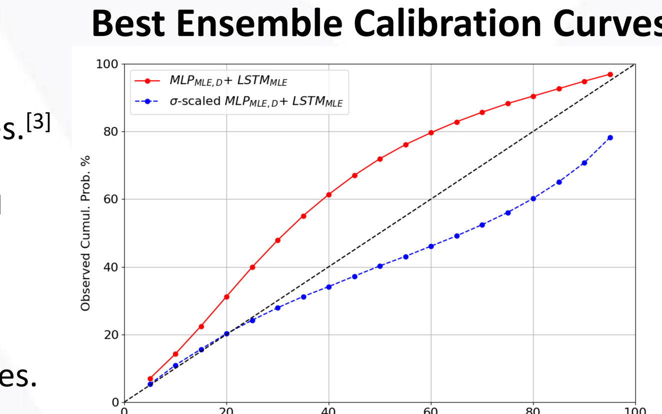
* Indicates improvement over the baseline

Results: Robustness and Reliability of Uncertainty Estimates

- Mean error (bias) indicates temporal uncertainty.
- CES quantifies uncertainty estimates made by ensembles.[3]
- σ scaling applies correction factor using validation set.[4]
- Uncertainty estimates are robust and reliable.
  - ➢ Slight under prediction seen at higher percentages.

Calibration Error Score (CES)			
Model	*Test (2006-2020)*	*Validation (1994-2004)*	*Training Sample (1964-1974)*
$MLP_{MLE\ D} + LSTM_{MLE}$	13.4%	14.34%	18.20%
$MLP_{MLE\ D} + LSTM_{MLE}$ (σ scaling)	9.8 %	5.22 %	14.51%

Best Ensemble Calibration Curves



$$\sigma_S = \sqrt{S} * \sigma = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \sigma_i^{-2} * (y - \hat{y})^2} * \sigma$$

$$CES = \frac{100\%}{m}\sum_{i=1}^{m}\left|p(\alpha_j) - p(\hat{\alpha}_j)\right|$$

Conclusions

- AR data can be manipulated to improve performance.
- Mean Error (bias) is decreased at higher solar activity levels.
- NN ensemble techniques allow for probabilistic forecasting.
- Uncertainty estimates from probabilistic forecasts can be evaluated; both quantitatively and qualitatively.
- *The best NN ensemble method provided more than a 50% improvement over SET's linear forecasting method.*

Future Work

- Investigate advanced ensemble combination techniques.
  - ➢ Temporal debiasing
  - ➢ Performance based weighting
- Apply ensemble techniques to other space weather indices and/or proxies.
  - ➢ Simultaneous multivariate forecasting
- Investigate te coupling of different sources of uncertainties.
  - ➢ Model uncertainty & Driver uncertainty

References

[1] Licata, R. J., Tobiska, W. K., & Mehta, P. M. (2020). "Benchmarking forecasting models for space weather drivers." (2020).

[2] Stevenson, Emma et al. "A deep learning approach to solar radio flux forecasting." (2022).

[3] Licata, R. J., et al. "Machine-learned HASDM thermospheric mass density model with uncertainty quantification." (2022).

[4] Laves, Max-Heinrich, et al. "Recalibration of Aleatoric and Epistemic Regression Uncertainty in Medical Imaging." (2021).

Acknowledgements

JDD and PMM acknowledge the support under NSF ANSWERS award # 2149747575.

†The results presented in this work rely on data collected by the Solar Radio Monitoring operated by the National Research Council and Natural Resources Canada. These data were accessed via the LASP Interactive Solar Irradiance Datacenter (LISIRD).

West Virginia University

* Department of Mechanical and Aerospace Engineering

NSF

SPACE SCIENCE SPACE TECHNOLOGY ASTRODYNAMICS