Environnement Canada Environment Canada

Canada

The XM Tool Project:

The use of non-linear statistical models for Air Quality Forecasting

Andrew Teakles[1], Sean Perry[1], Aiming Wu[1], Qian Li[1], Stavros Antonopoulos[1], Ken Lau[1], Harry Yau[1], Rita So[1], Johannes Jenkner[2]

IWAQFR – Geneva

Dec 12th – Dec 14th, 2012

1 – Environment Canada, Canada
2 – University of British Columbia, Canada

XM Tool Project

Mandate:

- Develop and evaluate new non-linear tools for post-processing of air quality forecasts of O_3, NO_2, and $PM_{2.5}$

Purpose:

- Supports AQHI forecast program

- Key Objectives:
  - Improved guidance for air quality episodes.
  - Improve timing of air quality episodes.
  - Improve overall model forecast skill.

List of Predictors

Predictors

84 AQ Model Predictors

3 **Persistence Predictors**

27 **Antecedent Predictors**

Persistence Predictors

OBS at Hr 00

Antecedent Predictors

Lag 24/48/72 hrs
Max
Min
OBS

AQ Model Predictor types

- Meteorological Variables
- Sine Julian Day
- Day of the week
- NO2, O3, PM2.5
- Spatial Average, Min, Max

6 AQ Sites Tested

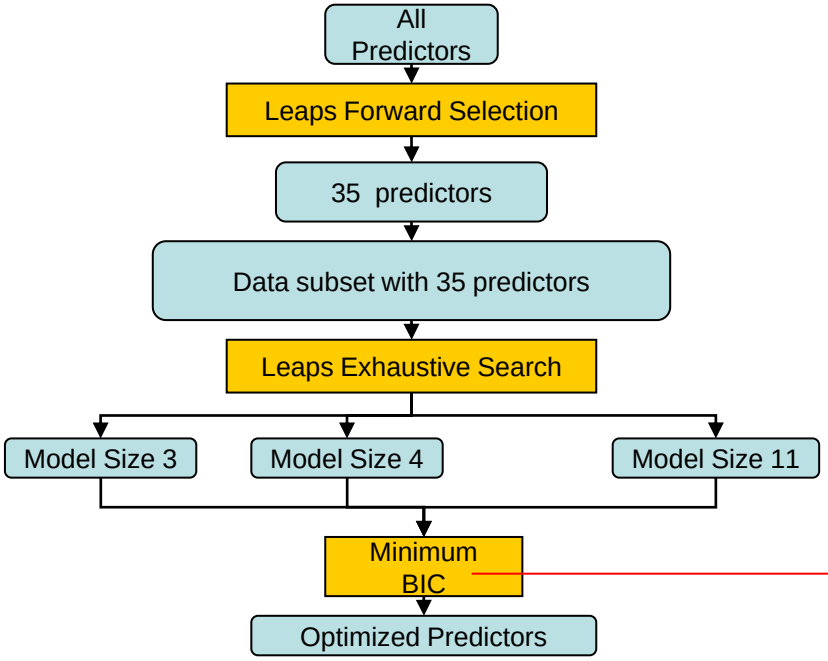
- Halifax
- Montreal
- Toronto
- Winnipeg
- Edmonton
- Vancouver

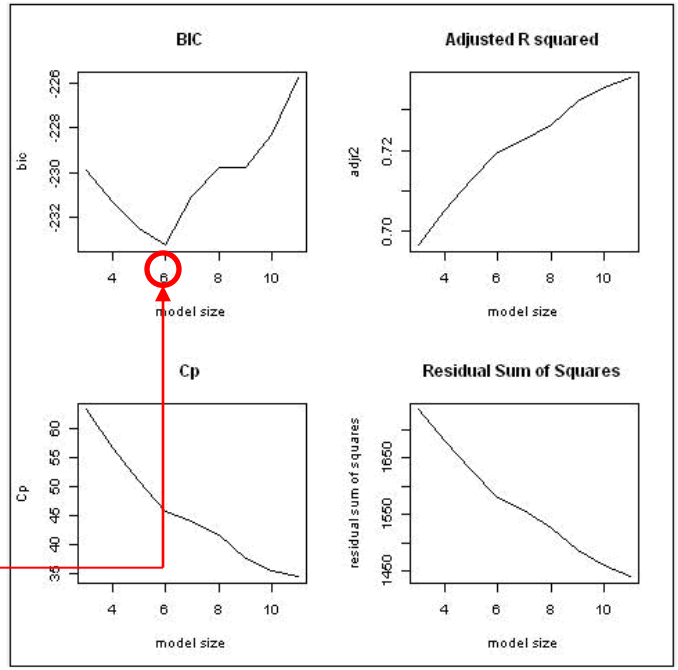
MLR approaches

	UMOS-AQ	XM version
Technique	MLR	MLR
Predictors	AQ model + Persistence	AQ model + Persistence **and/or Antecedents**
Predictor Selection	Forward Stepwise Regression	Leaps Minimum BIC (Ken Lau)
Seasonality	2 season	1 season
QA/QC	Min, Max, rate	Min, Max, rate
Minimum cases	250	250
Updatable	Yes	Not yet

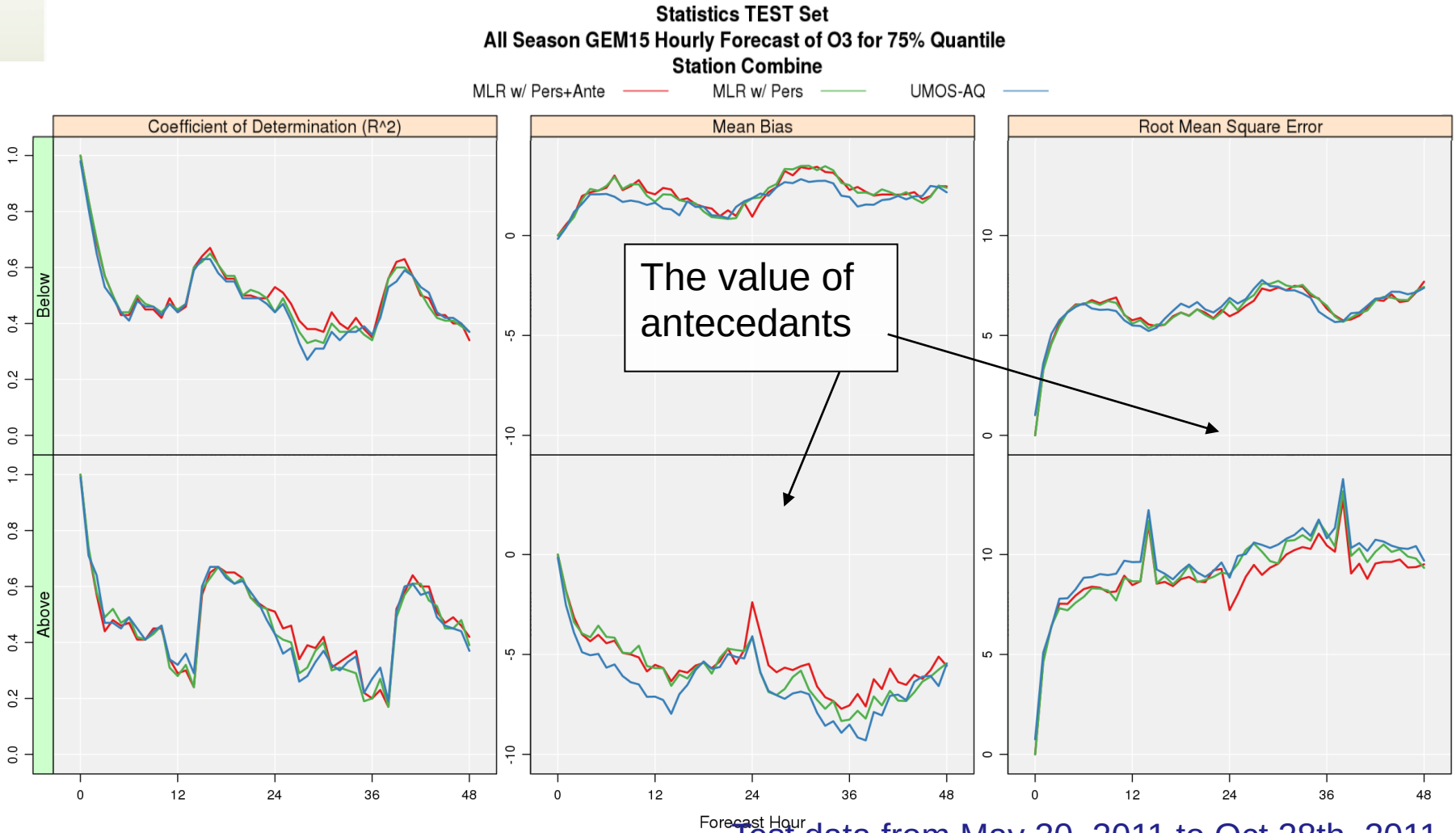
Flow Diagram for Leaps BIC Selection

- Leaps Package in R Statistical programming language
- Use combination of Forward and Exhaustive Search





MLR for O3: Persistance + Antecedents



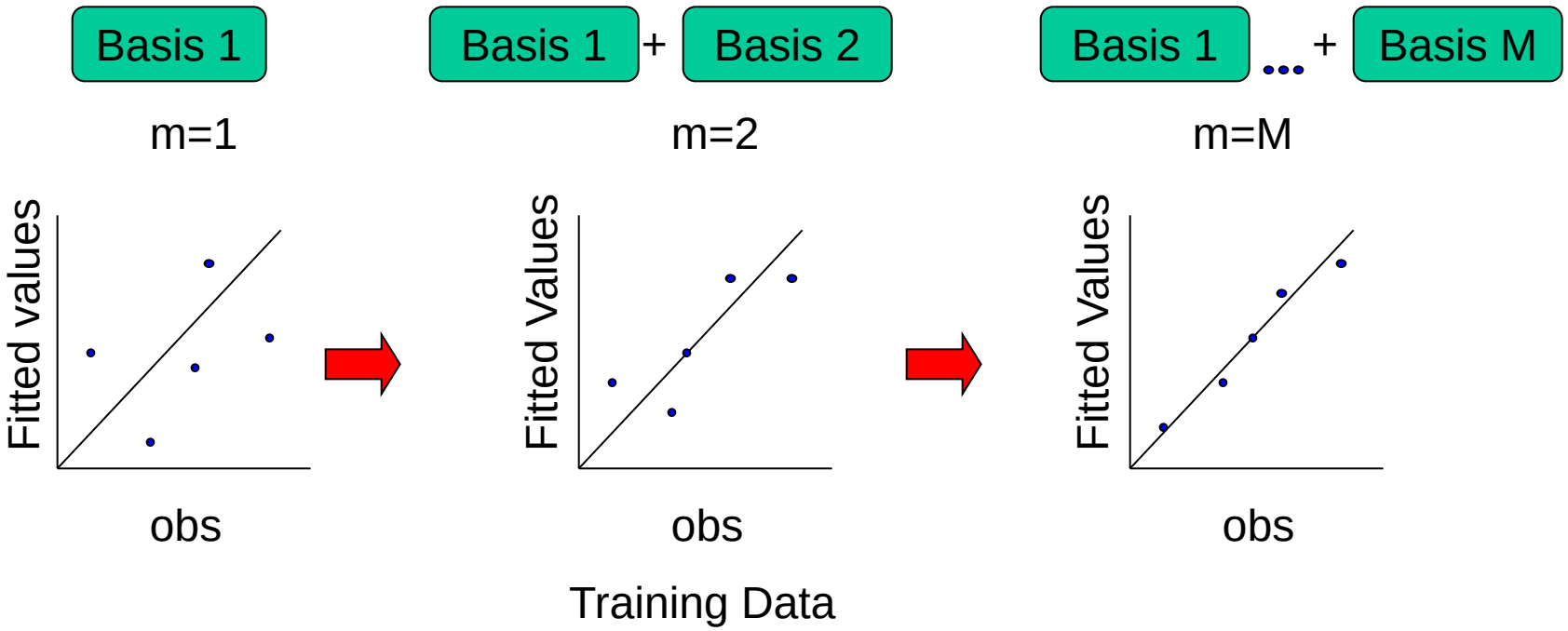
Test data from May 30, 2011 to Oct 28th, 2011

Techniques Tested

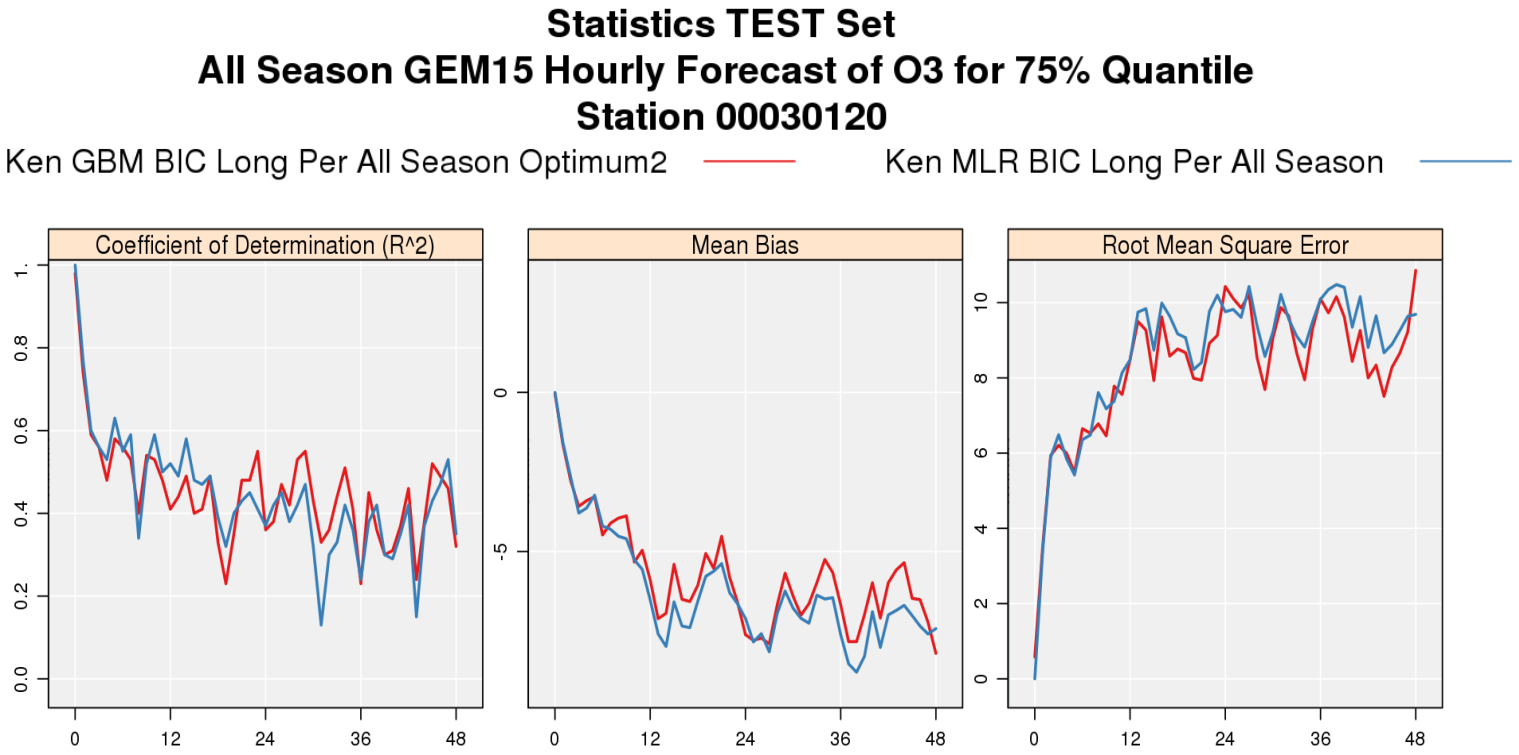
Method Tested	Status	Comments
Modified MLR with BIC predictors	Prototype	
Boosted Regression	Minor refinements	Optimization routine needed
Bayesian Neural Network	Needs Refinements	Revise predictors Matlab code
Support Vector Regression	Needs Refinement	
Neural Network with cross-validation	Rejected	Data loss
MDA	Rejected	
CART	Rejected	

Generalized Boosted Regression Model (GBM)

- Iterative Regression method
- Minimizes a certain loss function
- Applicable to many techniques [MLR, CART, SVR, …]

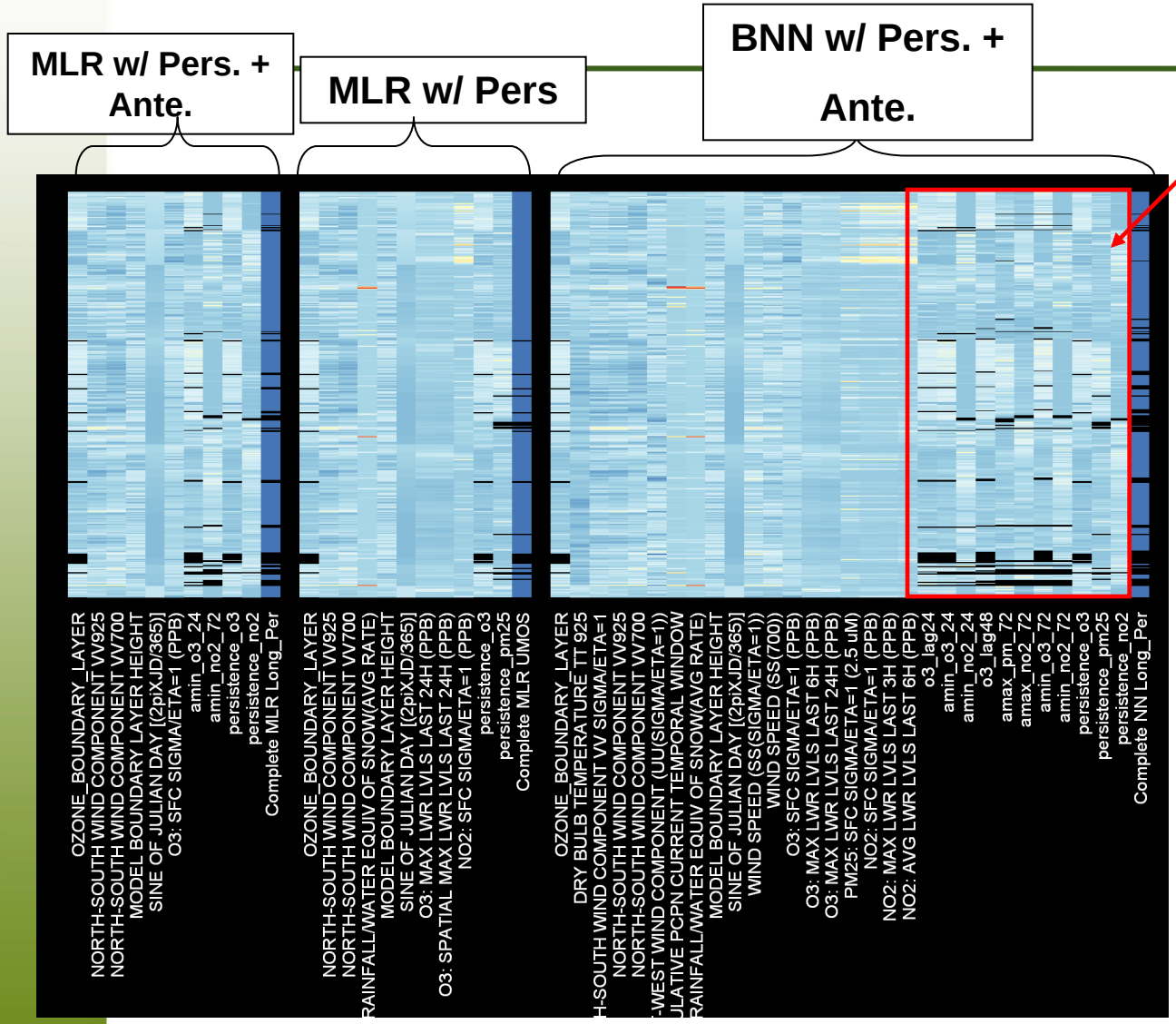


Verification Results for Above 75th Percentile cases



Shrinkage: 0.001 Iterations: 6000 Interactions: 9

Impact of Data Loss on Predictor Selection



Availability of antecedent preds.

Loss of training data due to its chose of antecedent predictors

*Last column indicates data availability for a given model (black space = loss of data)

Non-Linear Auto-Regression (prelim)

Approach: Relates trends in the Predictors and Predictand to improve the forecast accuracy

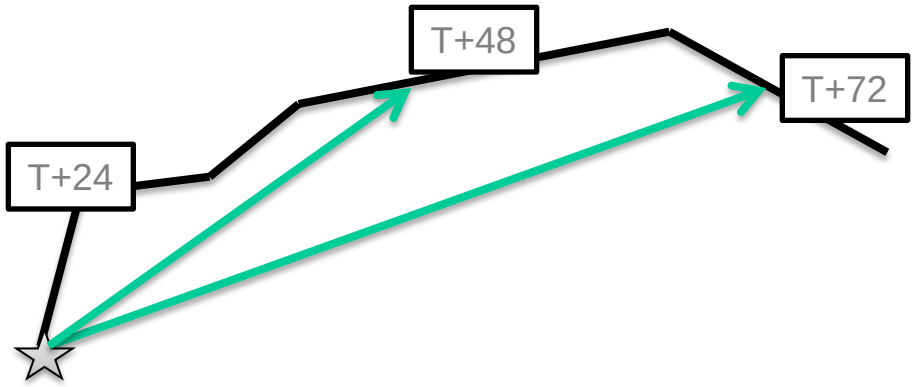
- Poster by Sean Perry

Trajectories (prelim)

Approach:

- Improve pattern recognition for transport events

- 7 launch heights (below 3km).
- Trace forward and backwards trajectories (00Z)



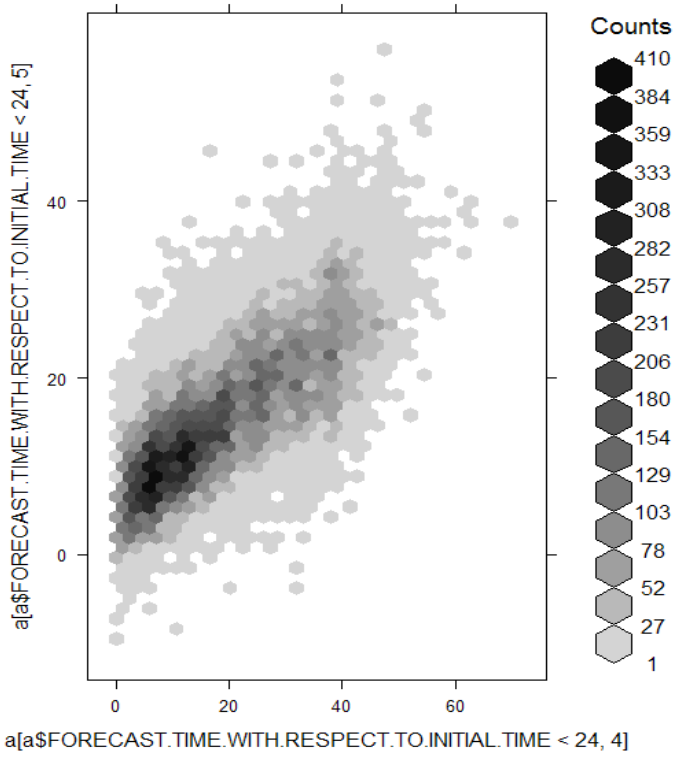
New Predictors

- Height
- Distance
- Bearing

Ozone for Halifax

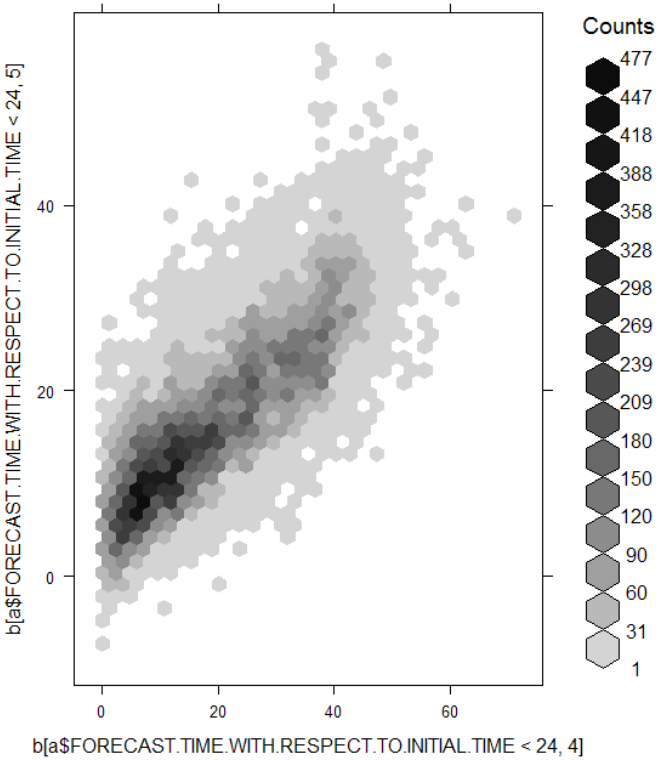
Trajectory MLR

$R^2 = 0.81$ MB= -2.473



MLR

R = 0.75 MB= -3.487



Lesson Learned

- Air quality forecast can be improved using statistical techniques
- Linear methods are not the only resource available for this type of work
- Improvement can be achieved by:
  - Applying non-linear techniques
  - Using better predictors
  - Refinement of your data treatment
- Conditional bias still a problem and there is more work needed

Path Forward

- Address data reliability issues
  - Has a greater impact on the non-linear methods
- More predictor testing
  - new GEM-MACH fields (chemistry)
  - Further trajectory testing
- New techniques
  - Recursive NN
  - Fuzzy Logic
- Evaluate learning rates
- Prototypes (early 2013)