# Predicting MDA8 Ozone using a Deep-Learning Model

# Outline

1. Previous work by Tailong He
2. Current work and results
3. Remaining questions

# Deep Learning

- Large “deep” neural networks
- Automatic feature extraction/learning
    - Learning hierarchical representations
- Can utilise large data and compute

- *“Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction.”* - Yann LeCun

# Recurrent U-net: Deep learning to predict daily summertime ozone in the United States - Tai-Long He et al

- Hybrid model (Conv-LSTM) to predict summertime (JJA) daily maximum 8-h average (MDA8) surface ozone concentrations in the US

- Meteorological fields from ERA-Interim reanalysis + monthly mean NOx emissions from CEDS

- Ozone measurements from EPA

- Years 1980-2014. 30 summers to train model (1980 - 2010)

# Model schematic

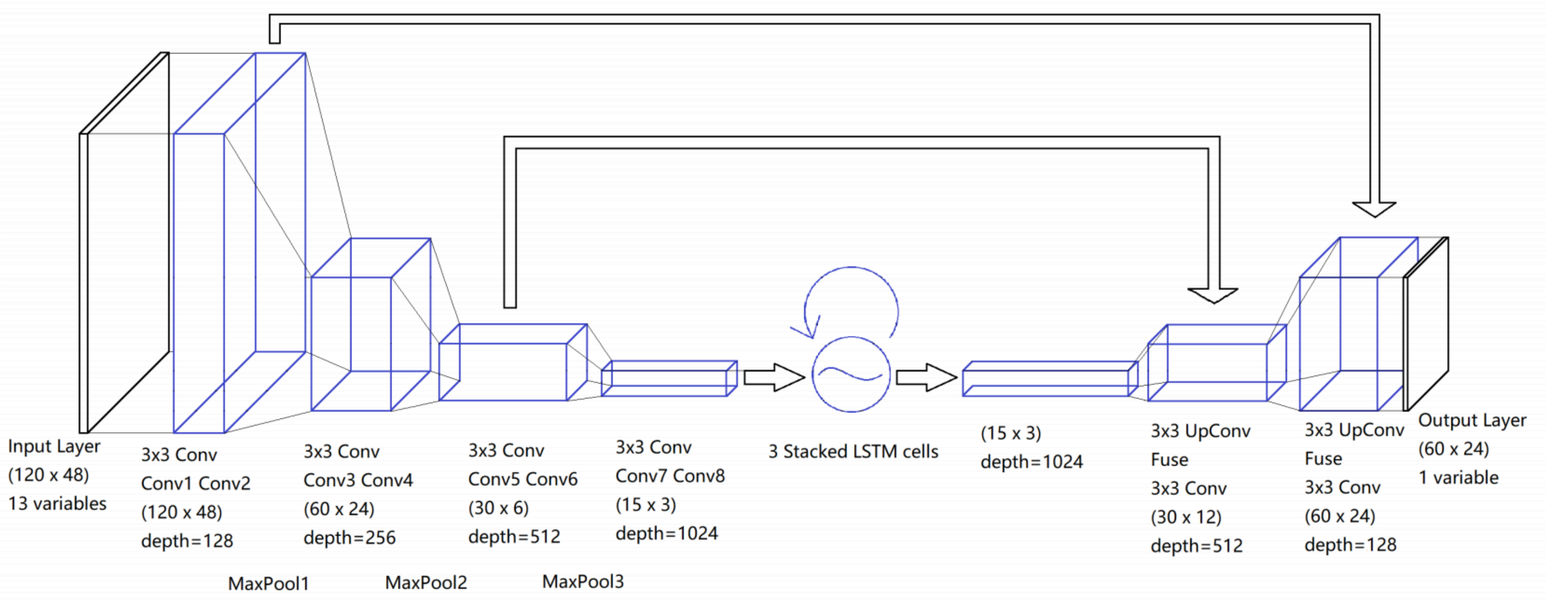



Figure 1: Deep learning model to predict JJA MDA8. The model consists of an input layer with 13 channels for the ozone predictors, 8 convolution and 3 max pooling layers to extract the dominant features in the data, and 3 stacked LSTM cells to capture the dynamics in the data. Compressed data are then passed to transposed convolution layers for projection to the output layer. The two arrows at the top indicate the skip connections that forward the high-resolution features extracted by the encoder to the decoder for better localization of the features.

# Predictors

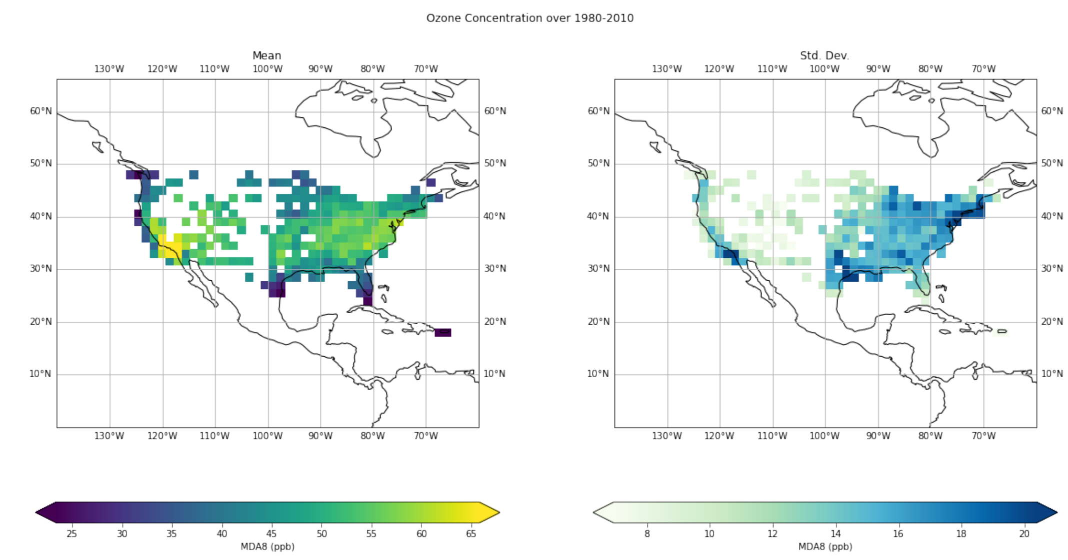
**Meteorological fields**

Regridded at 1.5˚ x 1.5˚

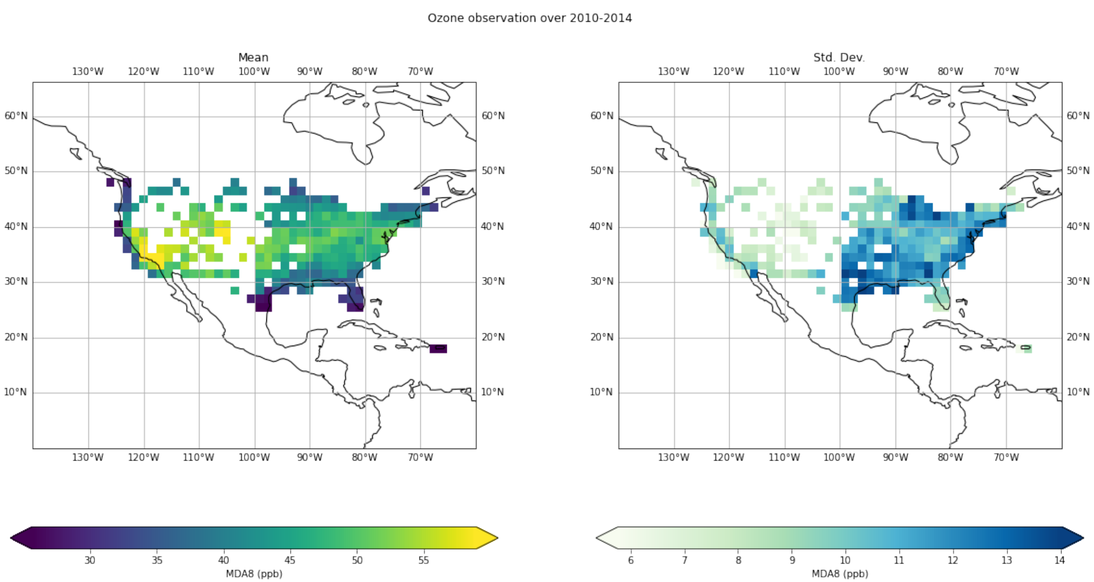
1. Geopotential at 500 hPa (Z)
2. Dewpoint at 2m (d2m)
3. Mean sea level pressure (mslp)
4. Temperature at 2m (t2m)
5. Shortwave radiation downwards (ssrd)
6. Sea surface temperature (sst)

**NOx emission sectors**

7. agriculture (AGR)
8. power industry (ENE)
9. manufacturing (IND)
10. residential and commercial (RCO)
11. international shipping (SHP)
12. surface transportation (TRA)
13. waste disposal (WST)



Ozone Concentration over 1980-2010
Mean
Std. Dev.
MDA8 (ppb)
MDA8 (ppb)

Ozone observation over 2010-2014
Mean
Std. Dev.
MDA8 (ppb)
MDA8 (ppb)

# Results

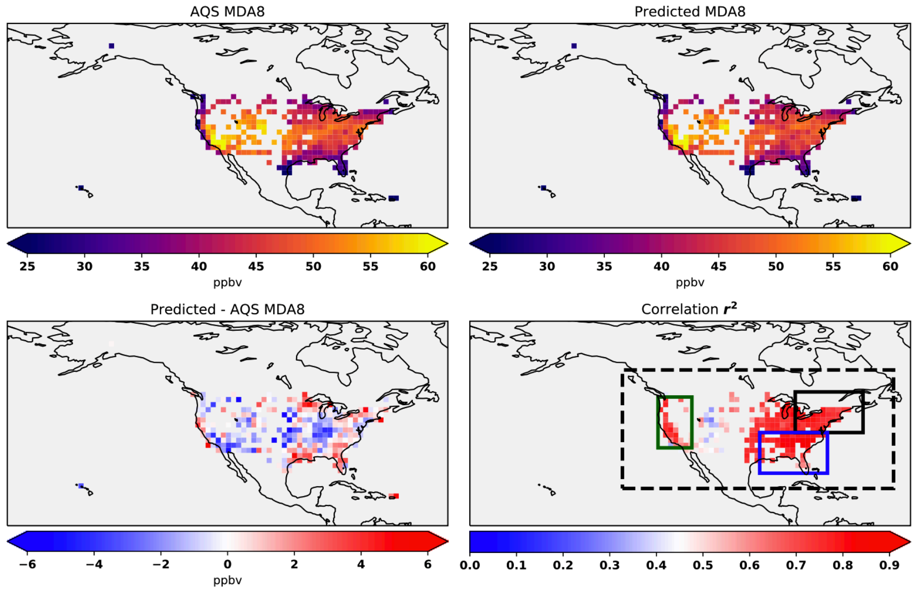



Figure 1. Observed (top left) and predicted (top right) mean JJA MDA8 ozone during 2010– 2014. Also shown (bottom left) is the absolute error (in ppb) for the predicted minus observed MDA8 ozone. The errors are calculated where the AQS observations are located. Correlation (r2) between the observed and predicted MDA8 ozone in each grid box is shown in bottom right. Also shown in bottom right are the definitions of the CONUS, Northeastern US, Southeastern US and the West coast domains in blacked dashed box, black solid box, blue box, and green box, respectively.

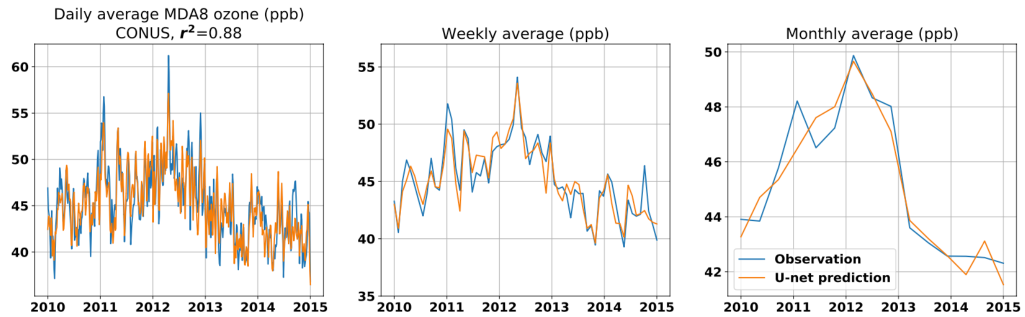



Figure 2. Observed (blue line) and predicted (orange line) daily (first column), 7-day averaged (second column), and monthly averaged (third column) JJA MDA8 ozone (in ppb) during the testing period (2010–2014). Shown are the time series for the CONUS (first row) …

**Tailong's GRL paper**

# Feature Visualisation

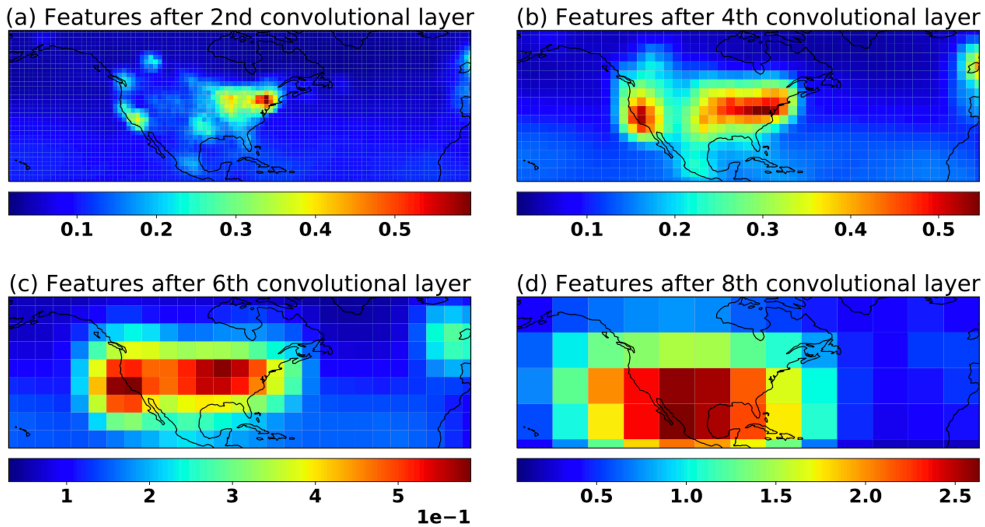



Figure 9: Feature maps from the 2nd, 4th, 6th, and 8th convolutional layers as illustrated in Fig. 4, averaged within all training data from 1980 to 2009.

**Figure 9**
**from Tai-long's preprint, Appendix D**

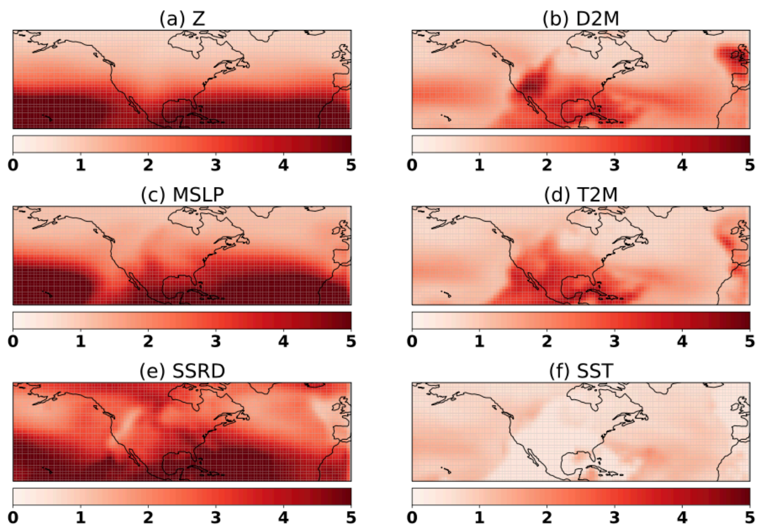



Figure 10: Averaged feature maps for the meteorological predictors, calculated from the training data set.

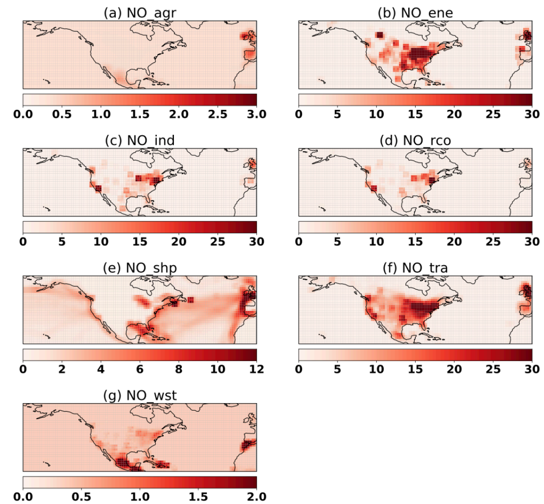



Figure 11: Average activation maps for NOx emissions, calculated from training data set. The seven NOx emission sectors are as follows: (a) agriculture (AGR), (b) the power industry (ENE), (c) manufacturing (IND), (d) residential and commercial (RCO), (e) international shipping (SHP), (f) surface transportation (TRA), and (g) waste disposal (WST).

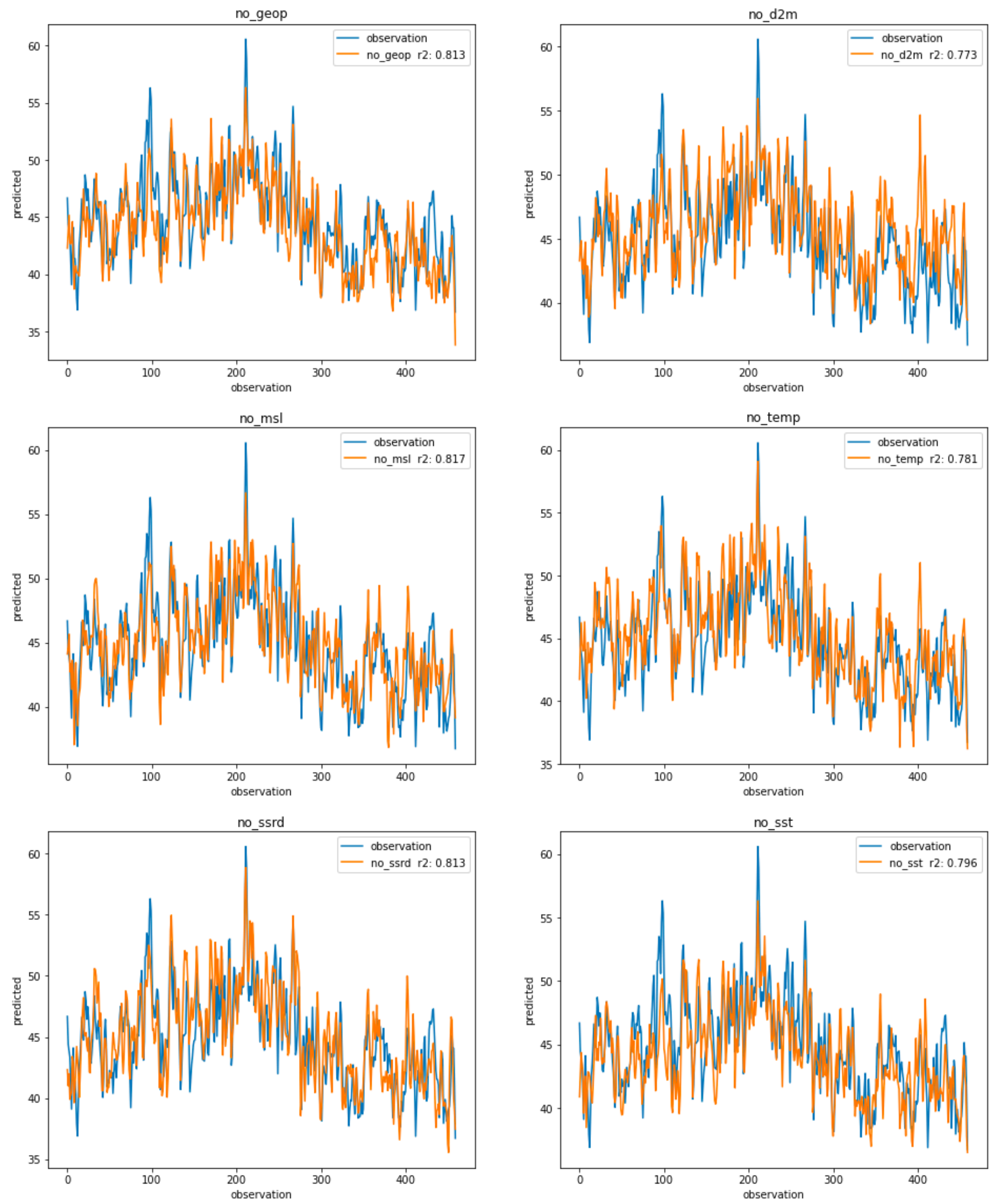
**Figures 10 and 11**
**from Tai-long's preprint, Appendix D**

# Questions

- Why is the model learning well?
- What gives the model its temporal skill?
- Region-dependence

# Training the model with different configurations

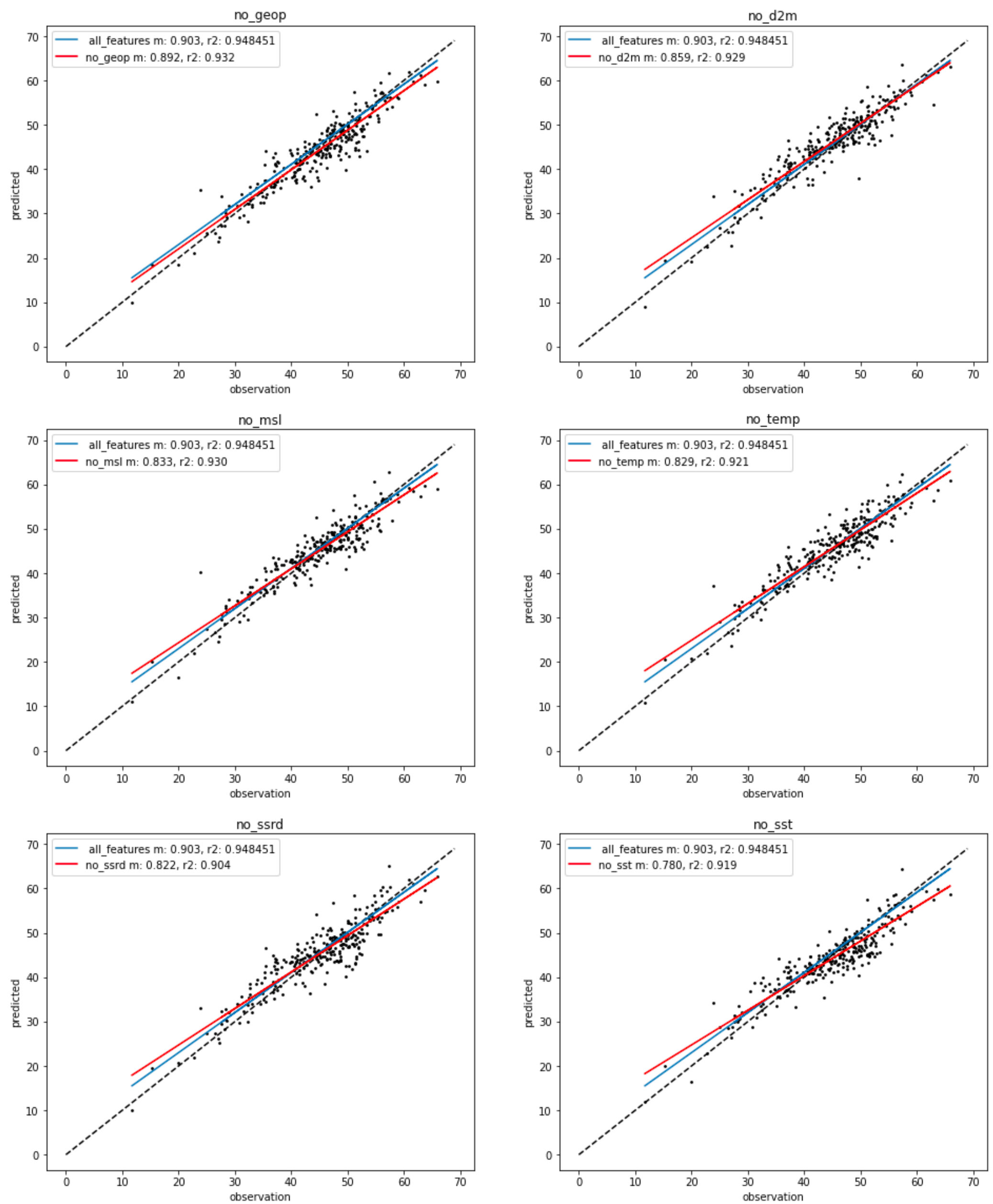
Training the model with all predictors but withholding one meteorological field




**Should be correlation coefficient
**r not r2**

# Training the model with different configurations

Training the model with all predictors but withholding one meteorological field




**should be correlation coefficient r
**not r2**

# Differences

- all_features model = model with all predictors
- featureless model = model without that meteorological field

Where do these two models differ?

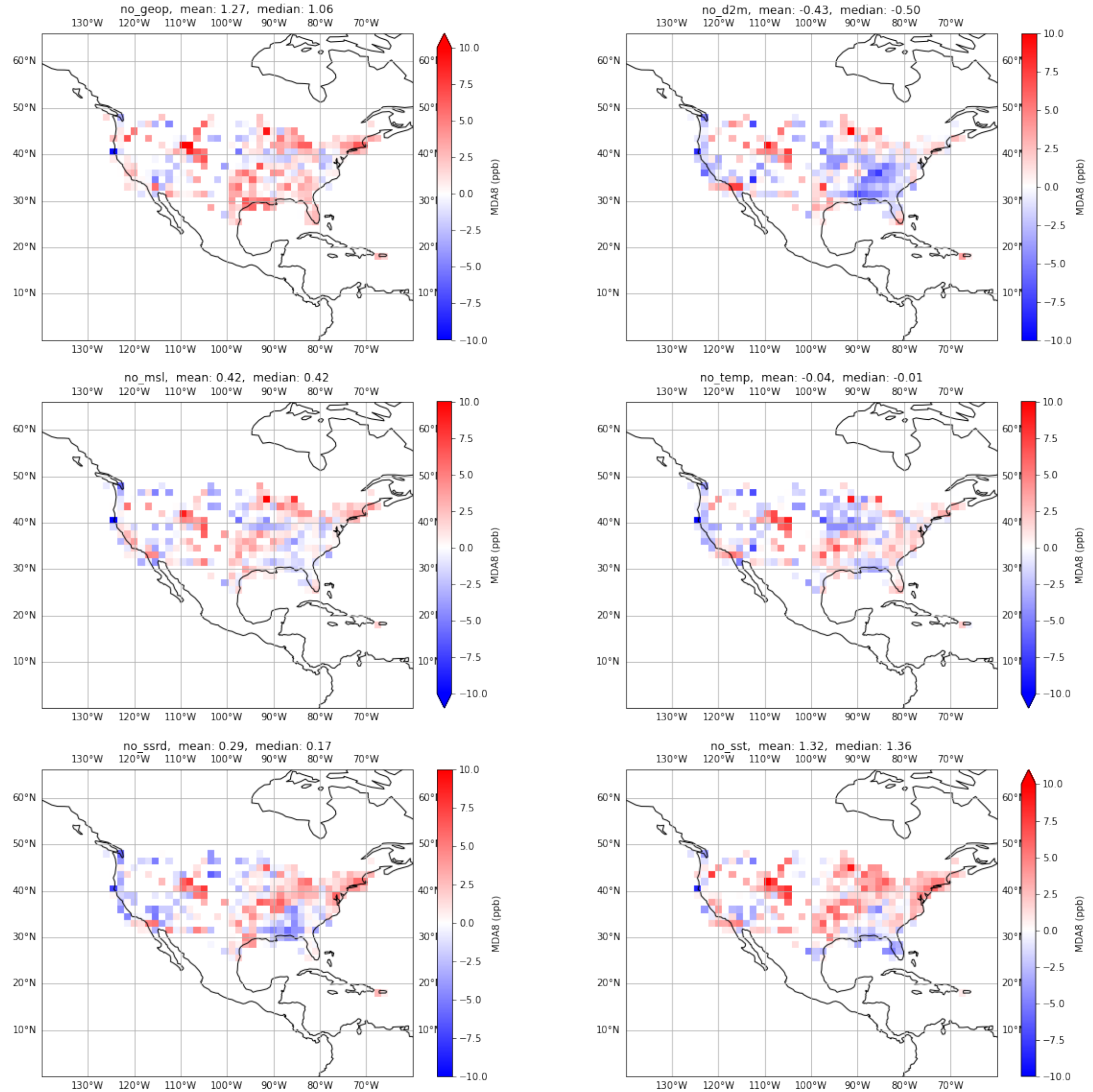
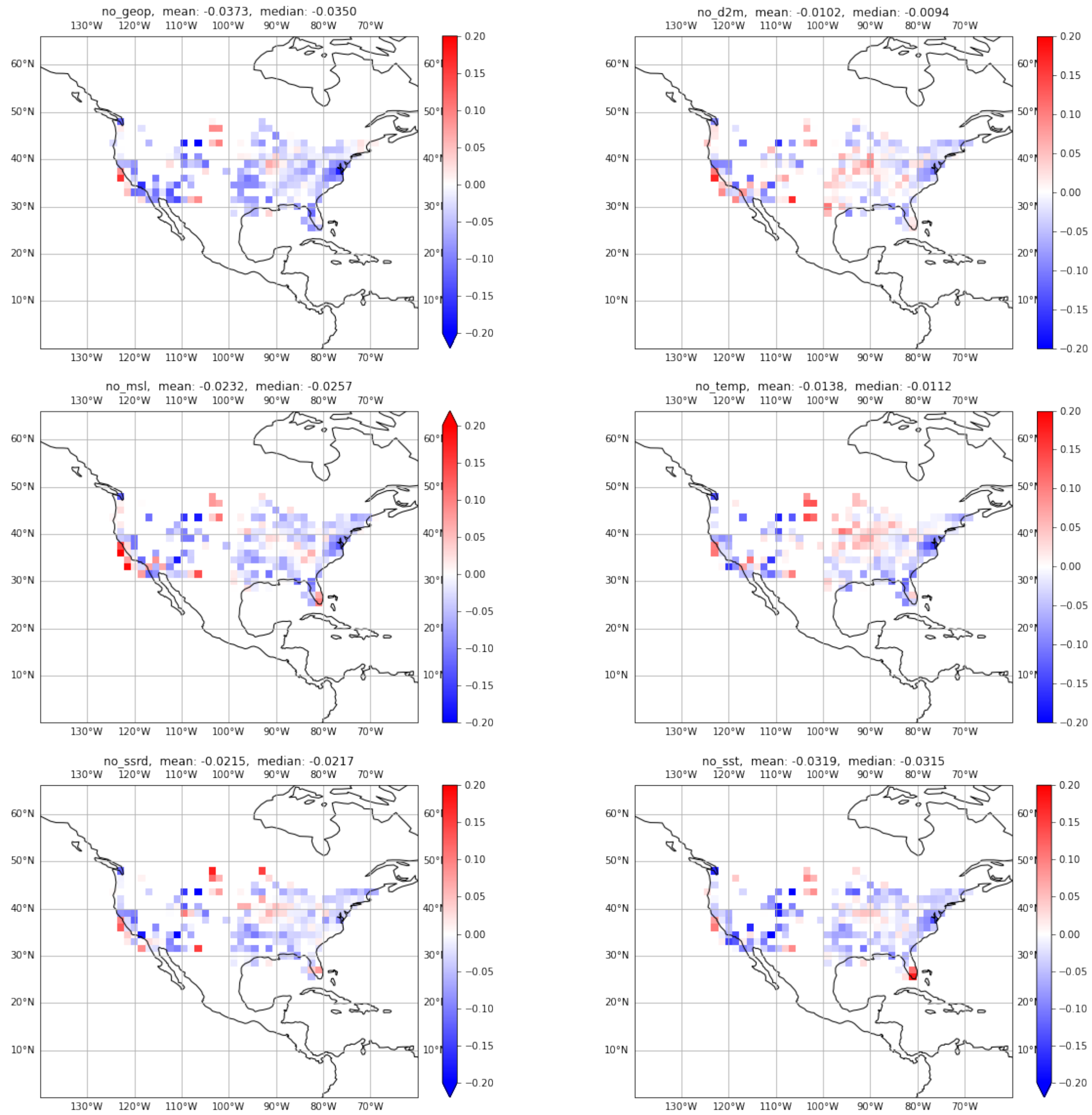



Figure shows the absolute differences between the all_features model predictions and no_met models predictions

# Difference in Correlation

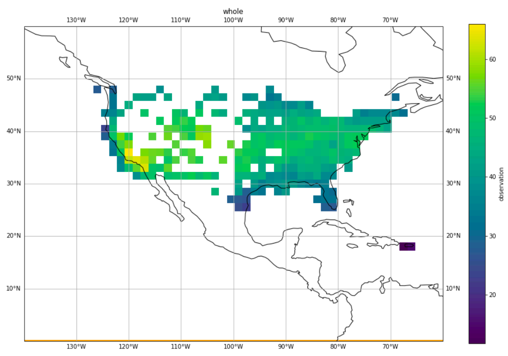
- all_features model = model with all predictors
- featureless model = model without that meteorological field
- all_features_corr - featureless_corr

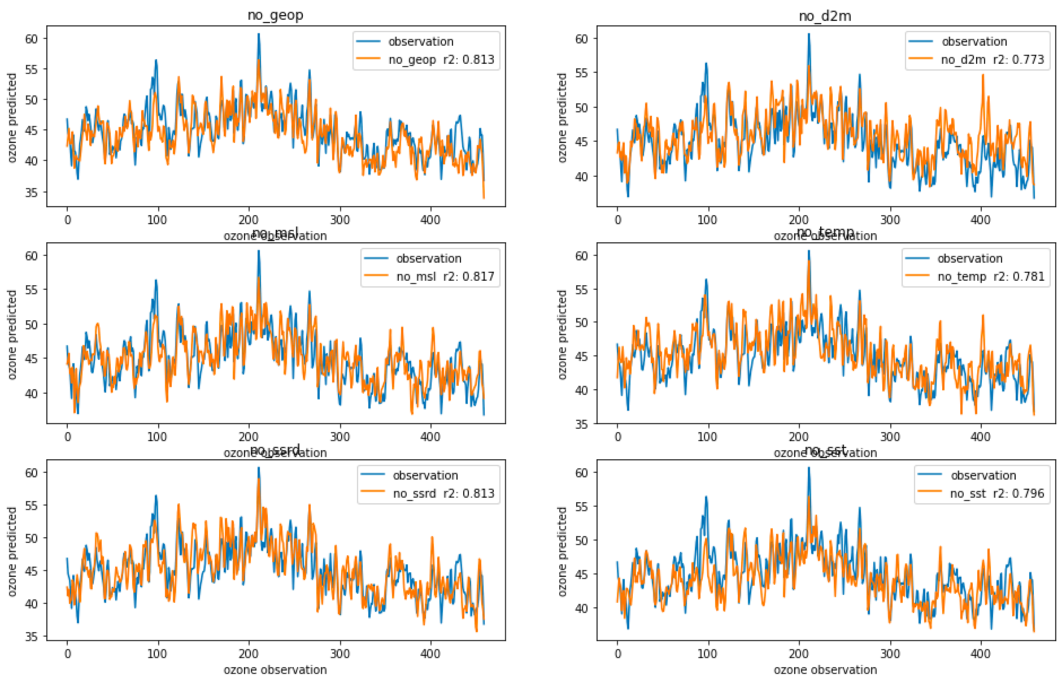
Where do these two models differ?

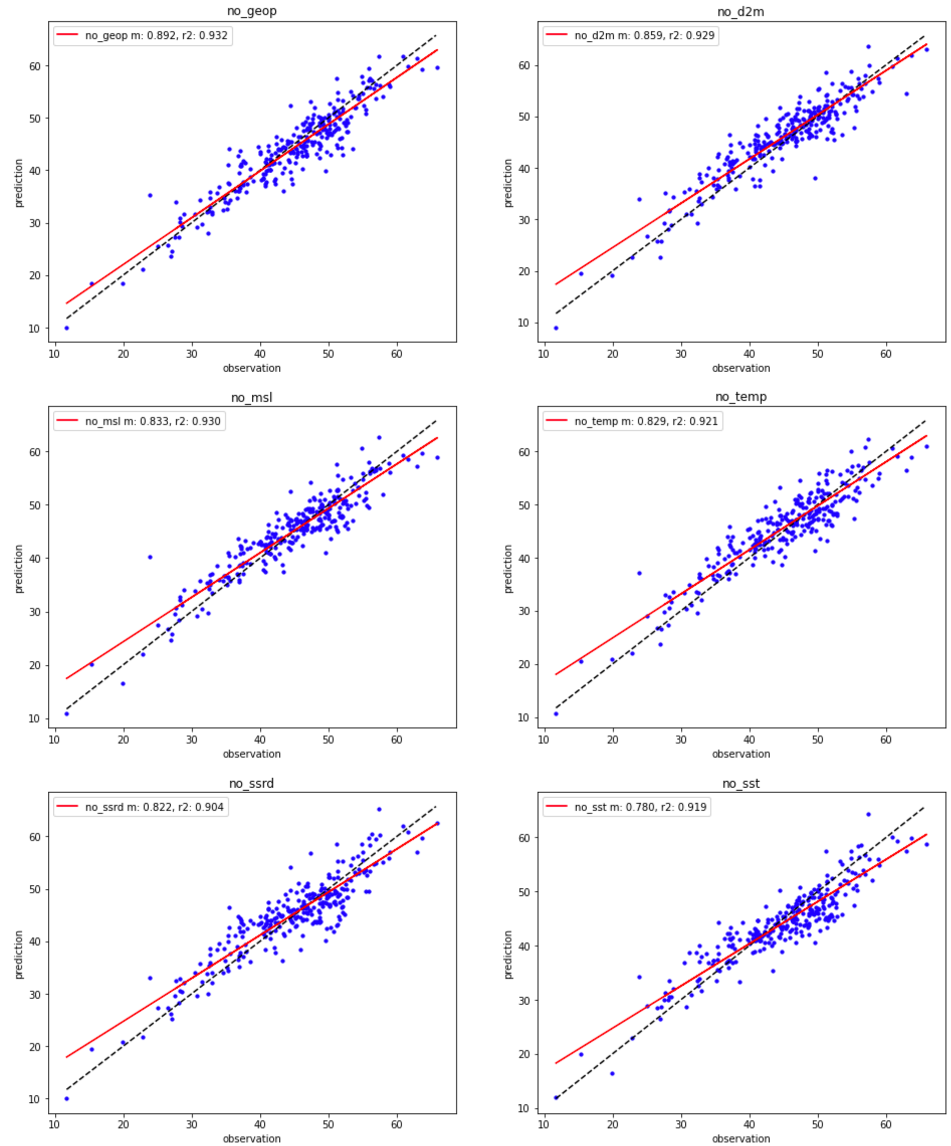


# Regional analysis CONUS

## 273 sites

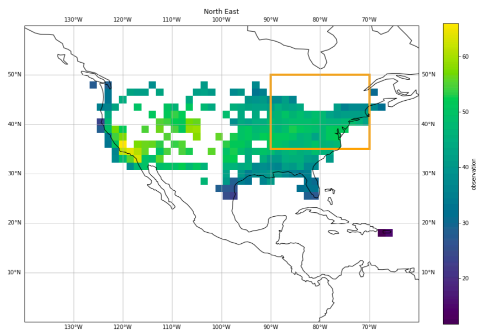


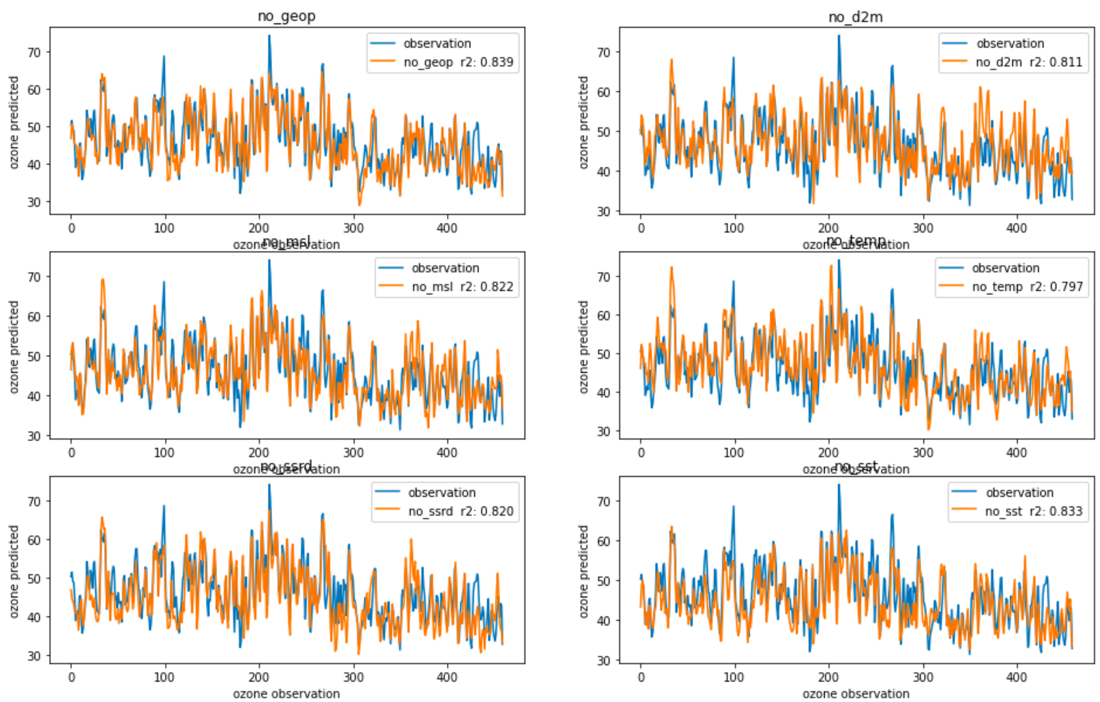



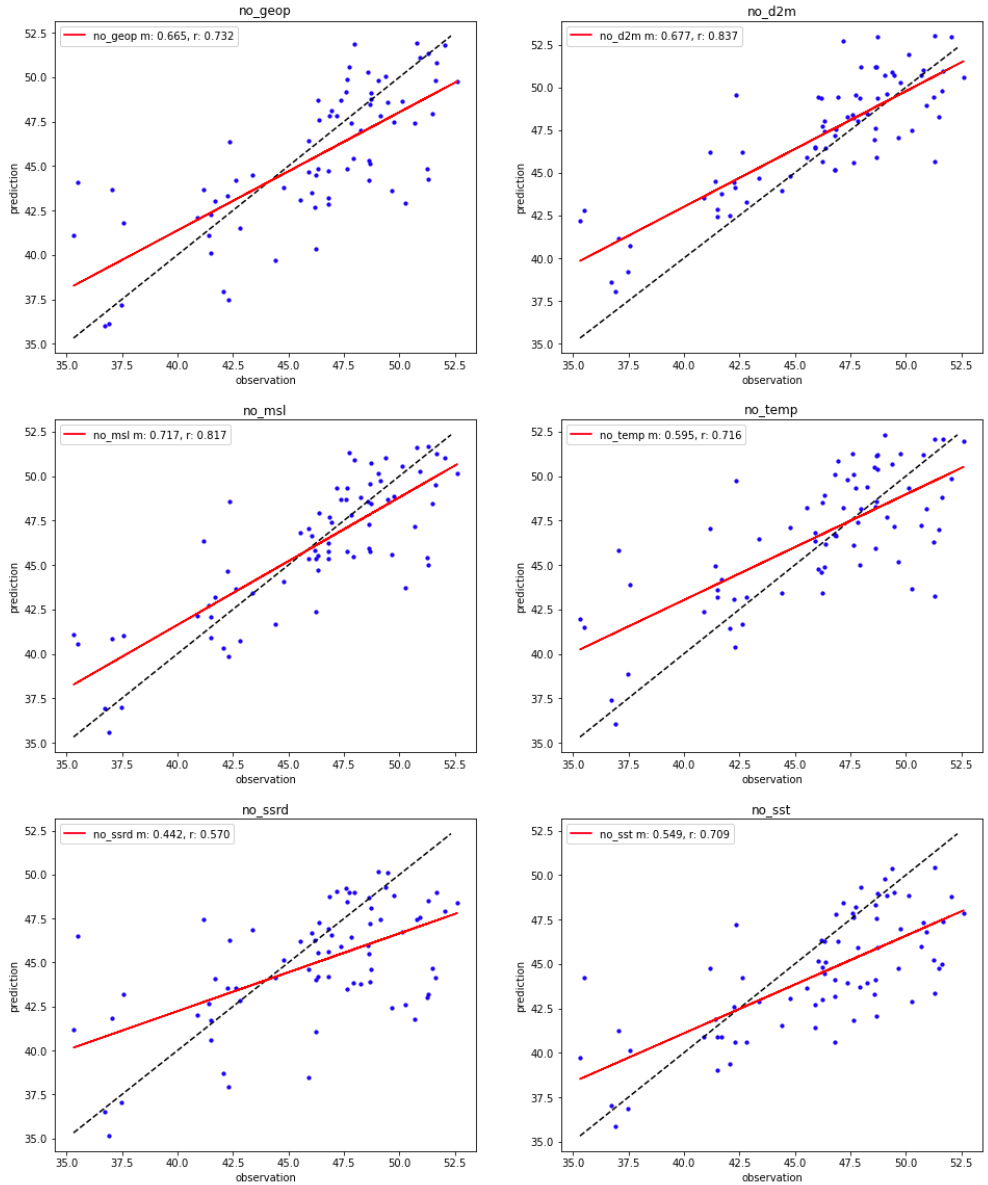



# Regional analysis Northeastern US

## 73 sites

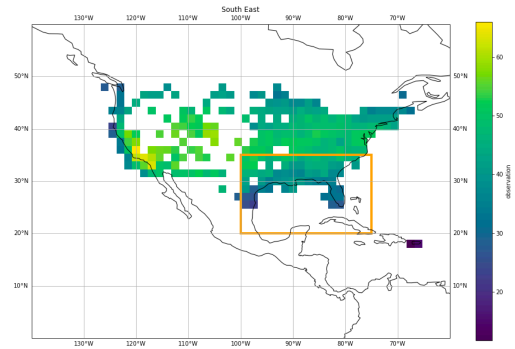


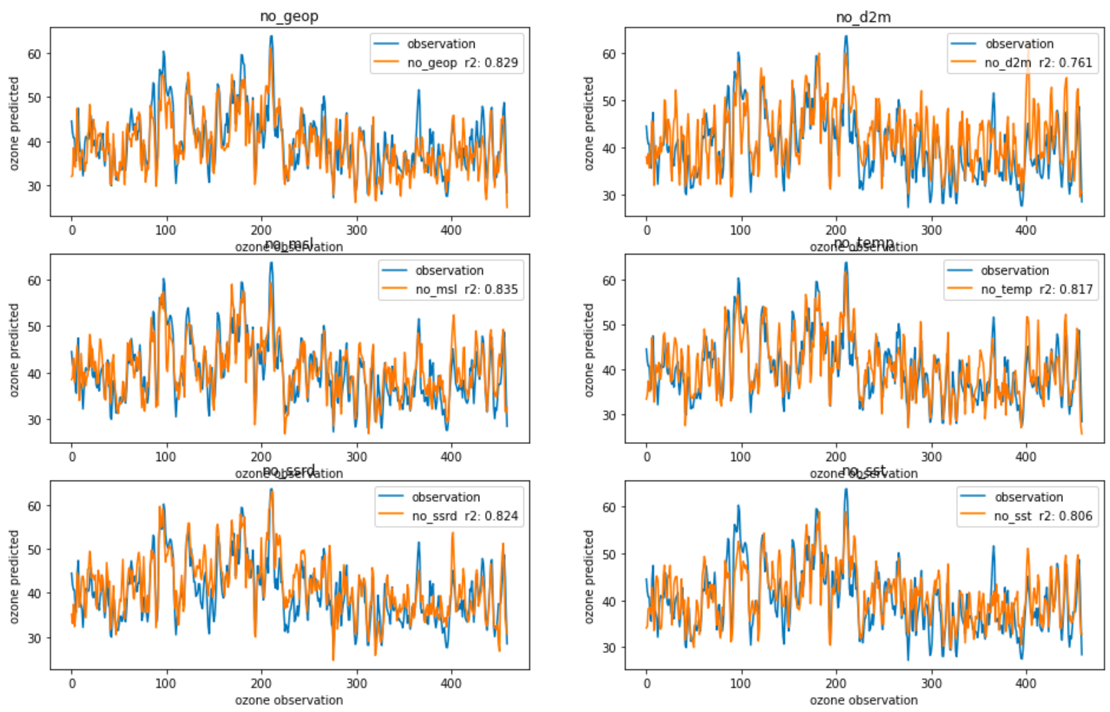



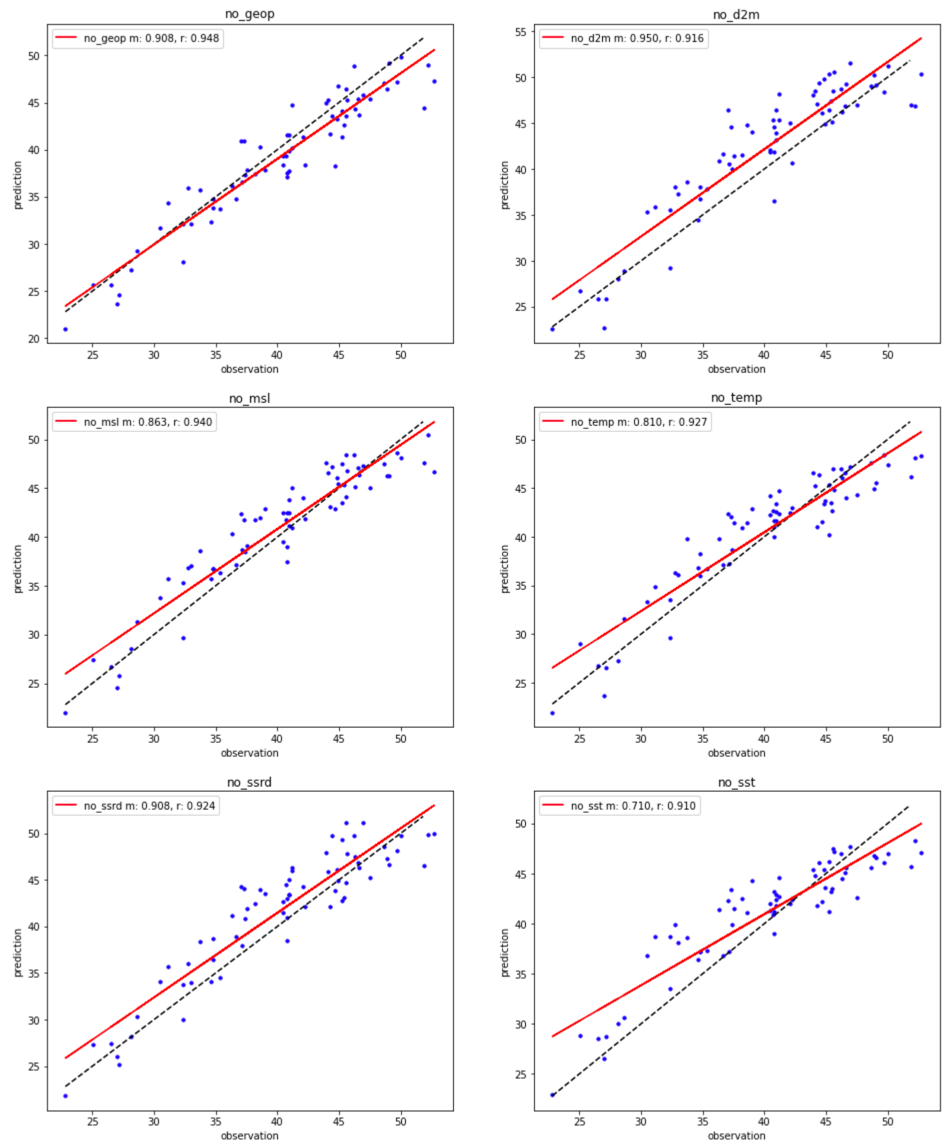



# Regional analysis
# Southeastern US

## 68 sites

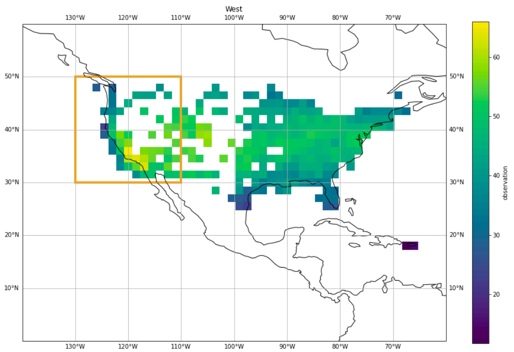


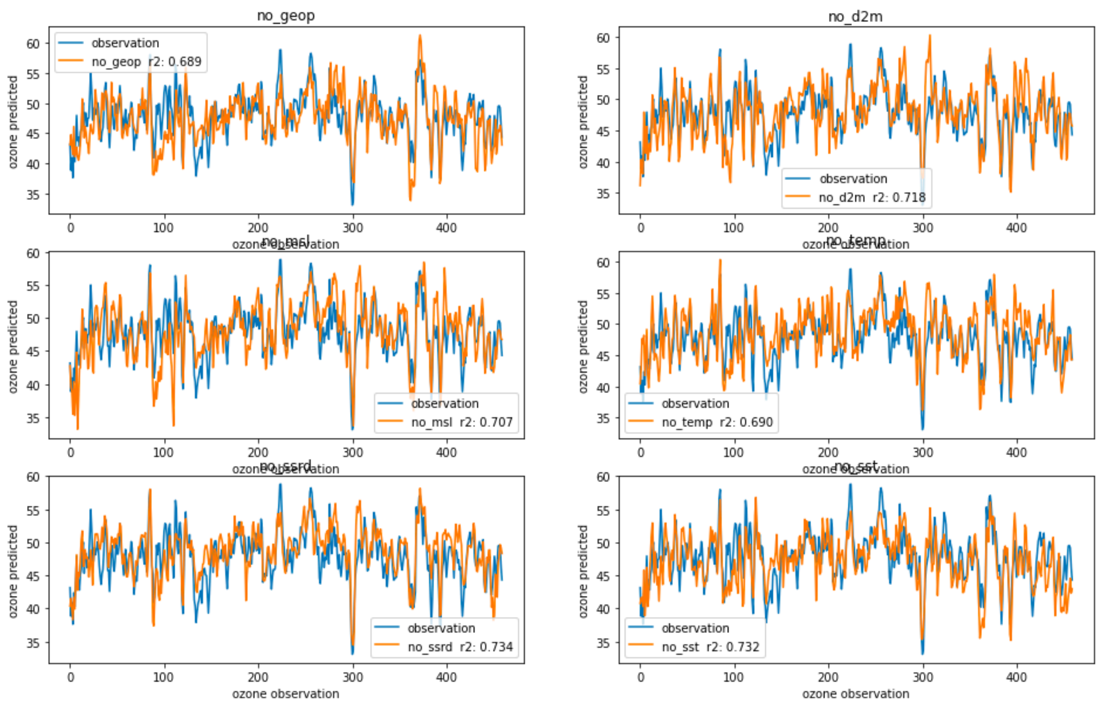



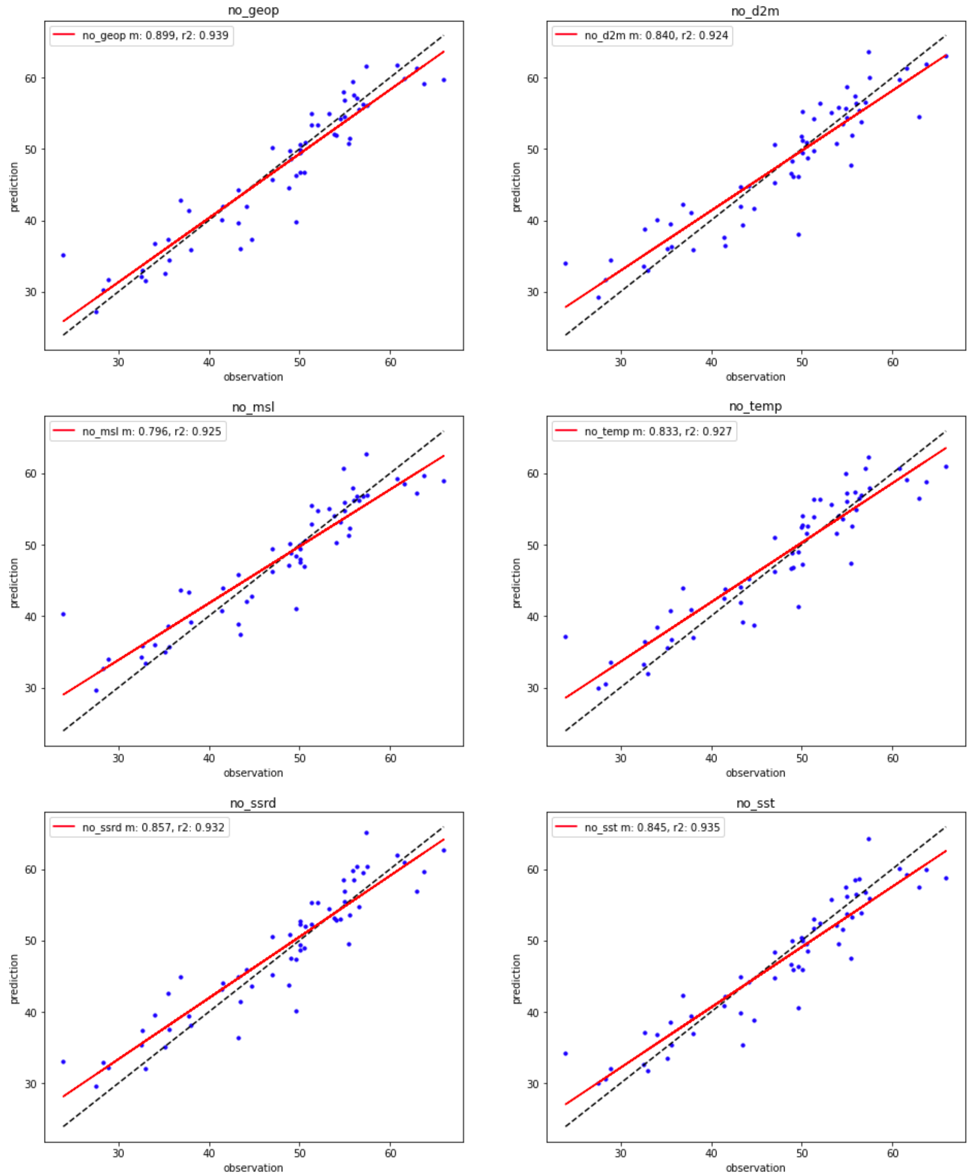



# Regional analysis
# Western Coast

## 58 sites

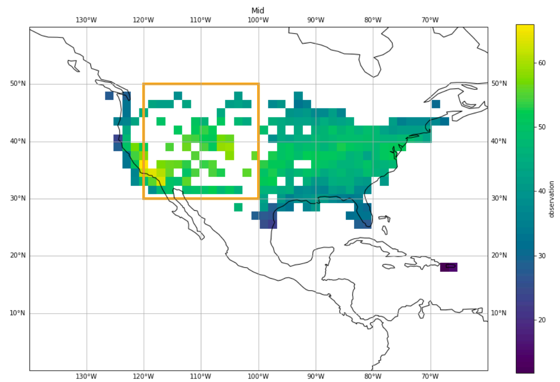


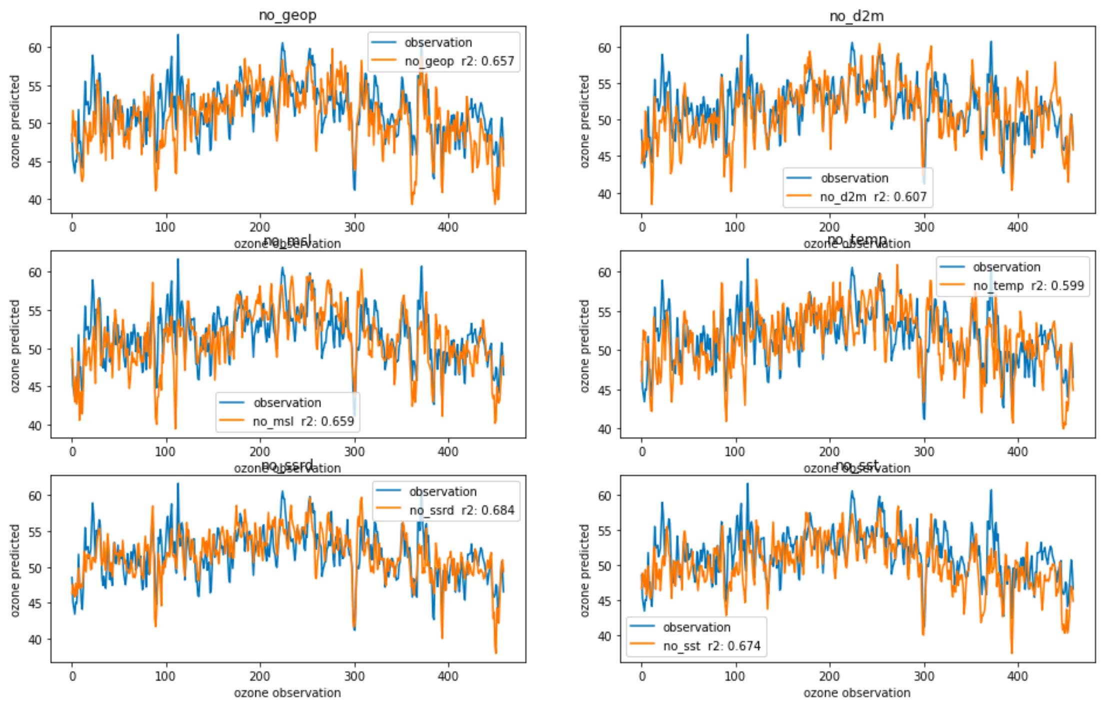



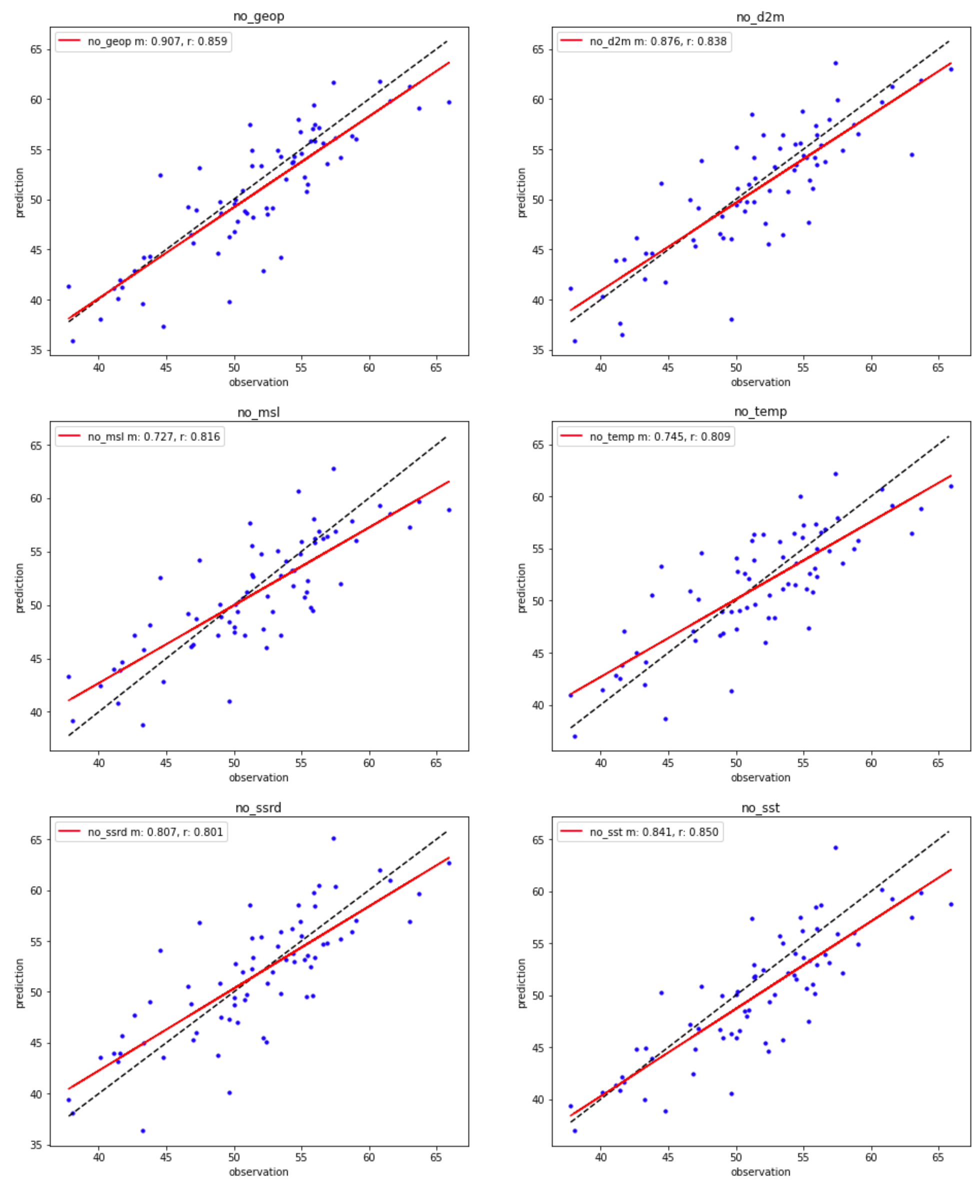



# Regional analysis Intermountain West

## 70 sites

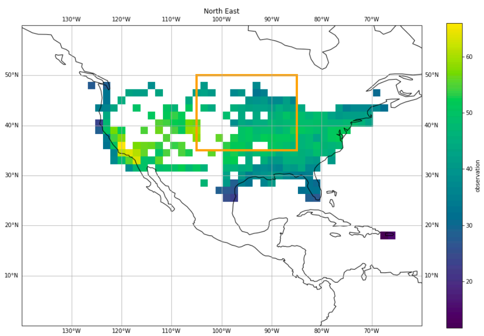


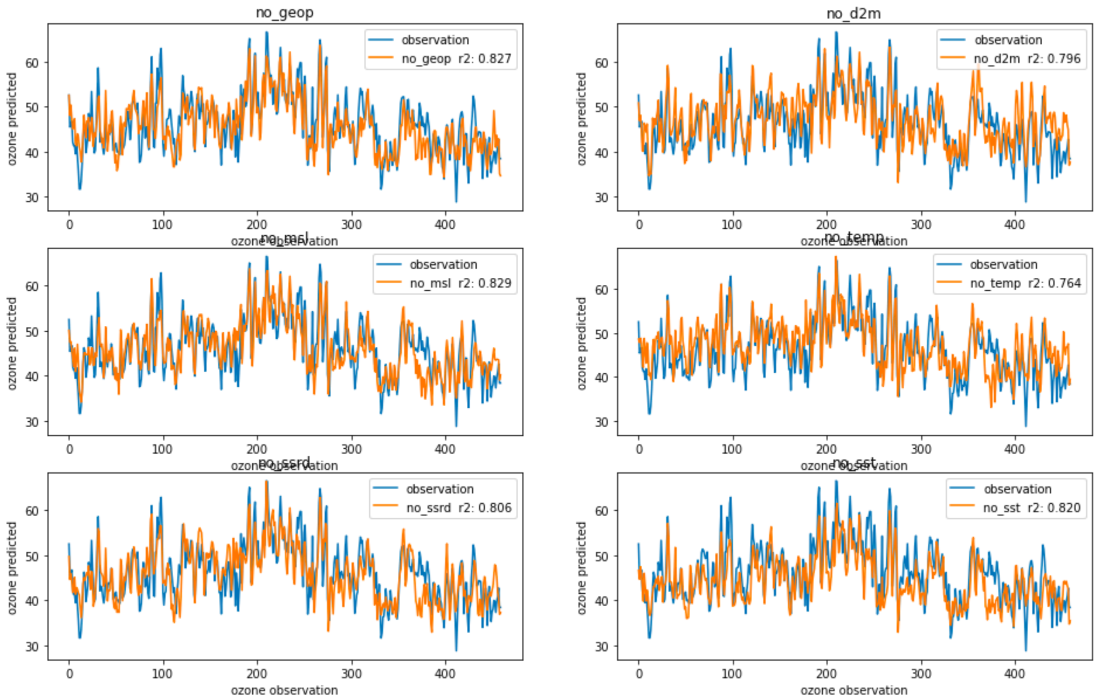



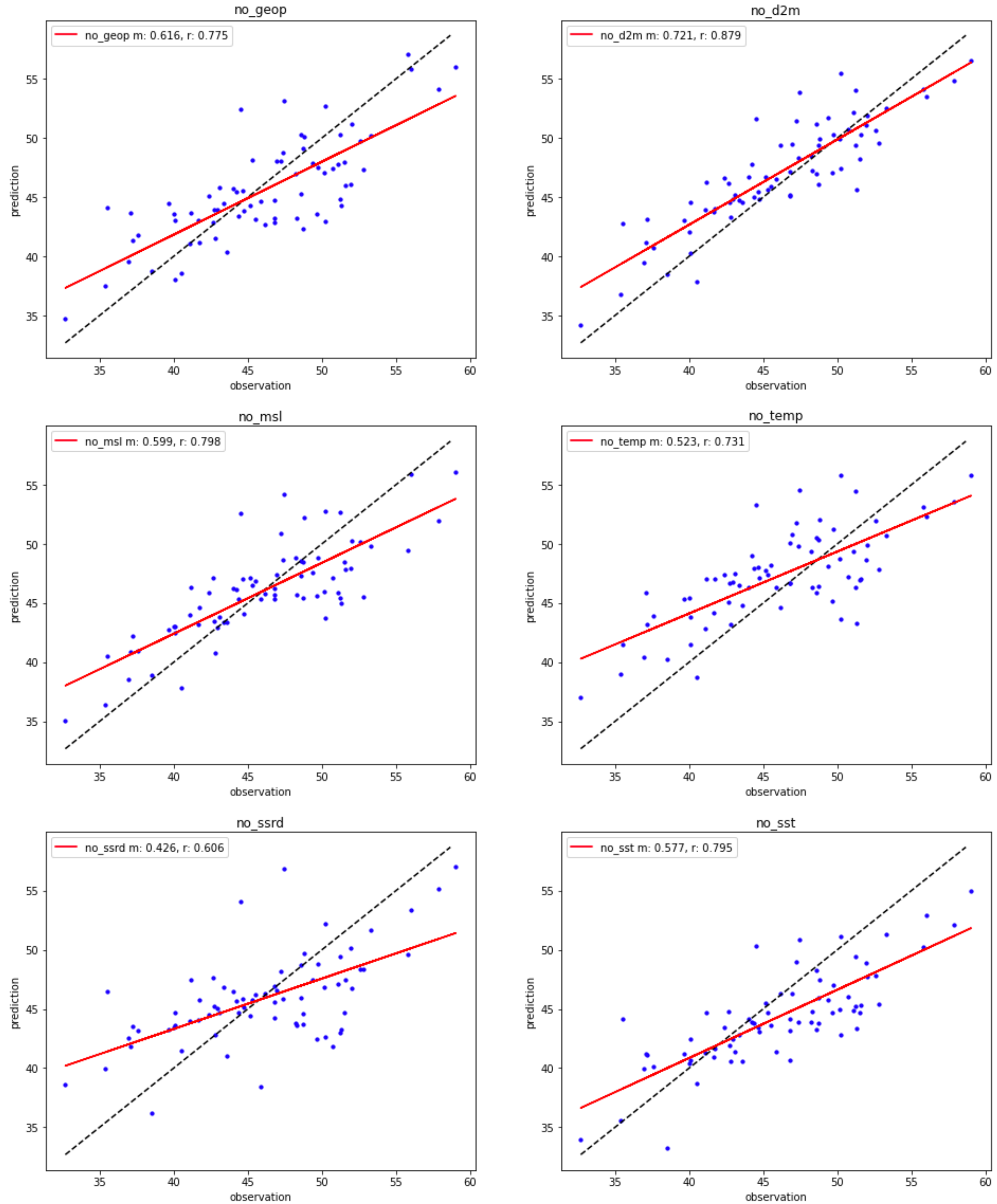



# Regional analysis
# North

# 72 sites



# Feature Visualisation

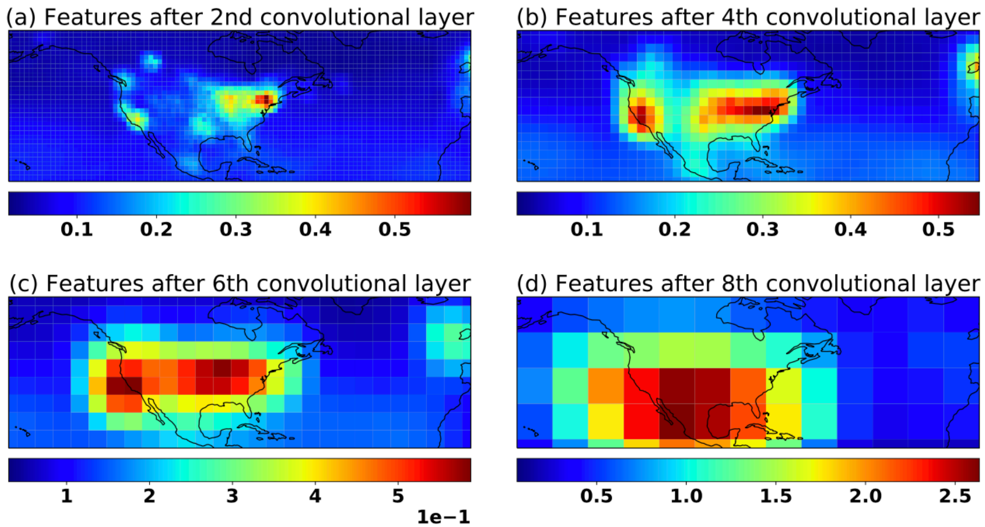
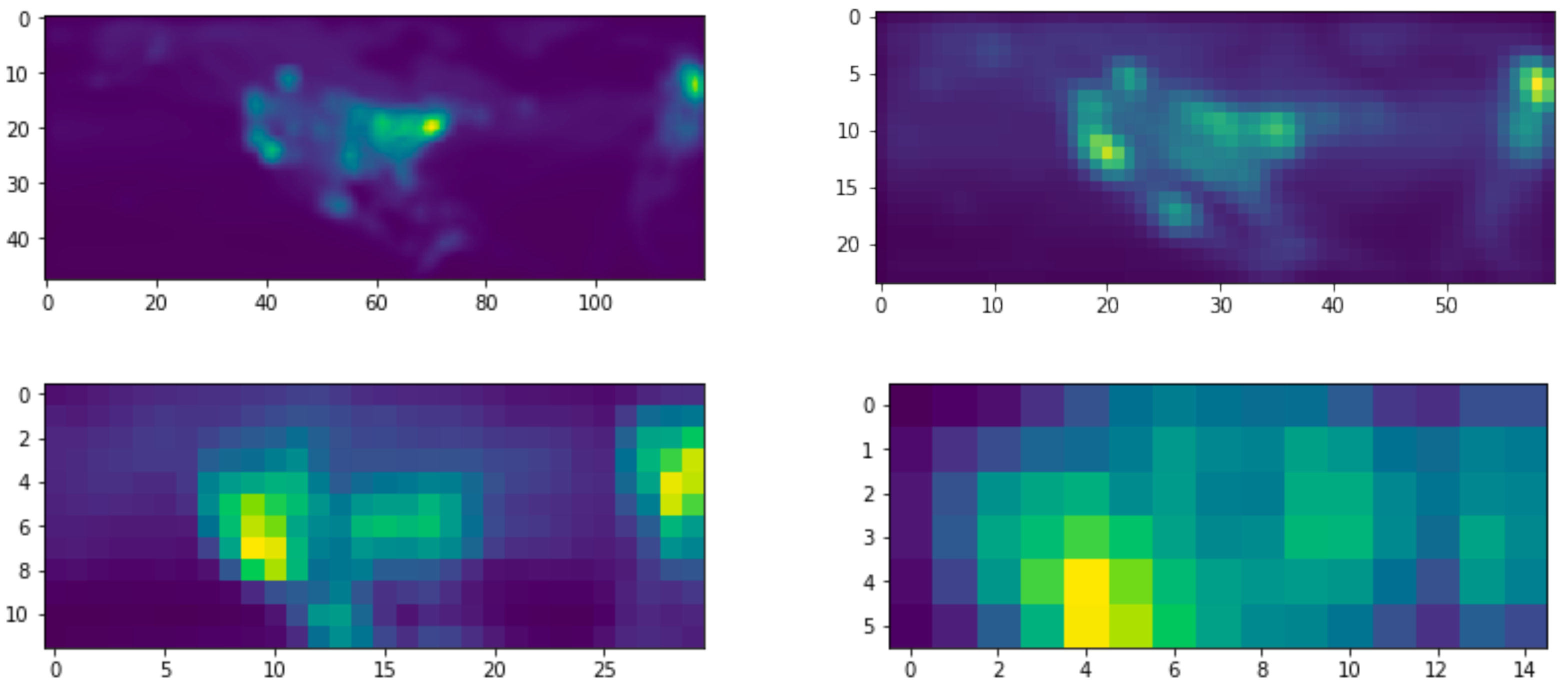



Figure 9: Feature maps from the 2nd, 4th, 6th, and 8th convolutional layers as illustrated in Fig. 4, averaged within all training data from 1980 to 2009.

**Figure 9**
**from Tai-long's preprint, Appendix D**

# Feature Visualisation
# Temp + NOx sectors



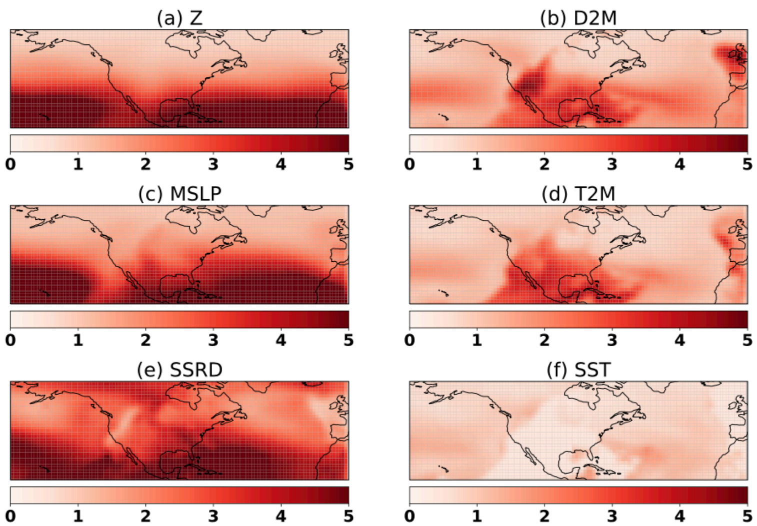



Figure 10: Averaged feature maps for the meteorological predictors, calculated from the training data set.

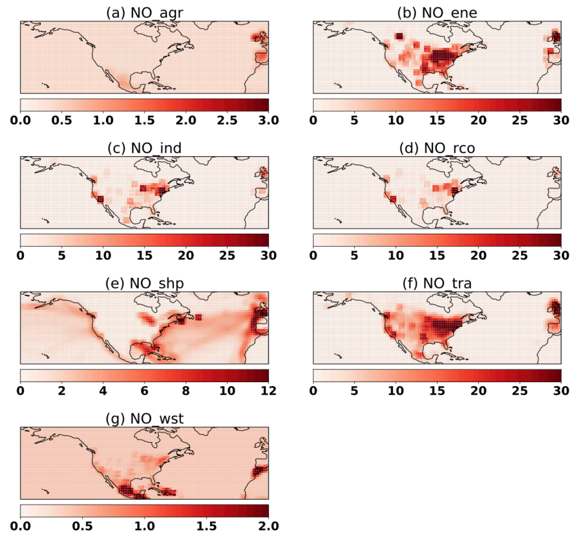



Figure 11: Average activation maps for NOx emissions, calculated from training data set. The seven NOx emission sectors are as follows: (a) agriculture (AGR), (b) the power industry (ENE), (c) manufacturing (IND), (d) residential and commercial (RCO), (e) international shipping (SHP), (f) surface transportation (TRA), and (g) waste disposal (WST).

**Figures 10 and 11**
**from Tai-long's preprint, Appendix D**

# Problems

- Does not account for the interactions between features when assessing model performance
- Hard to scrutinise temporal skill of models
- Need specific cases to asses (for feature visualisation)

# Towards Interpretability

- Adapt problem to classification - classifying ozone events?
- Analyse specific events / periods where model performs inadequately
- Two themes in Machine Learning research
    - Use less complex models —> reduced model skill
    - Develop model agnostic methods

# Shapley Values

- From game theory in Economics:

  *“Given a coalition, what’s the marginal contribution of each player?”*

- Given a set of features, what’s the effect of that feature on a particular prediction?

# SHAP

- SHAP (Python package) implements Shapley values, and integrates with certain ML frameworks (TF, scikit) and model-architectures smoothly.
- “Local” interpretability - explains particular instances
  - better developed for classification tasks and tree based architectures
- Aggregate shapely value statistics can be used for global analysis - feature importance

# Shapley Value Equation

Averaging over all possible subsets, orders of marginal contribution

Shapley value, $\phi_i$, for the local importance of the $i$ feature:

Marginal Contribution

$$\phi_i = \boxed{\sum_{S \subseteq F \setminus \{i\}} \frac{|S|!\ (|F| - |S| - 1)}{|F|!}} \boxed{\left[f_{S\cup\{i\}}\left(x_{S\cup\{i\}}\right) - f_S(x_S)\right]}$$

where:

Weighted average, equal weight for all combinations

Our model prediction with i

Our model prediction without i

$|S|$ size of the subset before we add the i[th] feature

|F| number of features

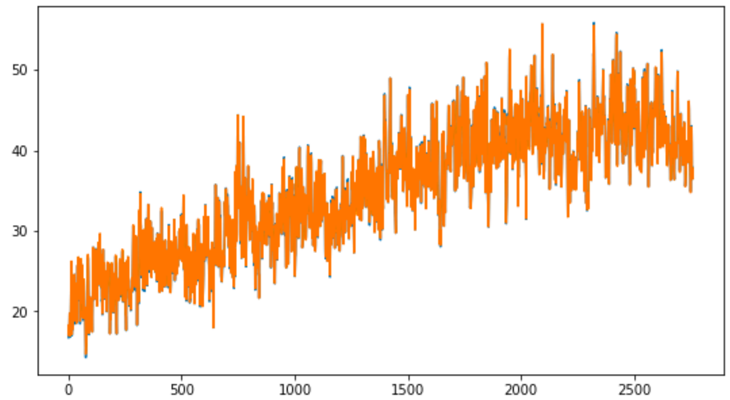
$\left[f_{S\cup\{i\}}\left(x_{S\cup\{i\}}\right) - f_S(x_S)\right]$ the marginal contribution

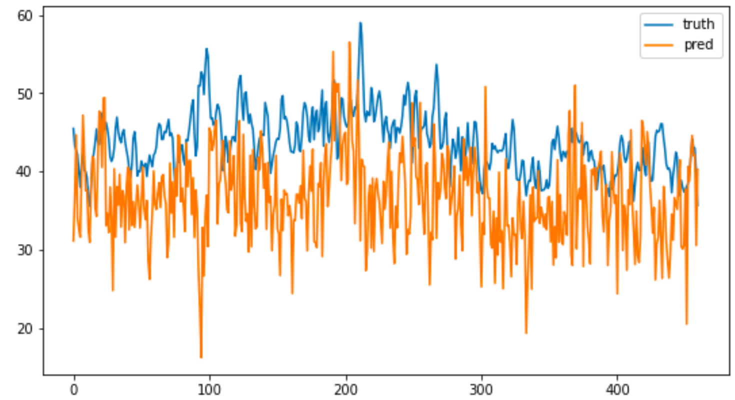
$\frac{|S|!\ (|F|-|S|-1)}{|F|!}$ is the weight for combinations for this occurrence

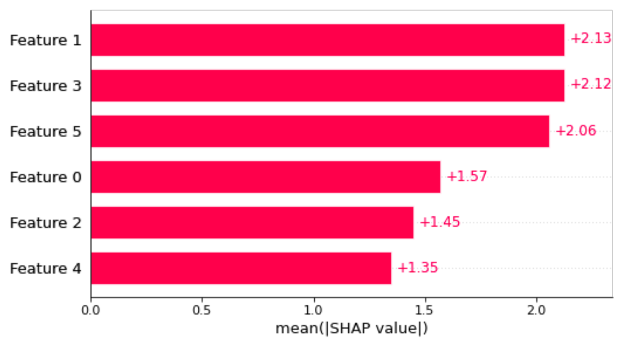
$S \subseteq F\setminus\{i\}$ is all possible subsets without $i$ feature, so we can add $i$

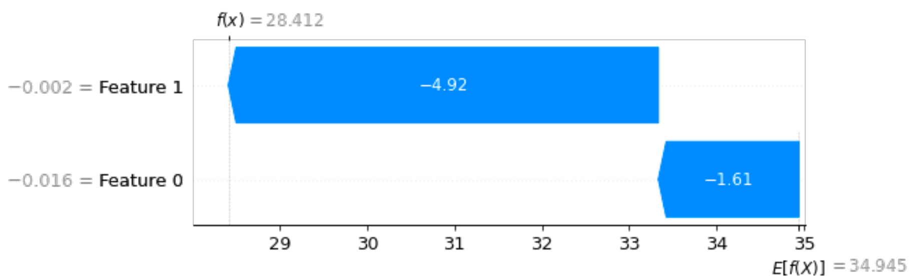
$S \cup \{i\}$ is subset $S$ with $i$ added and $S$ is a subset without $i$

# SHAP on XGBoost Model



# How to implement

- Shapley value for each observation
- Then perform aggregate statistics to see regional importance
    - computationally expensive

# List of resources

- Tai-Long He
  - arXiv preprint : Recurrent U-net: Deep learning to predict daily summertime ozone in the United States
  - GRL : Deep learning to evaluate US NOx emissions using surface ozone predictions

Christophar Molnar : Interpretable ML Book

Michael Pyrcz : Machine Learning lectures for Geostatics