# PROGNOS: A new MSC initiative to renew the operational statistical post-processing infrastructure


**Andrew Teakles[1], Jonathan Baik[1], Jacques Montpetit[2], Stavros Antonopoulos[2], Christian Saad[2], Naysan Saran[2]**

[1]Prediction Services Directorate, Meteorological Service of Canada (Halifax, Vancouver)

[2]Canadian Centre for Meteorological and Environmental Prediction Directorate, Meteorological Service of Canada (Dorval)


## Introduction

*PROGNOS is an initiative to renew the operational statistical post-processing infrastructure to accommodate evolving meteorological and air quality forecast program requirements and to improve compatibility with the NextGen forecast systems. Some PROGNOS features presented here are still under development and may differ in the final implementation.*

*RATIONALE:*

- *Difficult to adapt the existing system (UMOS) to frequent NWP updates due to design limitations.*
- *Reduce system maintenance costs*
- *Ability to apply new modeling strategies*
- *Better serve research and development projects.*

## Design Concept

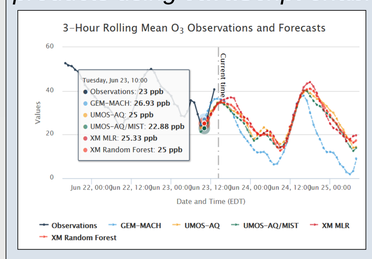
*PROGNOS is based on an experimental post-processing system (XM-Tool) designed in support of the Air Quality Health Index program and prototypes tested for the 2015 Toronto Pan Am and ParaPan Am Games*

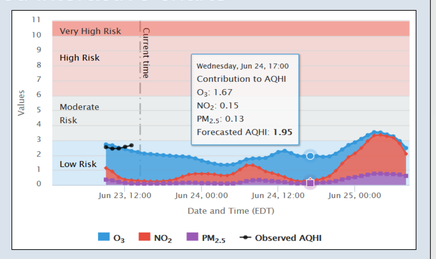
***FEATURES:***

- *Modular design with customizable components*
- *Extensive statistical modeling libraries*
- *Data-driven approach for increased automation*
- *Relational database usage for improved data treatment*
- *Diverse case weighting options to ease transitions between seasons and frequent updates to model designs*

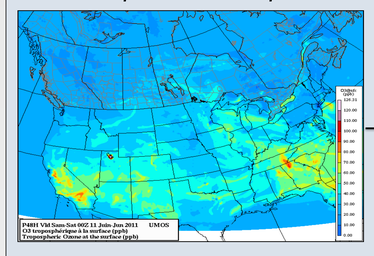
## Product Generation

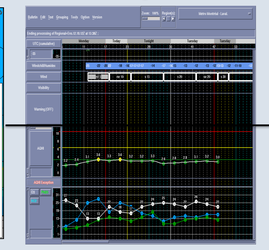
***Forecast dashboard*** *provides multi-product overview and AQHI products using JavaScript enabled interactive charts*




***Existing Operational Products*** *will also be supported by PROGNOS to provide* ***MIST Interpolated*** *forecast,* ***SCRIBE****, and* ***NextGen*** *products to operational meteorologists and clients.*



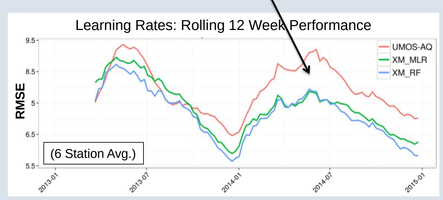


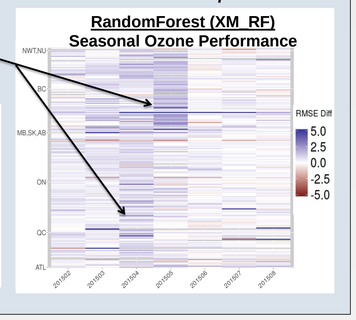
NextGen Systems

Forecaster Clients

***Verification and Diagnostics*** *tools in PROGNOS give additional products to evaluate statistical models and their learning rates. PROGNOS integration with operational EMET verification planned.*

*Improvement over UMOS-AQ given sufficient training*




## INTERNAL/EXTERNAL DATA SOURCES

**Meteorological Forecast Models**

**Real-time Observations**
Meteorology
Air Quality

CMC Archives

## PROGNOS Post-Processing Infrastructure

Data Ingest

DATA CENTRE

SQL Database

Gridded Data

CALCULATION ENGINE

Calculation Control

Config Files

MAESTRO Sequencer

Calculation Repository

Data Treatment

Machine Learning

Cluster Computing

Product Generation

## Data Ingest

***Data Driven:*** *Meteorological Forecast are interpolated to observation locations based recent availability and covariate predictor variable are also generated at those location as needed.*

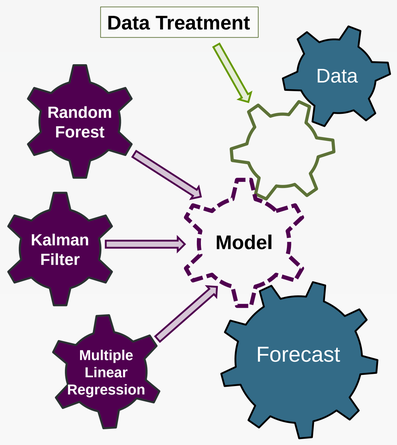
***Metadata*** *on the geographical location, observation type, data quality, and data source is maintained to facilitate further data treatment and record matching by PROGNOS.*

***Station TS records:*** *Time series (TS) of observation and predictors are matched over the record history and archived by station. Station TS records allow PROGNOS to have more control on data treatment and machine learning techniques*

## Modular Modelling

***The MAESTRO*** *suite provides a modular structure for statistical modeling.* ***Generalized templates*** *for data treatment, modelling and forecasting tasks can be customized using the MAESTRO configuration settings, leveraging a collection of PROGNOS scripts that use functions within the calculation repository.*

***PROGNOS R package*** *was developed to provide specialized functionality to tasks that fit within the framework of the PROGNOS system enabling* ***easy prototyping*** *with* ***plug-in*** *components.*

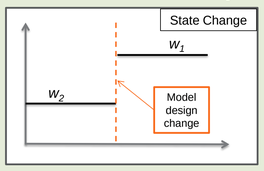


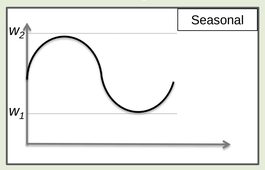

## Data Treatment

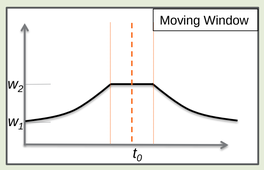
***Pre-Processing***

Filters (QA /QC, Completeness)
Predictor Generation
Antecedent (lag) Predictors
Transforms (if desired)
Data normalization
Variable Importance/Selection
Data Imputation
Data linkage/merging
Data pruning
Case weighting

**Case Weights** can adjust the importance of cases based on their number, seasonality, and/or forecast performance.




## Timeline

- **2015** • Pan Am Demo
- **2016** • PROGNOS Initiated • MAESTRO • Extended to Meteorology
- **2017** • New Data Ingest • Modeling Capacity • Document Demo
- **2018** • Experimental

## Machine Learning

*PROGNOS* adopts higher level languages (R, Python**)** with extensive statistical modeling libraries to support machine learning. Prototype methods were run twice-daily for air quality ($O_3$, $NO_2$, $PM_{2.5}$) in support of the 2015 Toronto Pan Am and ParaPan Am Games.

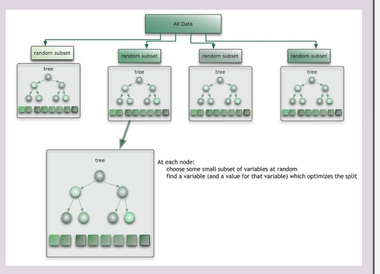
***Current Prototypes:***

- *Random Forest*
- *Linear Regression*
- *Kalman Filter*

***Predictor Selection****:*

- *Stepwise Regression*

*using the R "leaps" package*




**Batch updates** using the station TS record were done weekly. The current system (UMOS) using an **online learning** method that is more efficient (recent data updates terms) yet less flexible (only linear methods). Both batch and online updates will be possible with PROGNOS.

## Future Work

- *Experimental run with linear model designs*
- *Evaluate new modeling strategies for air quality and meteorology*
- *Prototype statistical ensemble forecast products*
- *Support the development of gridded post-processing methods*

Environment and Climate Change Canada | Environnement et Changement climatique Canada