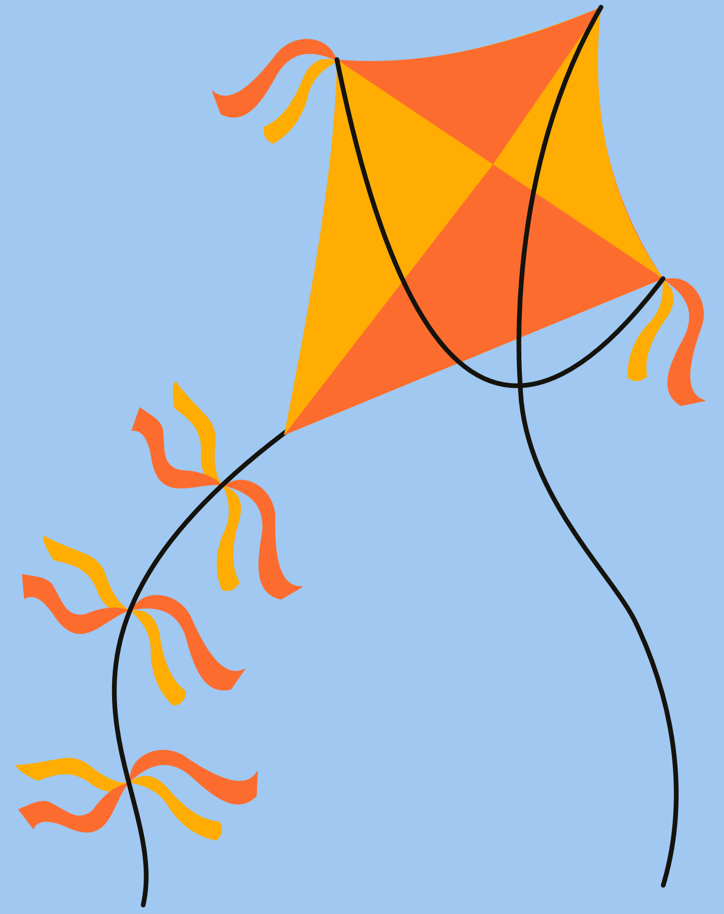


Click here to access the public view link of our slides!

TIME SERIES
WEATHER FORECASTING

PREDICTING RELATIVE HUMIDITY

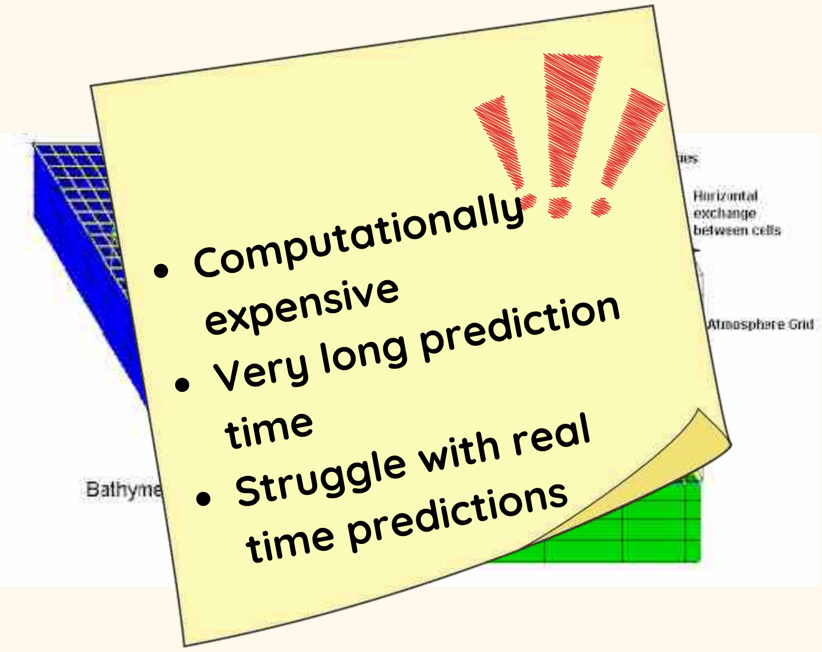
IT1244 Group 30
Jennifer, Yun Han, Shi Ying

INTRODUCTION

WHY IS WEATHER FORECASTING IMPORTANT?

- Atmospheric conditions
- Human comfort
- Agricultural productivity
- Disaster Management
- Making daily decisions

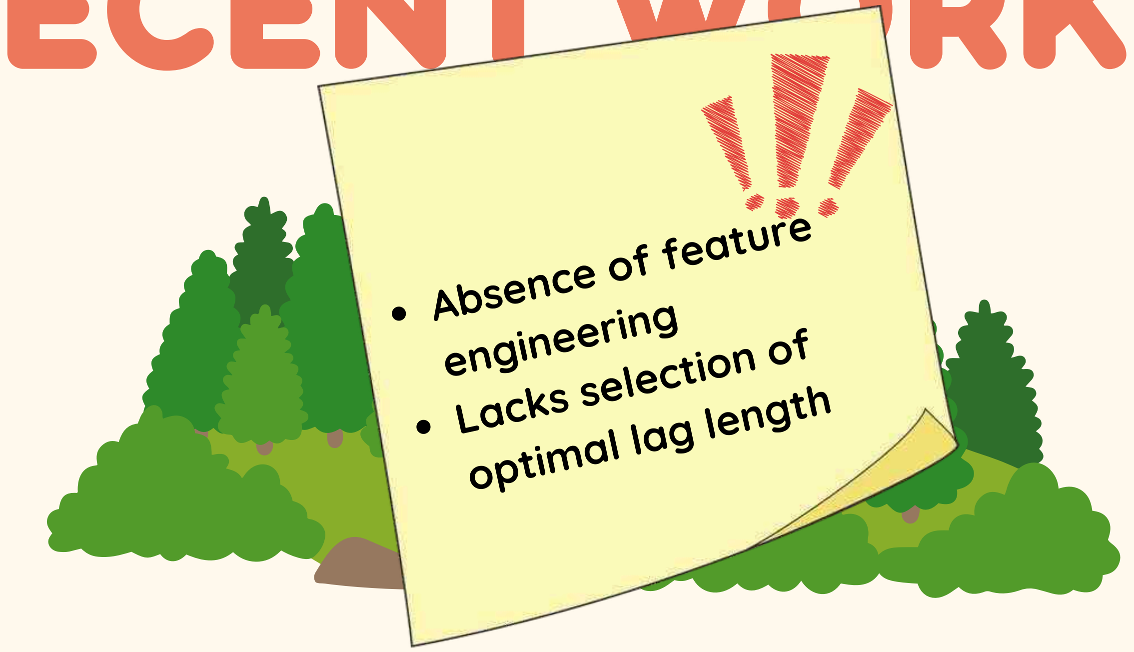
RECENT WORKS



- Computationally expensive
- Very long prediction time
- Struggle with real time predictions

NUMERIAL WEATHER PREDICTION (NWP) MODEL

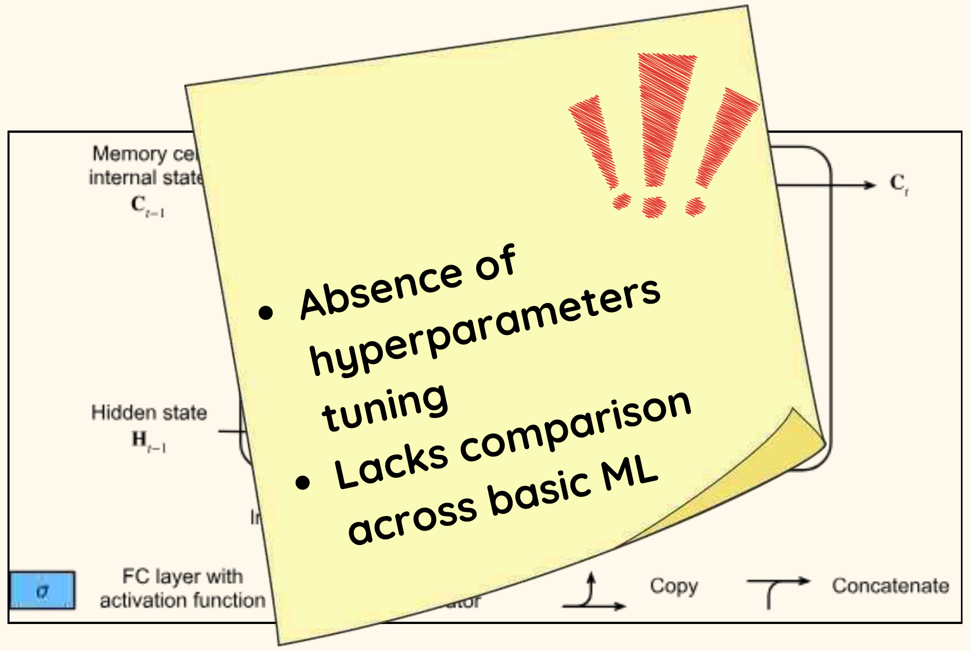
Computer simulations using complex equations.



- Absence of feature engineering
- Lacks selection of optimal lag length

ML MODEL: RANDOM FORESTS

Ensemble of decision Trees



- Absence of hyperparameters tuning
- Lacks comparison across basic ML

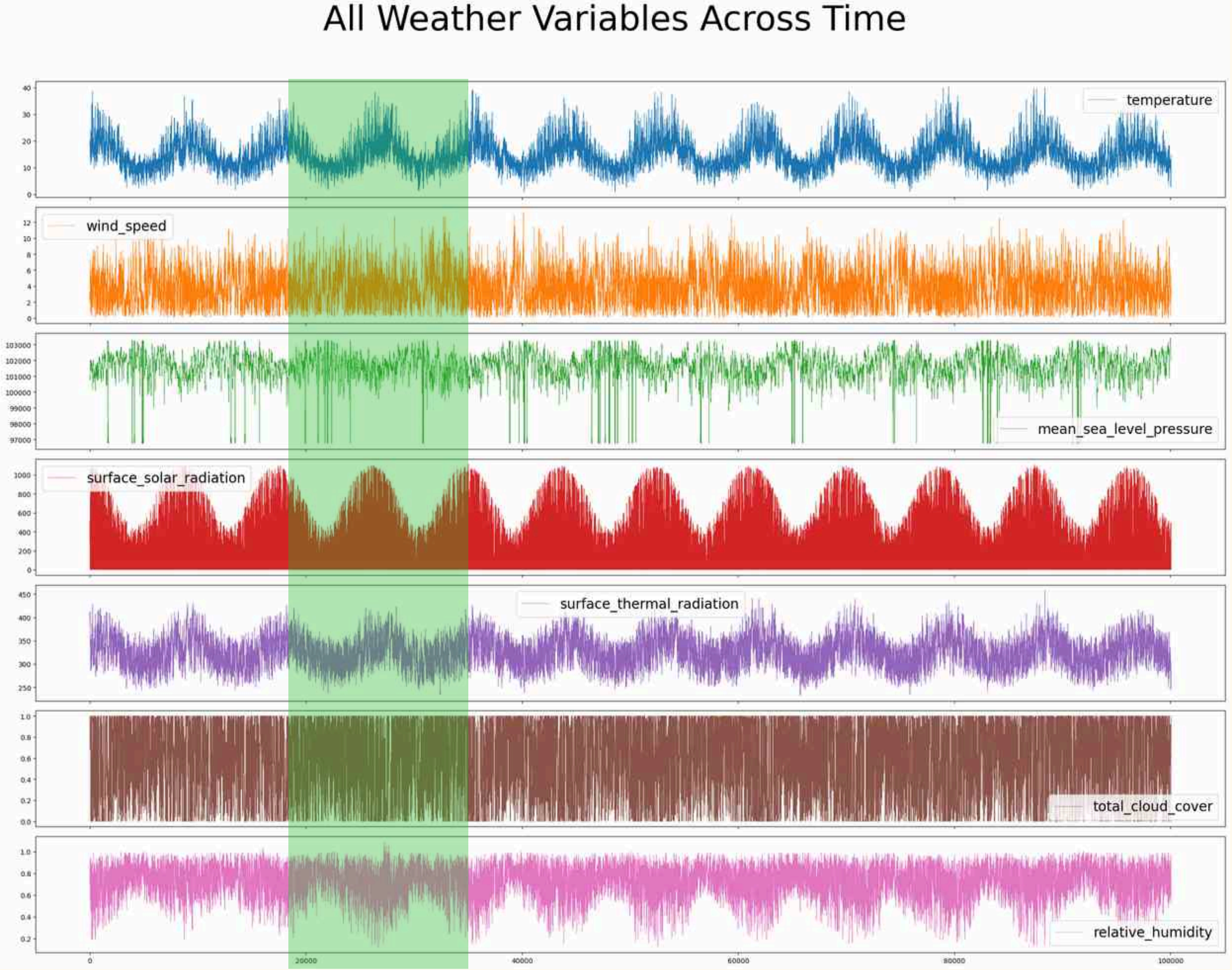
NEURAL NETWORK: LSTM

Artificial neural network

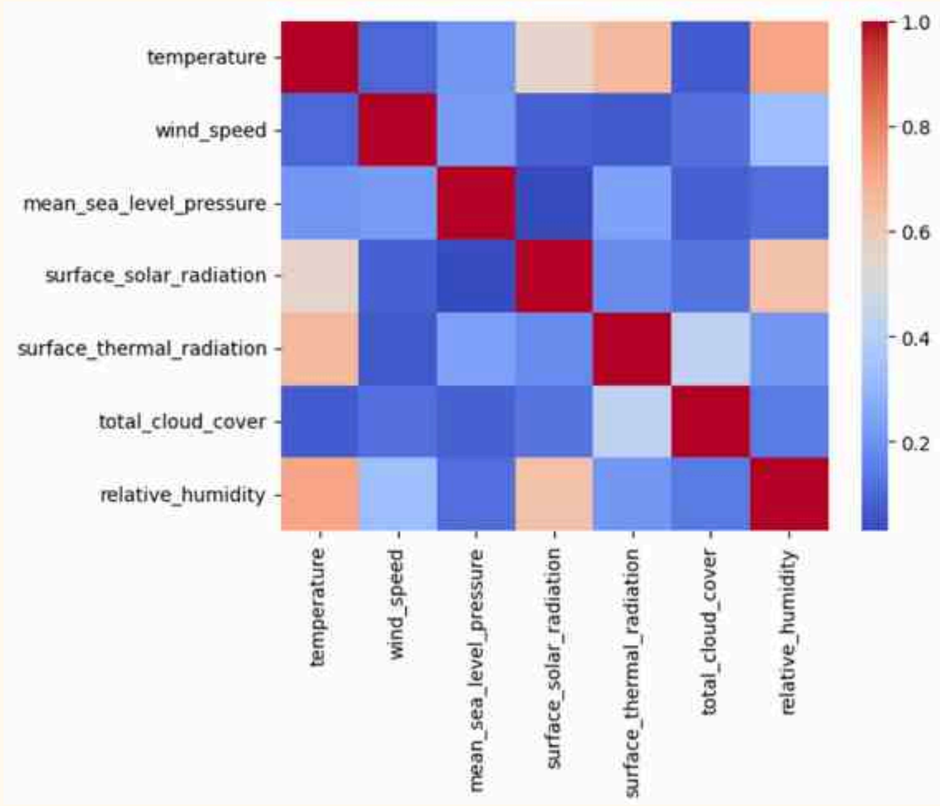
DATASET

- 100,057 rows and 7 weather parameters
- Recorded hourly from January 1, 2010, to June 1, 2021 at Monash University in Australia
- No missing values.

VISUALISATIONS



CORRELATION ANALYSIS



- Collinearity between features not significant
- Keeping these 7 features to allow the models to capture the non-linear relationships.

Seasonal pattern:

- Peaks in summer and falls in winter
  - temperature
  - surface_solar_radiation
  - surface_thermal_radiation
- Peaks in winter and falls in summer
  - relative_humidity
- 11.5 years → 11.5 peaks

EXAMINING DISTRIBUTION

HISTOGRAM OF DISTRIBUTION

Roughly normal distributions → performed standardisation

- temperature
- wind_speed
- surface_thermal_radiation

Created dummy values for anomaly values

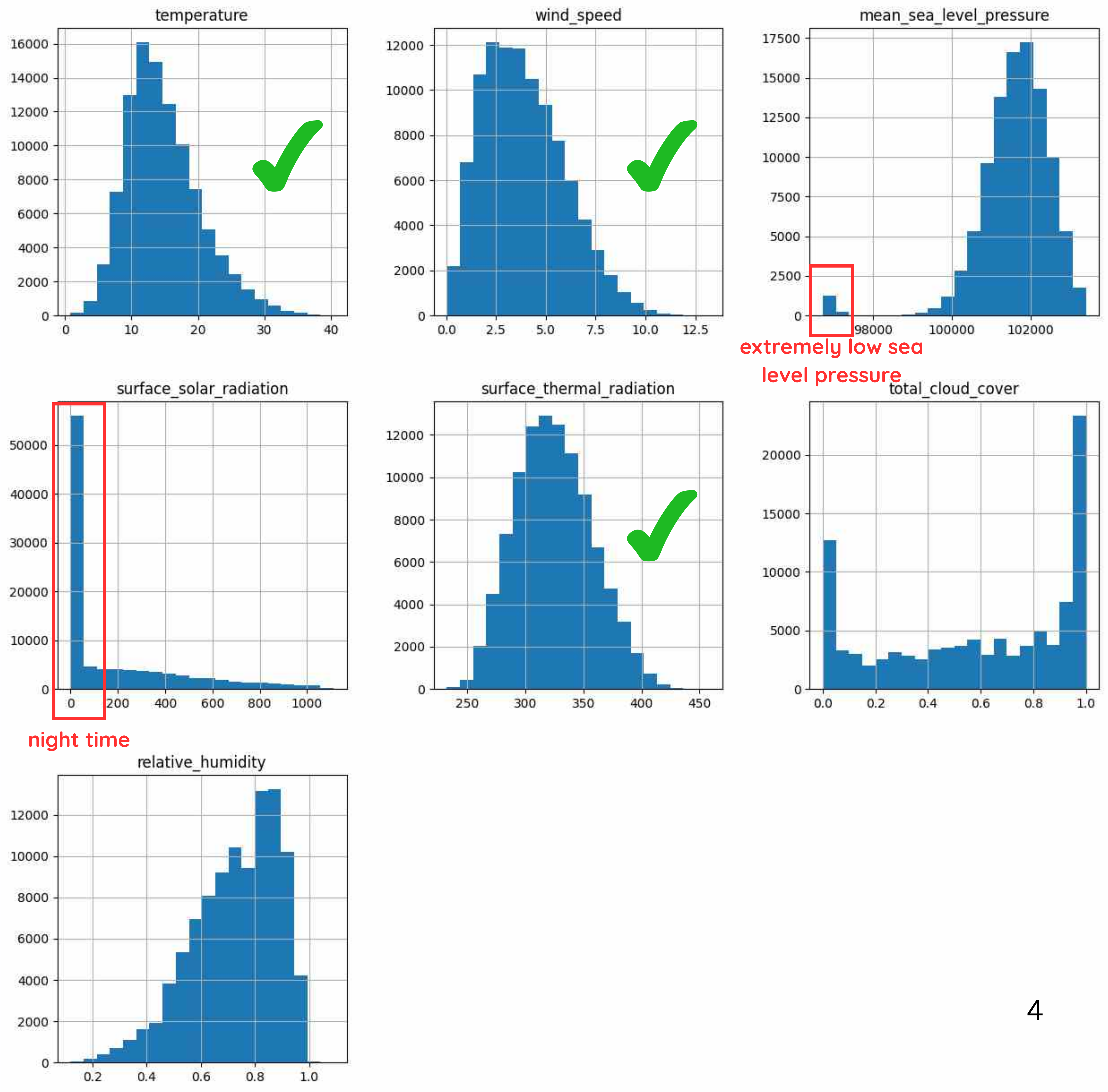
- mean_sea_level_pressure:
- surface_solar_radiation:

Data Cleaning

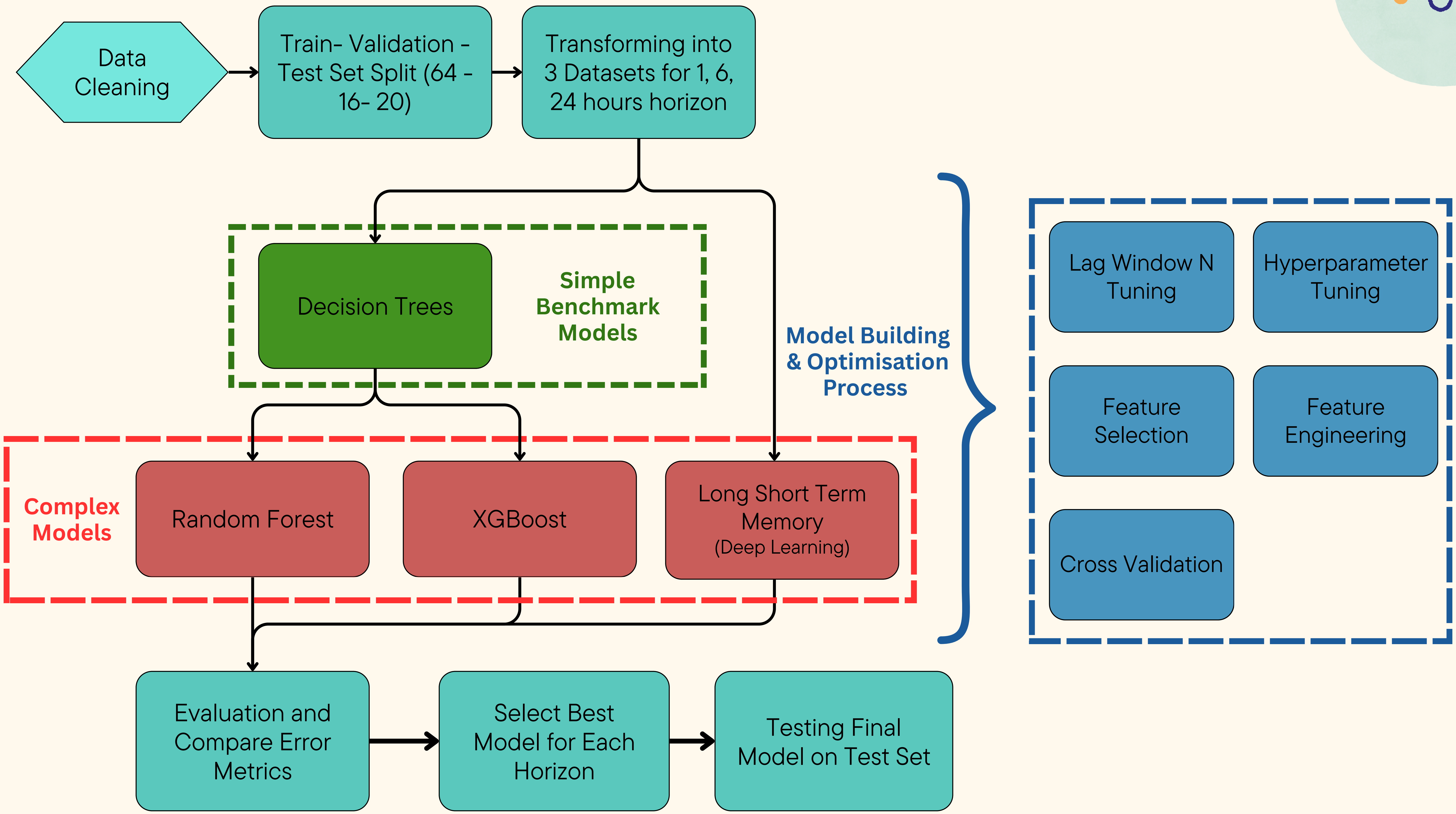
- relative_humidty
  - converted values > 1 to 1

Feature Enginnering

- Engineered temporal features like hour/ week of the day and rolling averages of certain continuous variables



METHODOLOGY



TRAIN-VALIDATION-TEST SPLIT

100k+ Observational Data

64% Training Set	**16% Validation Set**	**20% Test Set**
Used to train the models	Used to evaluate intermediate models and tune hyperparameters	Used only to evaluate the final model

When slicing the dataset

- Test set chronologically after training set (to avoid lookahead bias)
- Data points not shuffled to preserve the chronological order

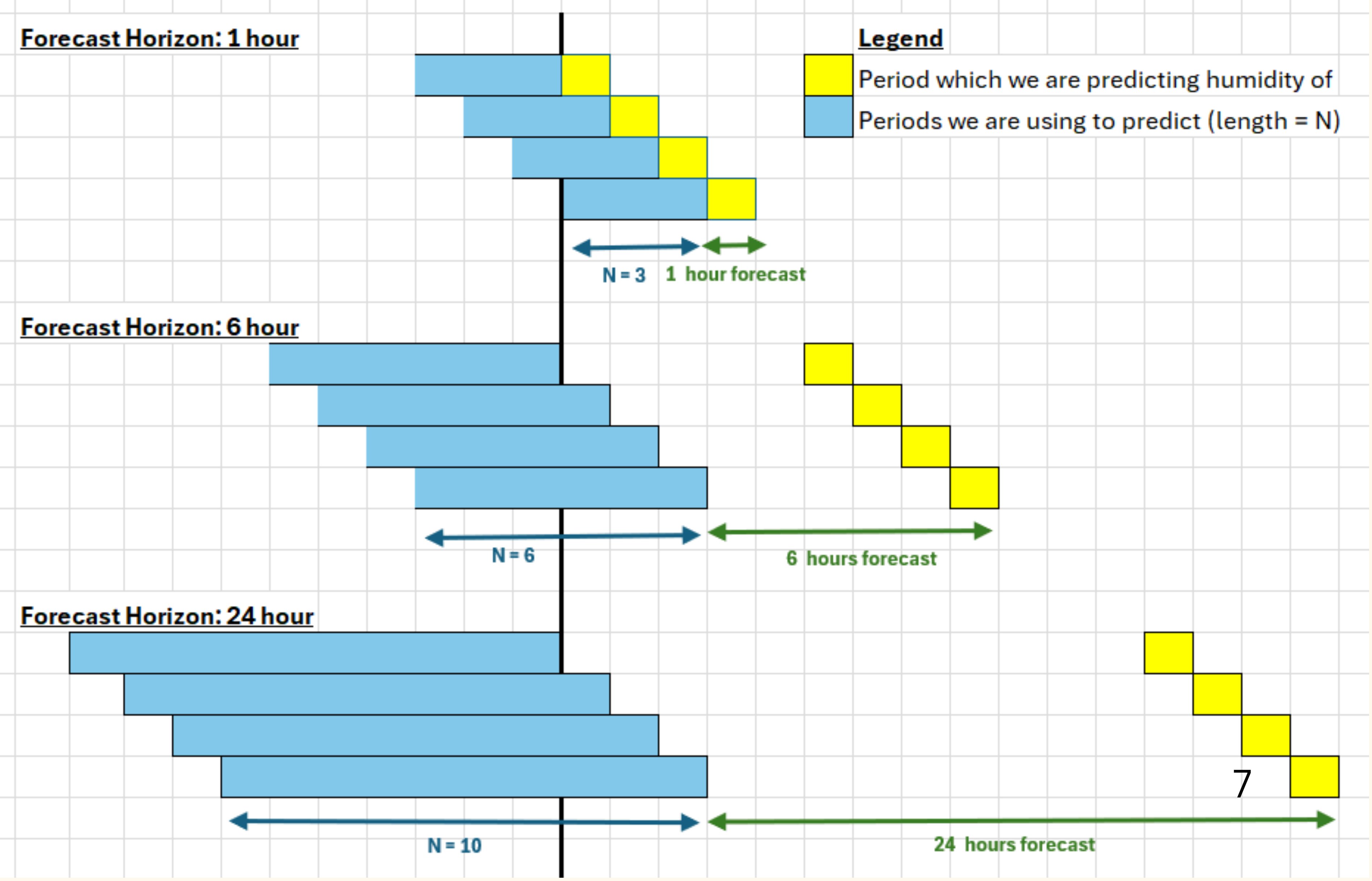
Error Metrics

- MAE/MSE
  - measures prediction error
- R^2
  - measures how well the model fits

DATA TRANSFORMATION

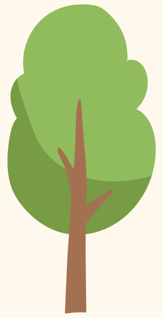
original dataset						lagged values								
temp	wind speed	sea level pressure	surface solar radiation	surface thermal radiation	total cloud cover	temp lag 1	temp lag 2	....	temp lag N	wind speed lag 1	wind speed lag 2	...	wind speed lag N	lagged values of other variables
t1	w1	...	...	...	...									...
t2	w2	...	...	...	...	t1				w1				...
t3	w3	...	...	...	...	t2	t1			w2	w1			...
t4	w4	...	...	...	...	t3	t2	...		w3	w2	...		...
t5	w5	...	...	...	...	t4	t3	...		w4	w3	...		...
t6	w6	...	...	...	...	t5	t4	...		w5	w4	...		...
t7	w7	...	...	...	...	t6	t5	...		w6	w5	...		...
t8	w8	...	...	...	...	t7	t6	...	t1	w7	w6	...	w1	...
t9	w9	...	...	...	...	t8	t7	...	t2	w8	w7	...	w2	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
t99	ws99	...	...	...	...	t98	t97	...	t92	w98	w97	...	w92	...
t100	ws100	...	...	...	...	t99	t98	...	t93	w99	w98	...	w93	...

Removed these rows due to NA values



DECISION TREE AND RANDOM FOREST – WORKFLOW

DT1

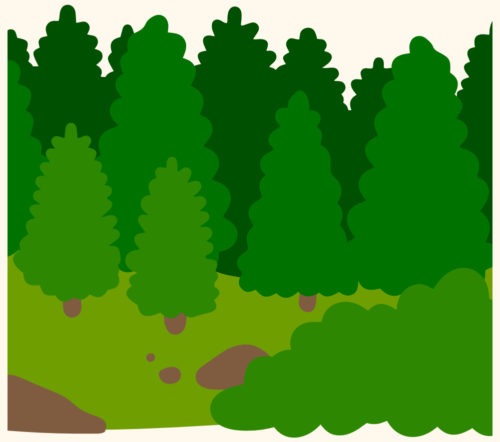


Baseline Decision Tree
(without hyperparameter tuning or cross-validation)

DT2

Tuned Decision Tree
(with hyperparameter tuning and cross-validation)

RF1



Random Forest Model
(with feature engineering, hyperparameter tuning and cross-validation)

RF2

Optimised Random Forest Model
(with feature selection)

RANDOM FOREST

What did we do differently?

Manual Lag N Tuning

- tailored the lag_N_candidates tested for each forecast horizon
- selected optimal lag_N through manual evaluation on a set of ***criteria***

1h	1, 3, 6, 12
6h	1, 3, 6, 12
24h	1, 3, 6, 12, 24, 48

Considerations

- MAE / MSE
- Model interpretability
- Timeliness of signal
- Pattern sensitivity

Feature Engineering

- Created **time-aware features**
  - hour-of-day and day/night indicators
- **Rolling averages**
  - smooths out short-term fluctuations

=> better capture temporal patterns in weather data

Feature Pruning

- **Cumulative importance:**
  - total contribution to model variance
- **Permutation importance**
  - drop in model's predictive performance when values for each feature are shuffled

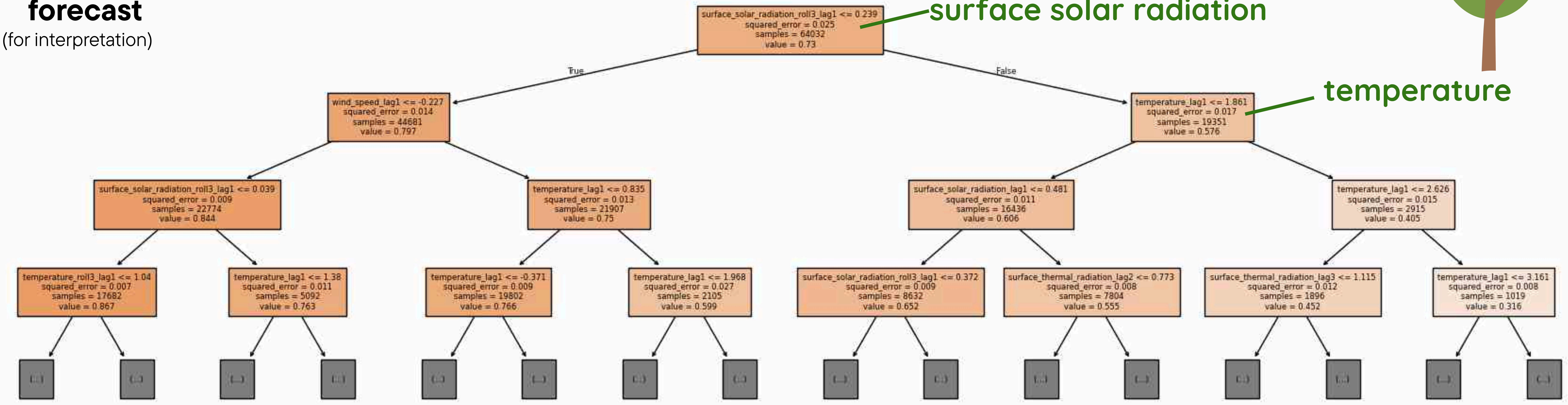
=> dimensionality reduction for a leaner model without significantly impacting model performance

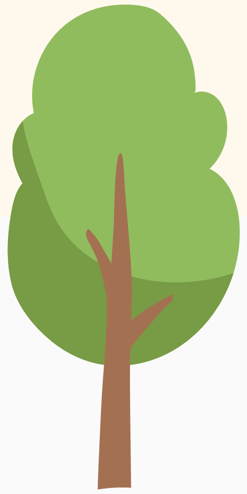
BASELINE DECISION TREE (DT1)

(without hyperparameter tuning or cross-validation)

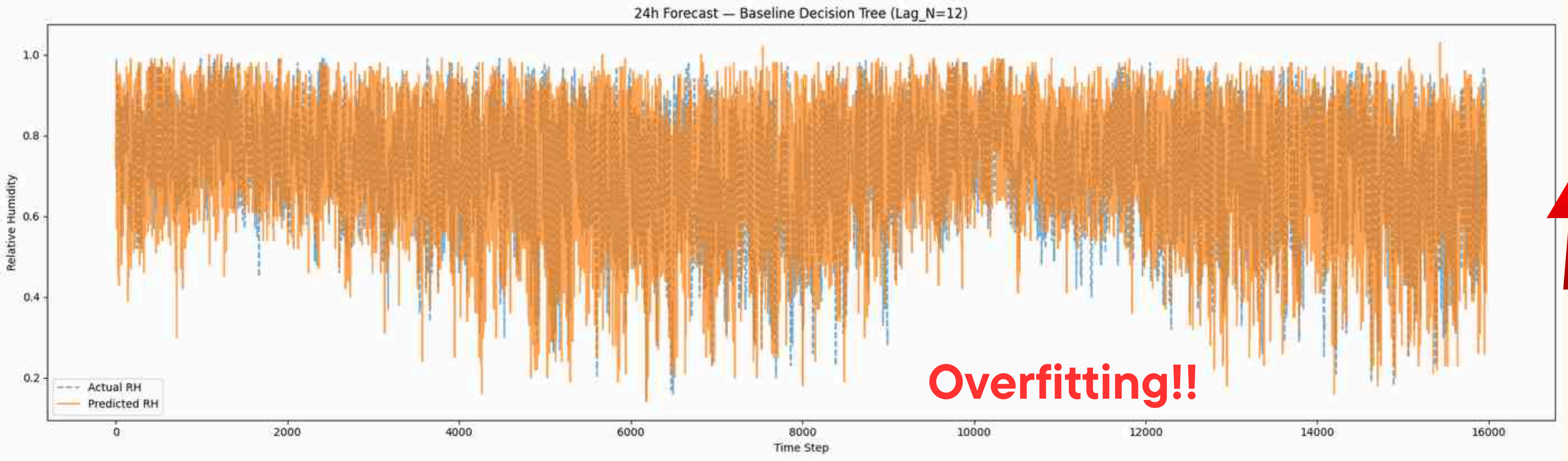
1h forecast

(for interpretation)





24h forecast



	MAE	MSE	R2
1h	0.071686	0.008725	0.669367
6h	0.095113	0.016291	0.382541
24h	0.115351	0.023220	0.119378

TUNED DECISION TREE (DT2)

(with hyperparameter tuning and cross-validation)

GridSearchCV to find optimal hyperparameters

results

	Best_Params	MAE	MSE	R2
1h	{'max_depth': 7, 'min_samples_leaf': 4, 'min_samples_split': 2}	0.064205	0.006901	0.738476
6h	{'max_depth': 7, 'min_samples_leaf': 1, 'min_samples_split': 2}	0.081929	0.011515	0.563536
24h	{'max_depth': 5, 'min_samples_leaf': 4, 'min_samples_split': 2}	0.089531	0.013569	0.485412

Improvements AFTER TUNING!

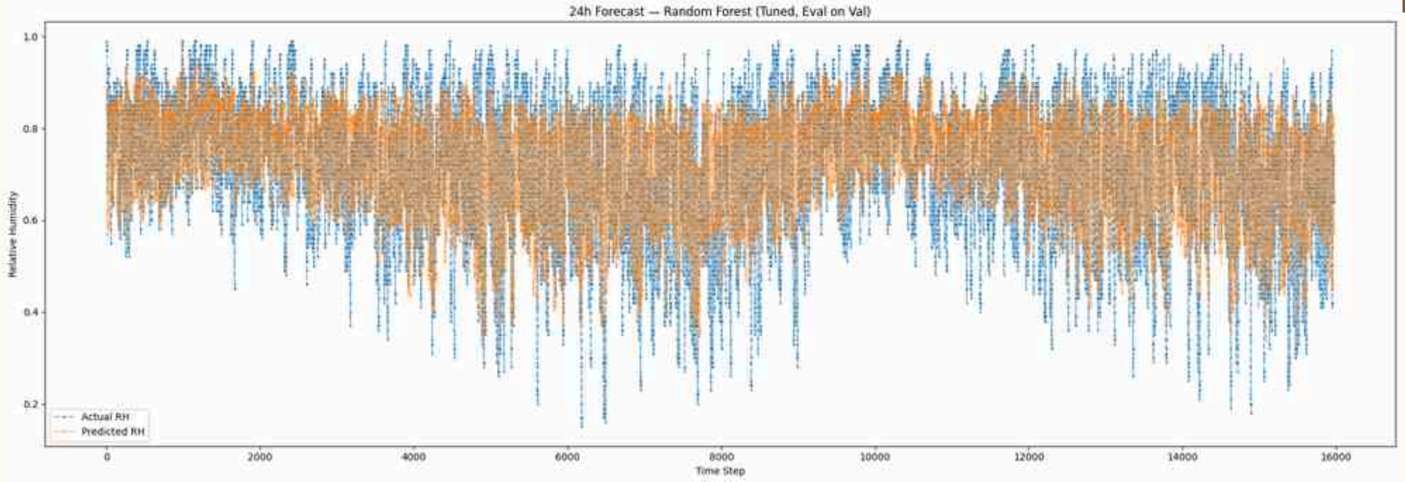
Forecast Horizon	MAE (%)	MSE (%)	R^2 (%)
1h	-10.43	-20.88	+10.33
6h	-13.27	-29.30	+47.30
24h	-22.42	-41.54	+306.59

Findings

- Hyperparameter tuning substantially improved performance across all horizons.
- The 24h model, initially underperforming, now shows viable forecasting potential.

RANDOM FOREST MODEL (RF1)

(with feature engineering, hyperparameter tuning and cross-validation)



Improvements (compared to DT 1)

Forecast Horizon	MAE (%)	MSE (%)	R^2 (%)
1h	-22.61	-40.49	+19.99
6h	-25.49	-46.99	+75.85
24h	-26.03	-46.62	+344.66

Overall interpretation:

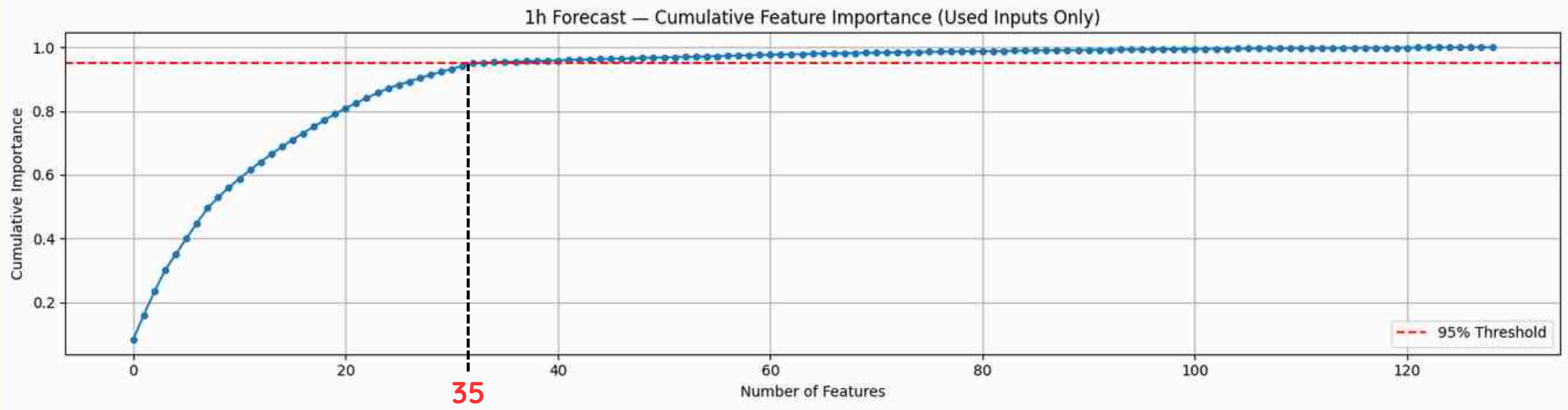
- Random Forest outperforms both decision tree models, especially at 6h and 24h forecasts.
- Ensemble learning via bagging improves robustness, reduces variance, and avoids overfitting seen in deeper trees.

OPTIMISED RANDOM FOREST MODEL (RF2)

(with hyperparameter tuning, cross-validation + Feature Selection/Pruning)

1 Cumulative Importance

(using raw RF importance values)



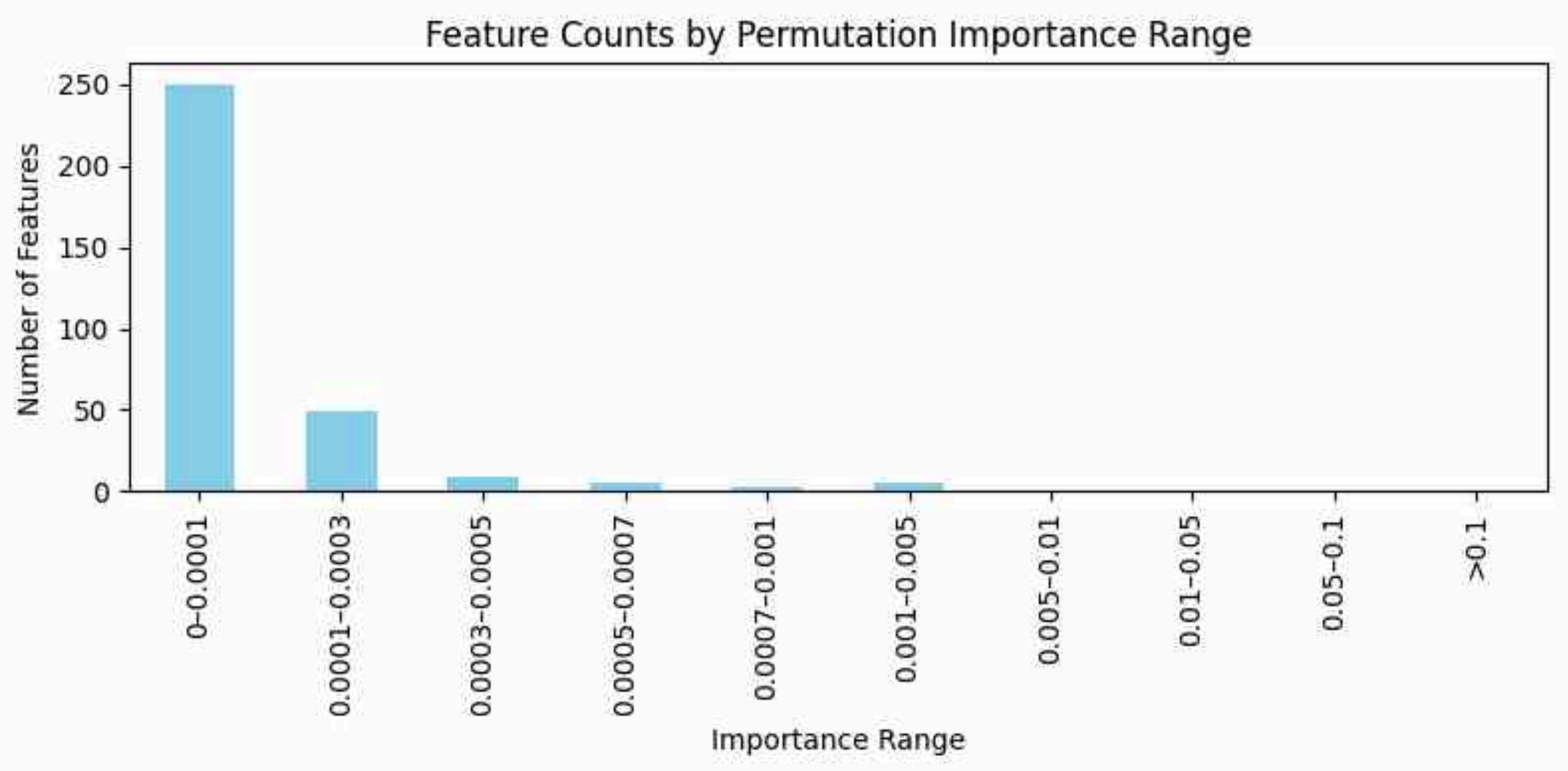
Time Horizon	No. of features that covers 95% total Gini importance
1	35
6	130
24	300

Horizon-specific thresholds

Lower no. of features cover 95% of total importance

=> higher threshold selected

2 Permutation Importance

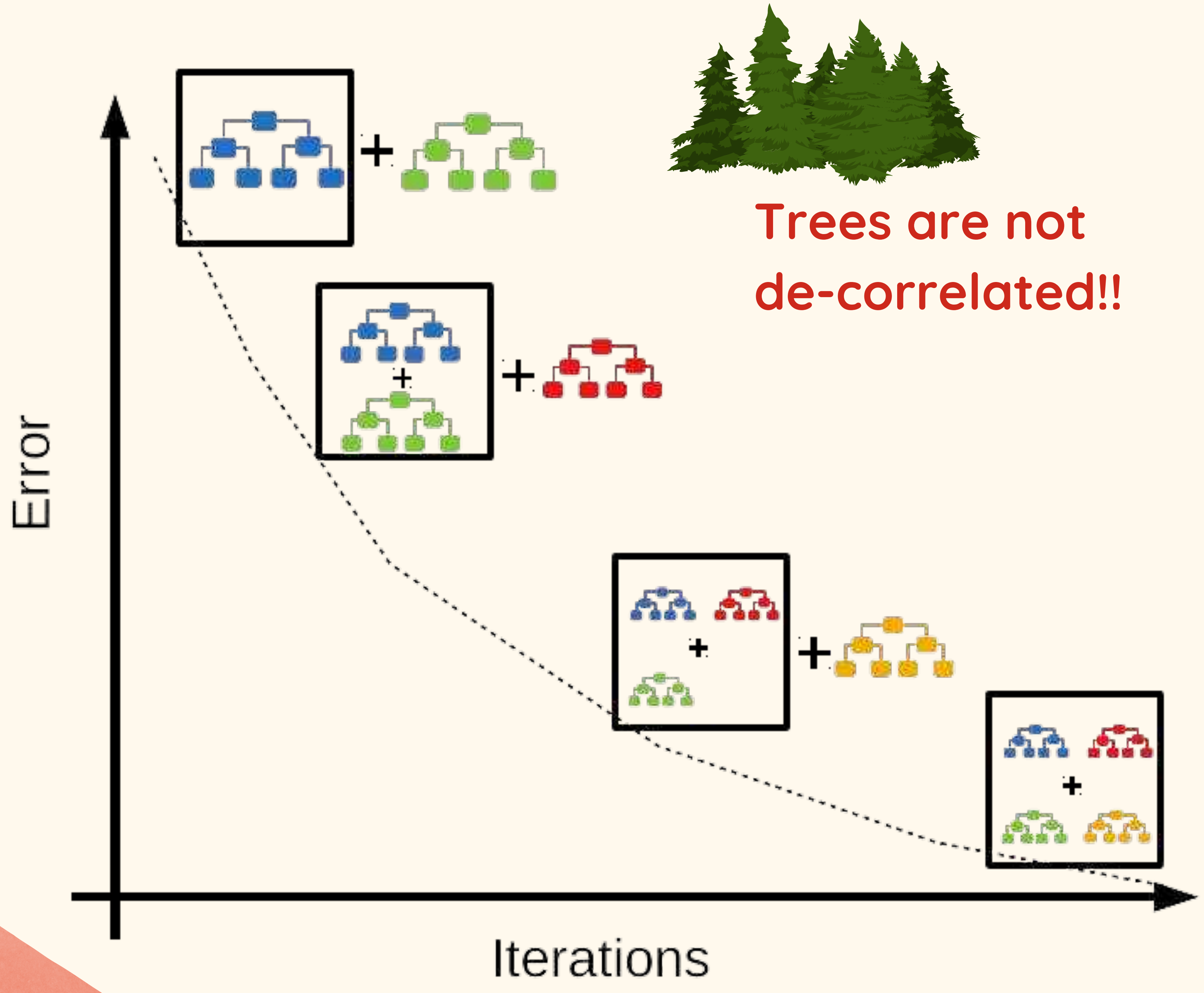


Dropping all features with permutation importance < 0.0001

✓ Eliminates over 75% of low-impact features

✓ Focuses the model on high-signal inputs

EXTREME GRADIENT BOOSTING XGBOOST



Trees are not de-correlated!!

How it works: Sequential Learning

Additive Tree boosting: Fixing previous tree's errors	Loss function & Gradient Descent: Minimum Absolute Error used

Generalizes better

Shrinkage: Small learning rate	Regularization: Penalizes complexity

Process

1. Build upon RF - similar input dataframes and features
2. Tune optimal lag window for each forecast horizon
3. Hyperparameters Tuning using Optuna
4. Features pruning test

GRADIENT BOOSTING - XGBOOST

Tune Lagging Window N

Pretty different from the previous!

Time Horizon	Optimal Lag Window N
1	Lag_24
6	Lag_72
24	Lag_24

Still have to input the lagged features, one for each horizon!

>> lag_3 previously!

- Leverage pattern from one day ago

>>>>> lag_6 previously!

- Harder to predict due to day-night transitions
- 3 full daily cycles helps XGB learn

== lag_12 previously!

- Captures 1 full diurnal lag
- XGB can learn from repeating daily pattern

"Leverage patterns across time without overfitting!"

XGBOOST

Tuning Hyperparameters

~~GridSearch CV with TimeSeriesSplit (5 fold) across 50-100 combinations per model~~

Optuna (Bayesian optimization) for 10 trials across 3 folds across a search range

Few hours → Few minutes

Objective: Minimizing MAE - better interpretability

Gamma (Penalization)

- Higher for 1h and 24h
- Relied on short-term or periodic trends more

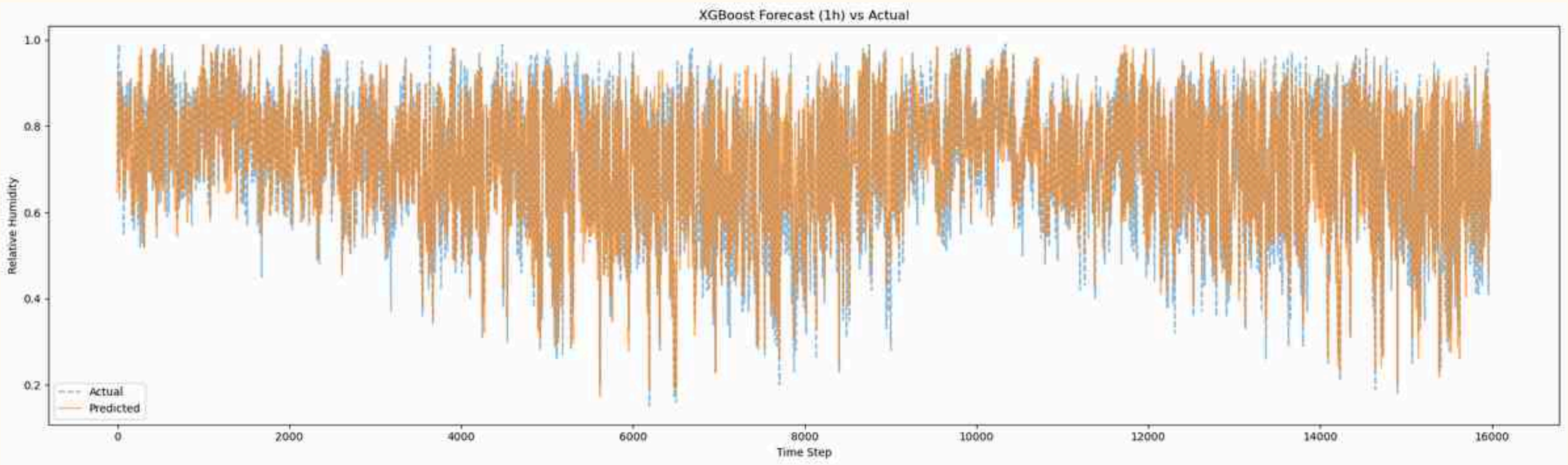
Learning rates (Shrinkage)

- Lower for longer horizons
- Pickup harder pattterns

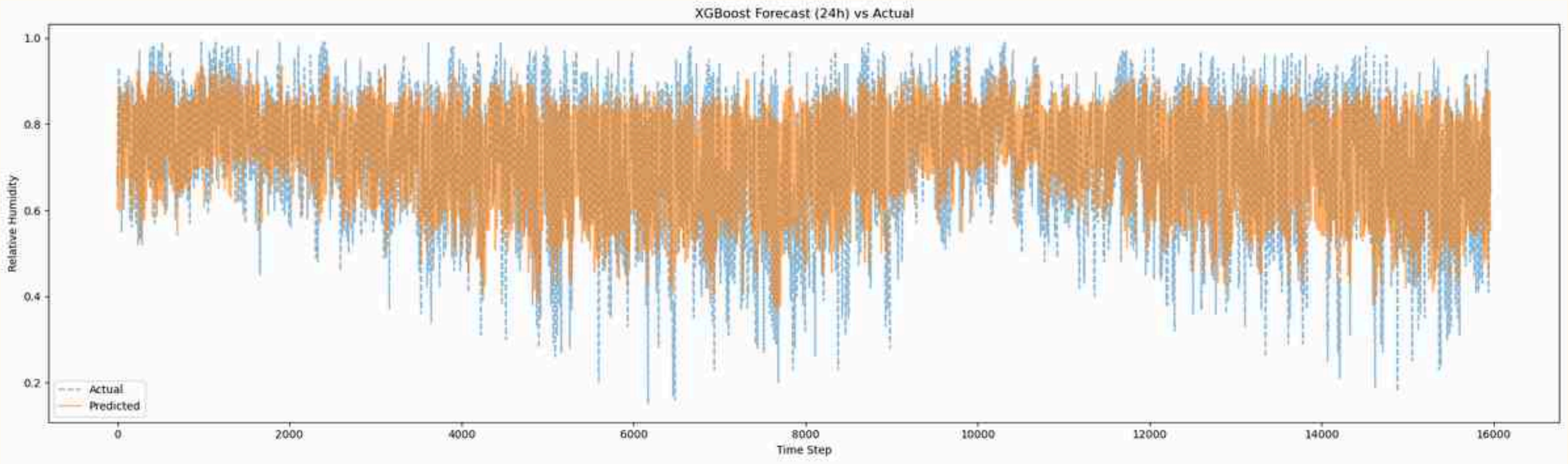
Pretty solid results as expected!
However, R^2 for 24h poorer than RF2

Performance Fit

1h



24h



Time Horizon	MAE	MSE	R^2	R^2 for RF2!	R^2 for DT2!
1	0.0504	0.0044	0.8331	0.8122	0.7385
6	0.0659	0.0078	0.7050	0.6483	0.5635
24	0.0851	0.0131	0.5028	0.5279	0.4854

XGBOOST

Feature Importance and Pruning

95% Cumulative Importance used to prune features

Barely improved performance

Performance of Pruned Model vs Full Model

Time Horizon	Dropped Features	R^2 (Full model)	R^2 (Pruned model)	MAE (Full model)	MAE (Pruned Model)
1	91	0.8365	0.8357	0.0500	0.0502
6	247	0.7050	0.7043	0.0657	0.0658
24	72	0.5028	0.5204	0.0839	0.0840

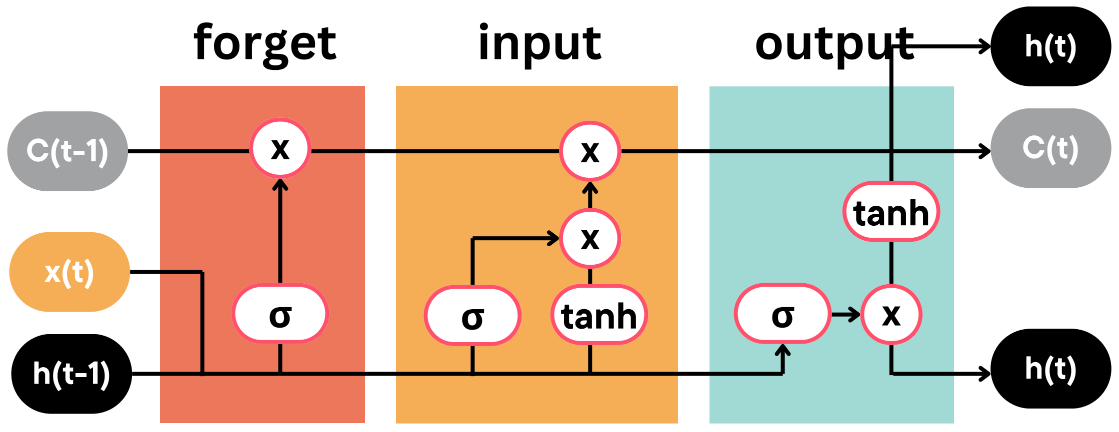
MAE Increased!

- Features with low importance ARE STILL IMPORTANT!!
- XGB benefits from feature redundancy, making pruning not effective

LONG-SHORT TERM MEMORY

LSTM1

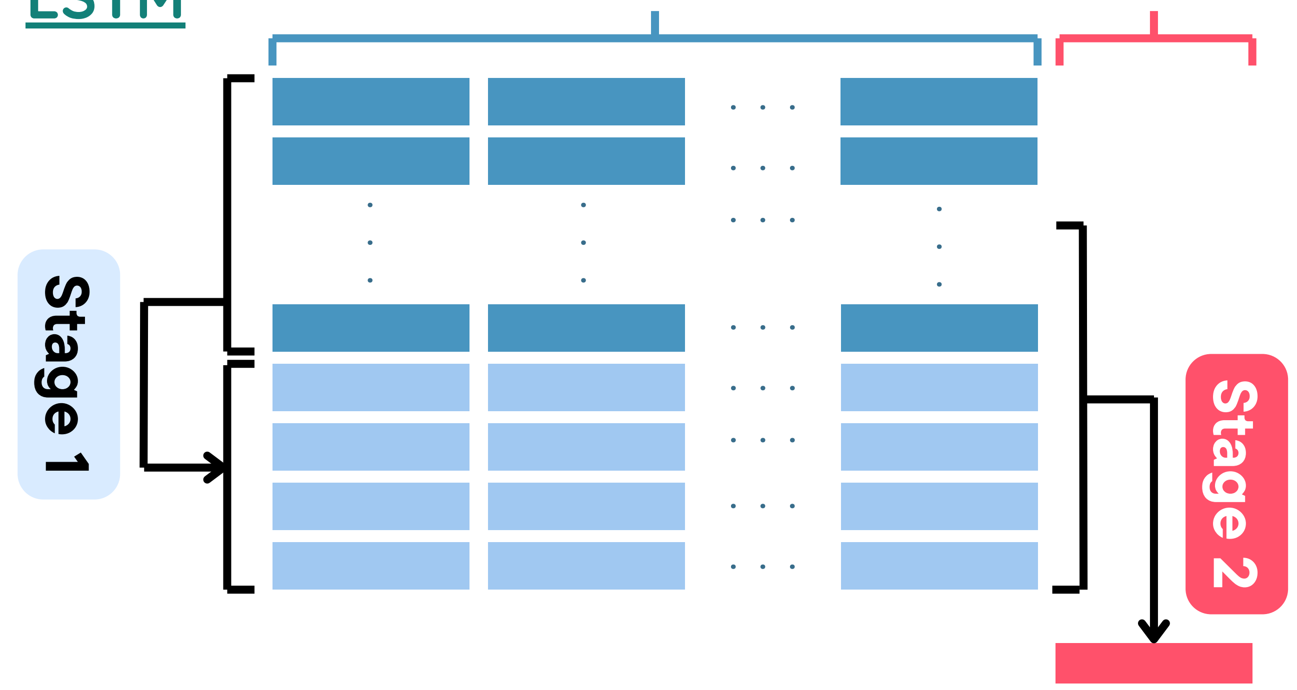
LSTM Memory Cell



Improved RNNs
Avoids vanishing/ exploding gradient
Regulates flow of long/ short term information.

LSTM2

Multi-step LSTM



Hyper parameter tuning

Manually perform random search through

lag	6, 24, 48, 72
sampling_rate	1, 2, 4
batch_size	32, 64, 128
neurons_per_layer	32, 64
hidden_layers	2, 3, 4

tried 20 random combinations

Feature Engineering

Yearly seasonality

$$\texttt{sin\_time\_8766} = \sin\left(\frac{2\pi \times timestamp_{hours}}{8766}\right)$$
$$\texttt{cos\_time\_8766} = \cos\left(\frac{2\pi \times timestamp_{hours}}{8766}\right)$$

Daily seasonality

$$\texttt{sin\_time\_24} = \sin\left(\frac{2\pi \times timestamp_{hours}}{24}\right)$$
$$\texttt{cos\_time\_24} = \cos\left(\frac{2\pi \times timestamp_{hours}}{24}\right)$$

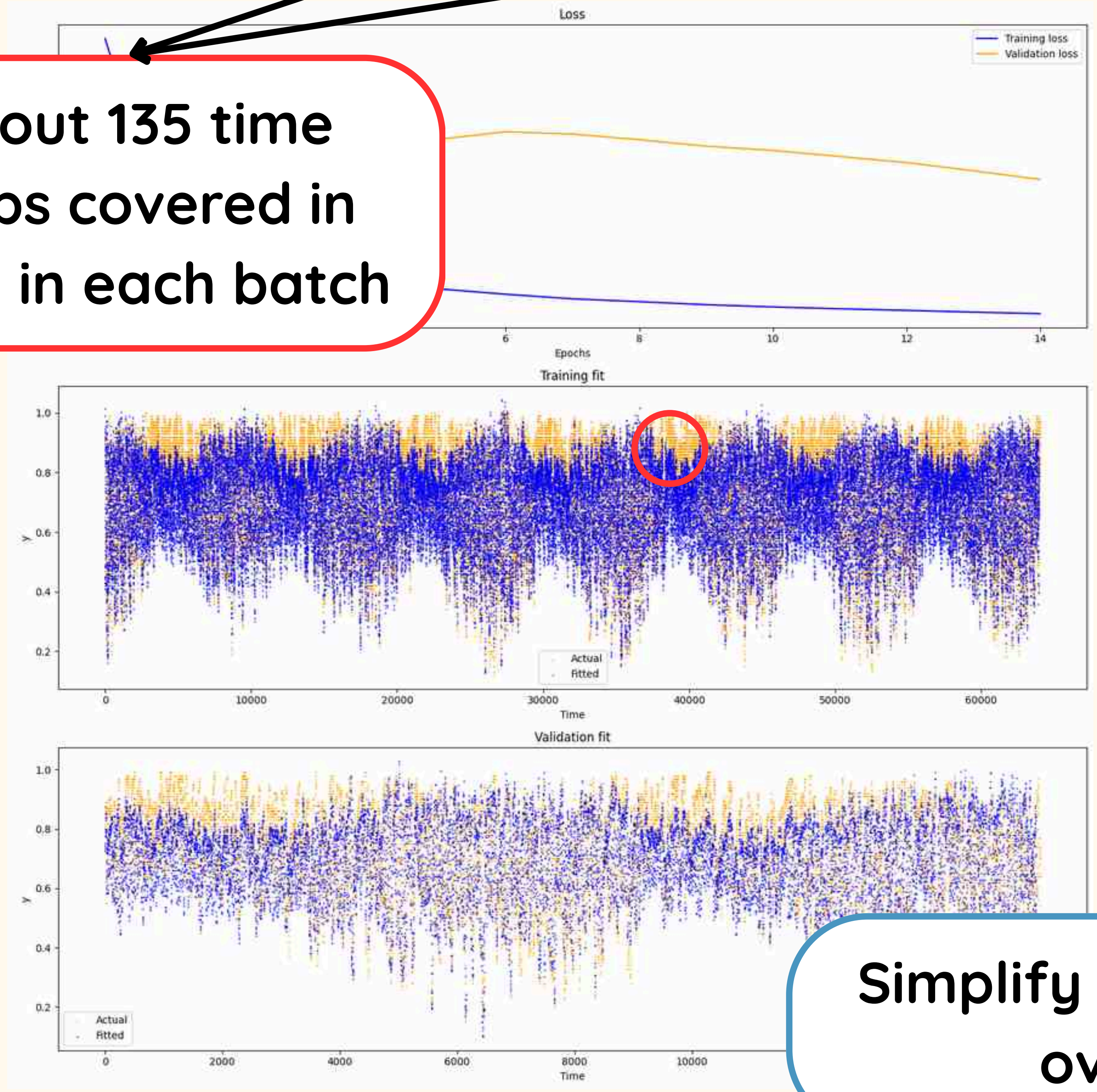
LSTM3

LSTM1: basic LSTM
LSTM2: multi-step LSTM
LSTM3: LSTM with added features

LONG SHORT TERM MEMORY HYPERPARAMETER TUNING

Before tuning (1 hour):

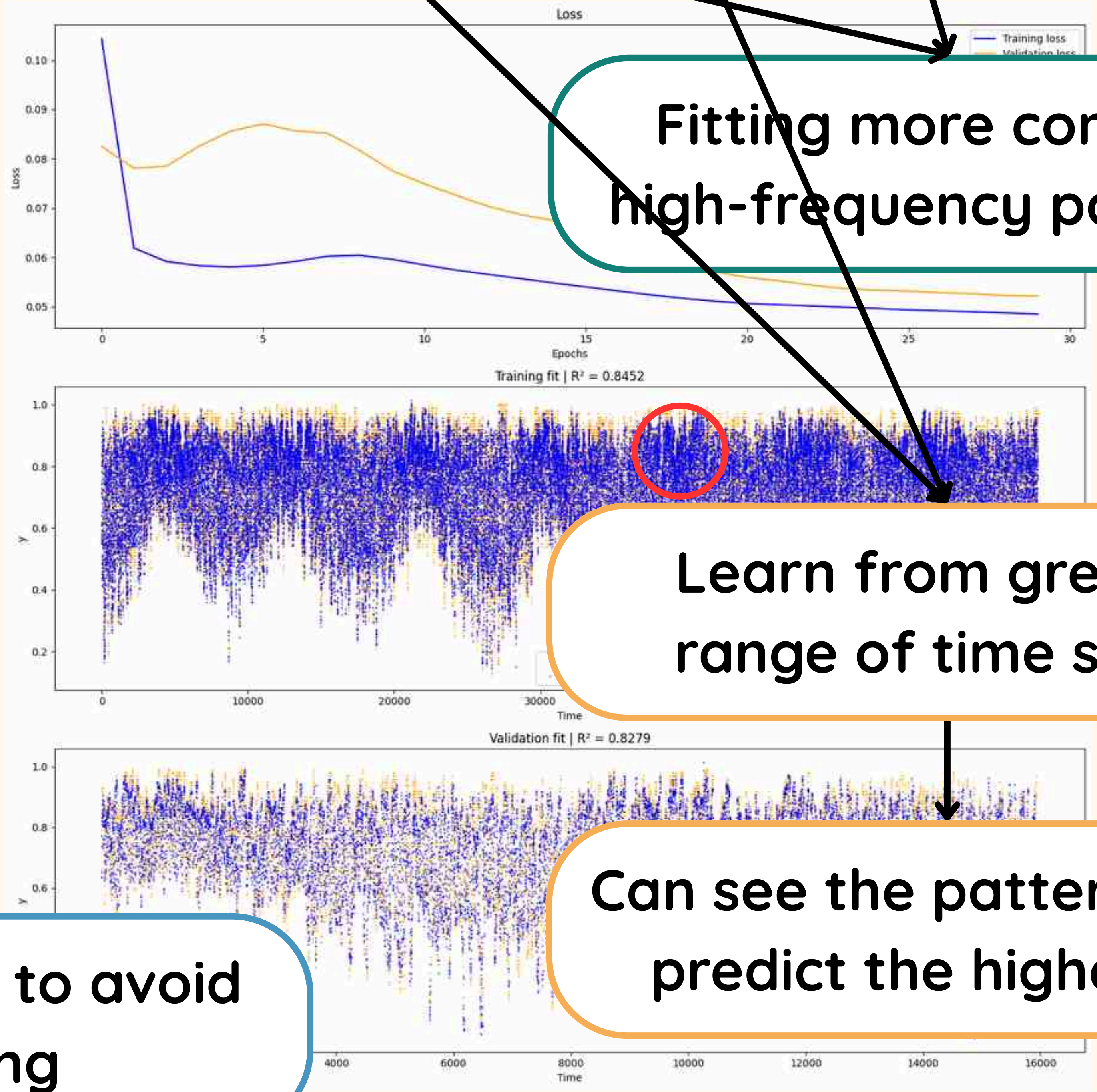
Settings: lag=72, sr=2, bs=64, npl=24, hl=3



about 135 time steps covered in total in each batch

After tuning (1 hour):

Settings: lag=48, sr=1, bs=128, npl=32, hl=2



Fitting more complex high-frequency patterns

Learn from greater range of time steps

Can see the patterns that predict the higher RH

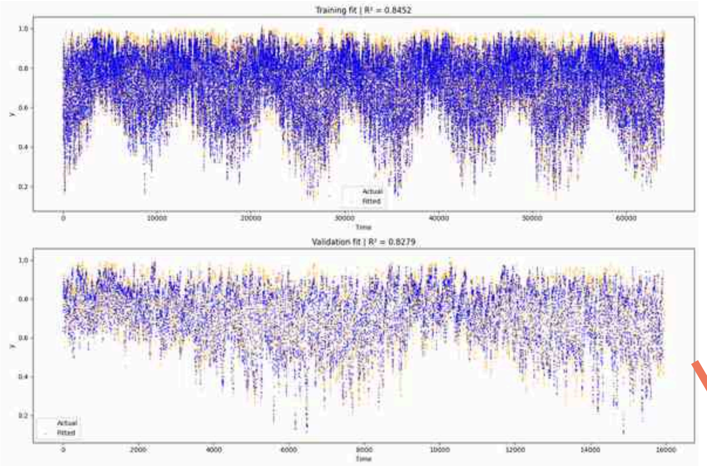
Simplify model to avoid overfitting

LSTM1: basic LSTM
LSTM2: multi-step LSTM
LSTM3: LSTM with added features

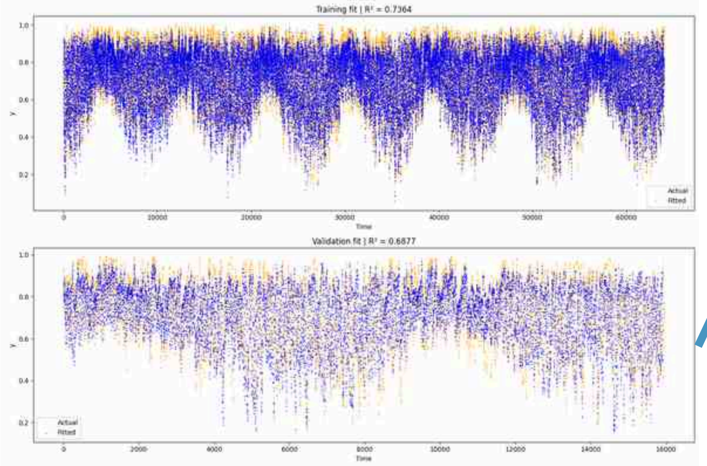
LONG SHORT TERM MEMORY SIMPLE AND MULTI-STAGE

Over the time horizons

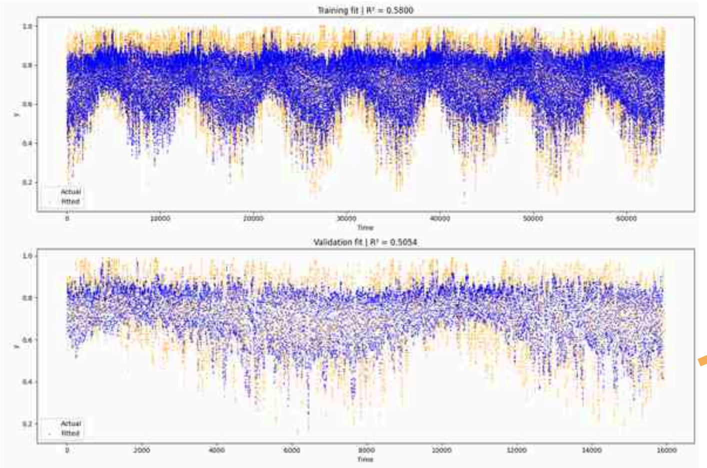
1 hour
R2 = 0.8279



6 hours
R2: 0.6877

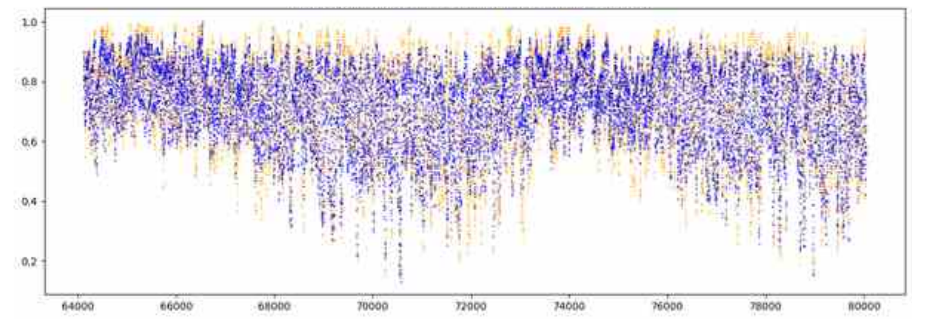


24 hours
R2 = 0.5002



Multi-stage LSTM

Validation fit for 6 hours (R2 = 0.7168)

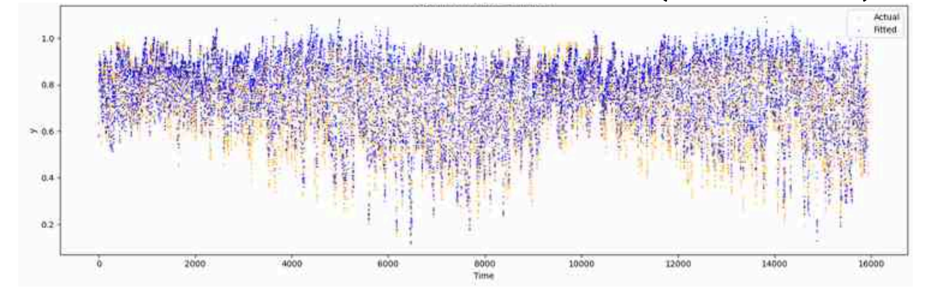


Multi-stage/multi-step LSTM:

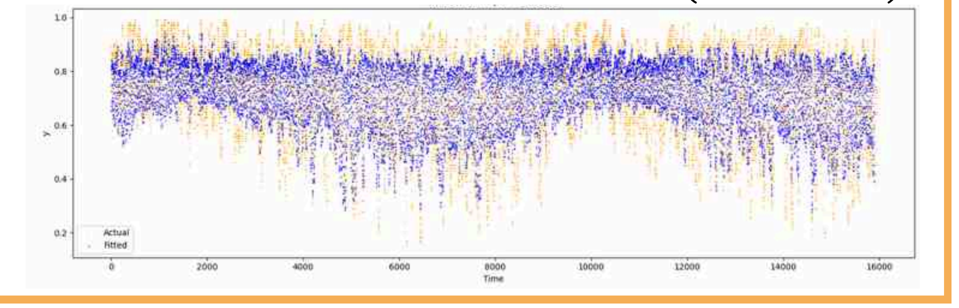
- accuracy of 1-hour predictions
- autocorrelative properties of weather variables

Added features

Validation fit for 1 hour with MSE (R2 = 0.6973)



Validation fit for 24 hours with MSE (R2 = 0.5193)



From shorter time horizon:

- Seasonality amplified
- More variable and extreme (many RH > 1)

For longer time horizon:

- Predicting more extreme values helps

- decreasing quality of prediction with respect to time (based on R2)
- less variable predictions (based on graphs)

LONG SHORT TERM MEMORY

A simpler model works best for the 1 hour time horizon

LSTM1: basic LSTM
LSTM2: multi-step LSTM
LSTM3: LSTM with added features

Seasonality would be overly influential

	1h			6h			24h		
	LSTM1	LSTM2	LSTM3	LSTM1	LSTM2	LSTM3	LSTM1	LSTM2	LSTM3
MAE	0.0521		0.0543	0.0685	0.0663	0.073	0.0854	0.0909	0.0857
R^2 (MAE)	0.8279		0.8184	0.6877	0.7145	0.6362	0.5002	0.4613	0.5048
MSE	0.0048		0.008	0.0089	[redacted]	0.0104	0.013	[redacted]	0.0127
R^2 (MSE)	0.8186		0.6973	0.6609	[redacted]	0.605	0.5054	[redacted]	0.5193

Intermediate predictions benefit the 6-hour time horizon

High total error of intermediate predictions

Daily and yearly seasonality become important for longer time horizons

CONCLUSION

Time Horizon	Model	R^2
1	XGB	0.8387
6	LSTM2	0.7145
24	RF2	0.5306

All different..

Accuracy
XGB,
Tuned RF

Runtime efficiency
XGB,
DT

Interpretability
XGB,
DT

Deployable

XGB Chosen!

Robustness

Time Horizon	Model	R^2
1	XGB	0.8387
6	XGB	0.7023
24	XGB	0.5280

On the test set!

Future Improvements

- Ensemble models
- Interactive features
- Possible external variables

Societal Impacts

- Support time-sensitive or low-resource weather forecasting applications
- Assist meteorologists and automate routine analysis