

A Machine Learning-based Approach for Bias Correction of Microwave Radiances in Regional NWP

Alice Abramowicz



Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Infrastructuur en Waterstaat

Introduction

HARMONIE-AROME

Observation biases

Variational Bias Correction (VarBC) linear model:

$$b(x, \beta) = \beta_0 + \beta_1 p_1(x) + \beta_2 p_2(x) + \dots + \beta_n p_n(x)$$

VarBC updates Betas every few hours inside NWP model

→ Predict β 's

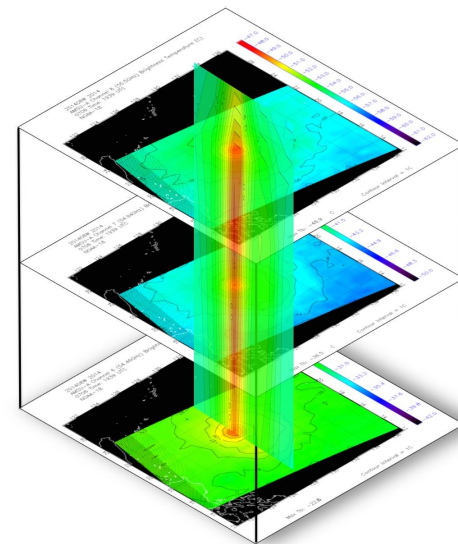
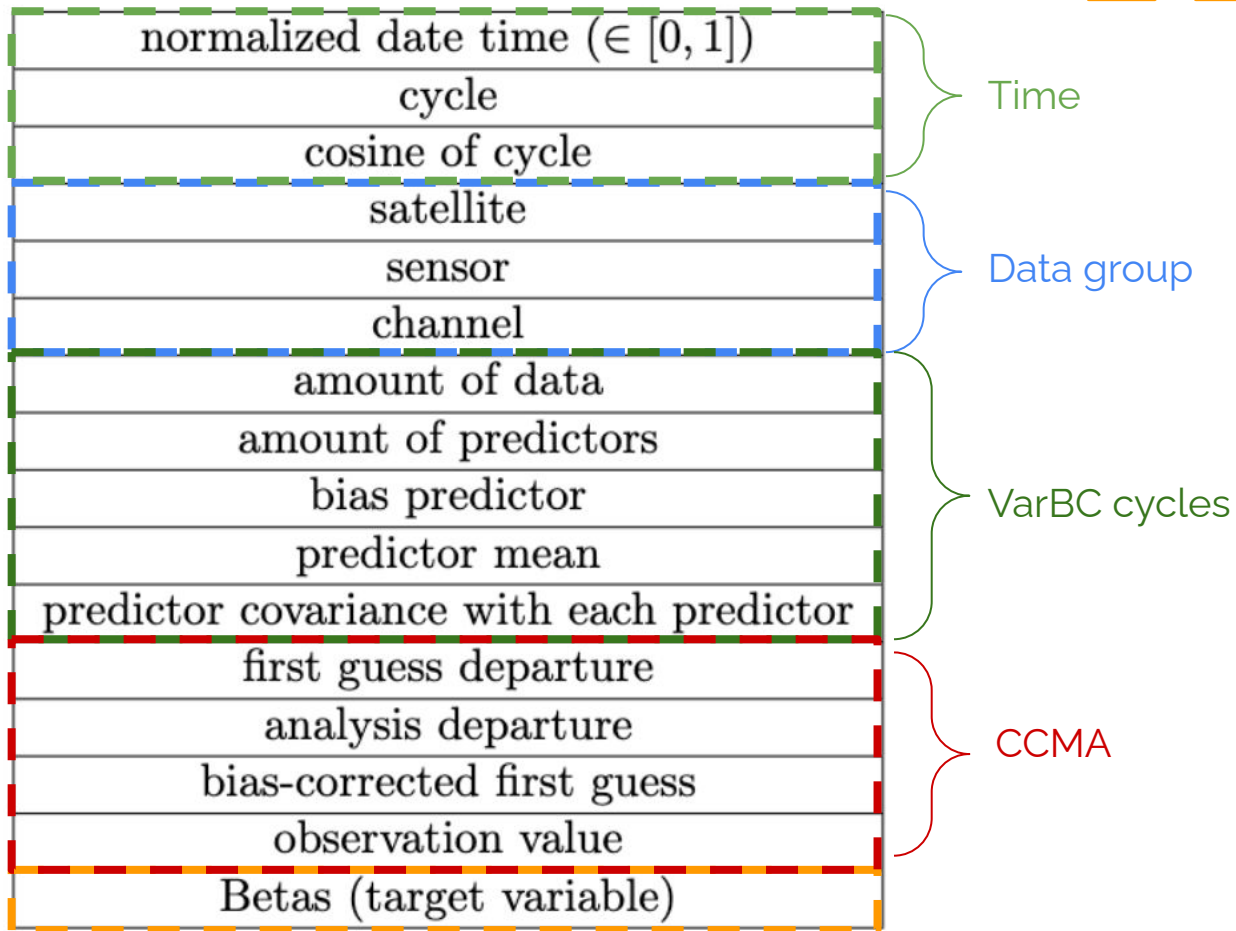
Research questions

$$b(x, \beta) = \beta_0 + \beta_1 p_1(x) + \beta_2 p_2(x) + \dots + \beta_n p_n(x)$$

1. Can ML models emulate VarBC in terms of estimating the betas outside of the data assimilation scheme?
2. Is knowledge transferable from one instrument to another?
3. How does the traditional HARMONIE-AROME 4D-Var compare with the ML-based HARMONIE-AROME 4D-Var?

Metadata

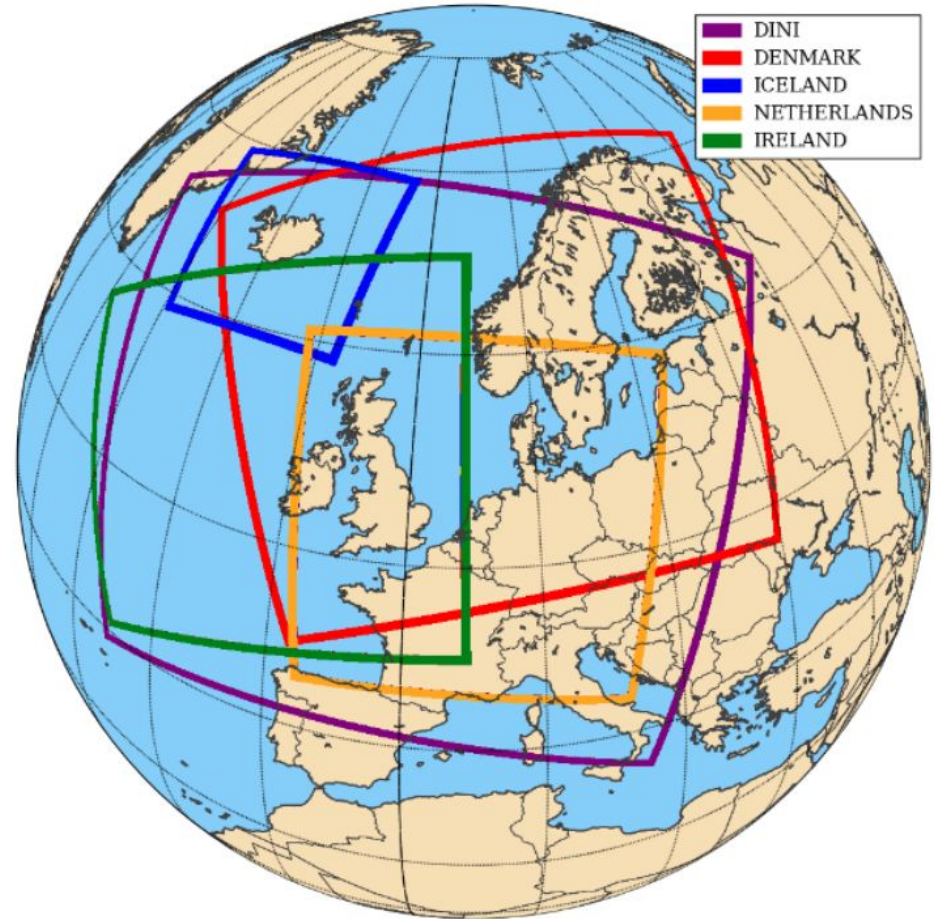
$$b(x, \beta) = \beta_0 + \beta_1 p_1(x) + \beta_2 p_2(x) + \dots + \beta_n p_n(x)$$



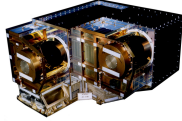
Research design

DINI domain (2023) for training

Dutch domain (2021) for testing



Data groups



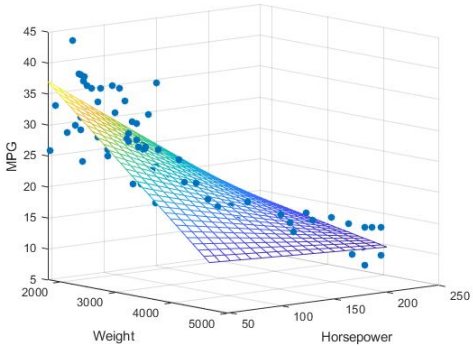
Sensors/ Satellites	AMSU-A	MHS	MWHS2
Metop-B	channels 7, 8, 9	channels 3, 4, 5	-
Metop-C	channels 6, 7, 8, 9	channels 3, 4, 5	-
NOAA18	channels 6, 7, 8	-	-
NOAA19	channels 6, 9	channels 4, 5	-
FY-3D	-	-	channels 4, 5, 6, 11, 12, 13, 14, 15

Baseline model & Machine learning (ML) models

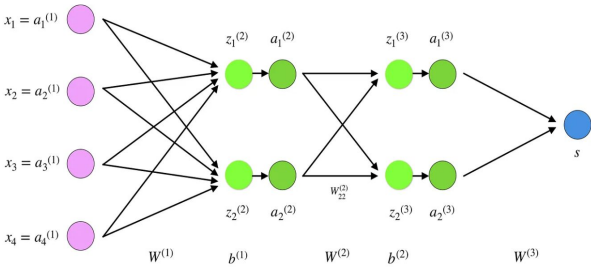
Theoretical overview

Models

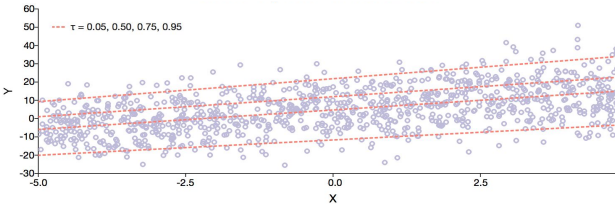
Multiple Linear Regression (MLR)



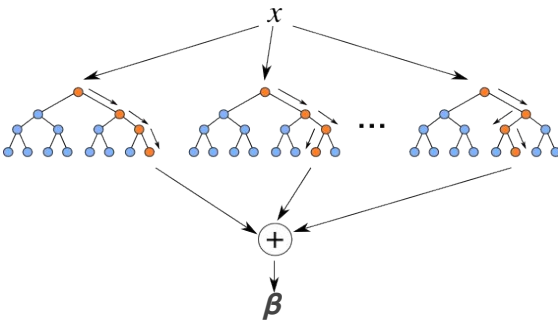
Artificial Neural Network (ANN)



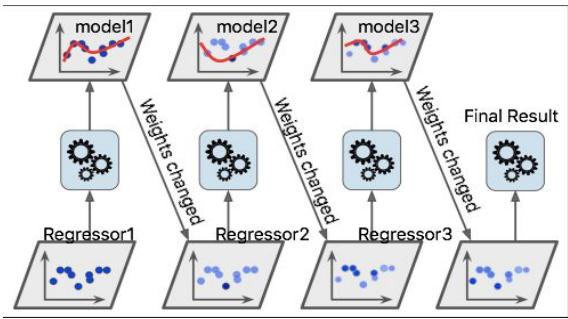
Quantile Regression Forest



Random Forest Regressor (RFR)



Adaptive Boosting (AdaBoost)



Not shown

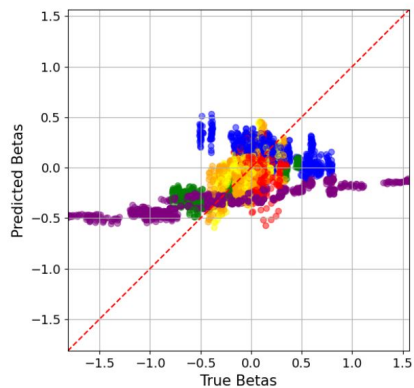
Regression models

Results

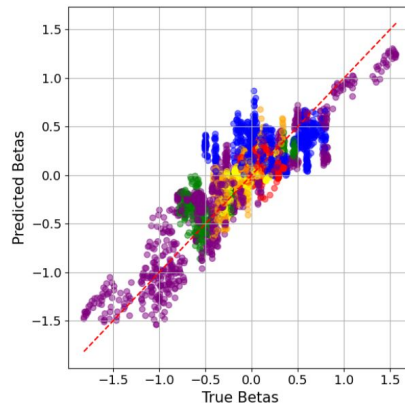
Can we predict the Betas using ML methods instead of VarBC?

Results of regression models

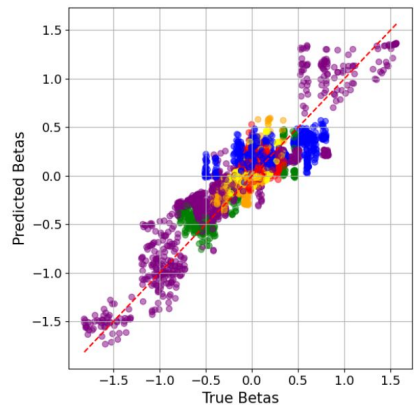
$$b(x, \beta) = \beta_0 + \beta_1 p_1(x) + \beta_2 p_2(x) + \dots + \beta_n p_n(x)$$



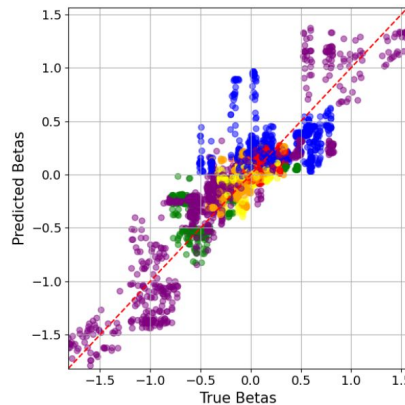
(a) MLR



(b) ANN



(c) RFR



(d) AdaBoost

Scoring metrics/Models	MSE	R ²
Artificial Neural Network	0.0337	0.6734
Random Forest Regressor	0.0257	0.7502
AdaBoost	0.0332	0.6771
Baseline model	0.1015	0.0140

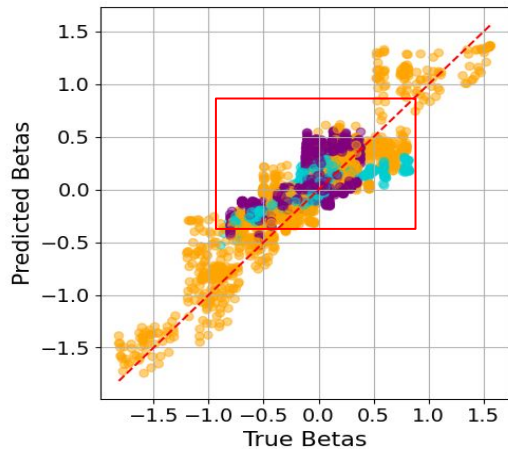
On-demand sub-km domains?

Random Forest Regressor (RFR): Sensitivity analysis

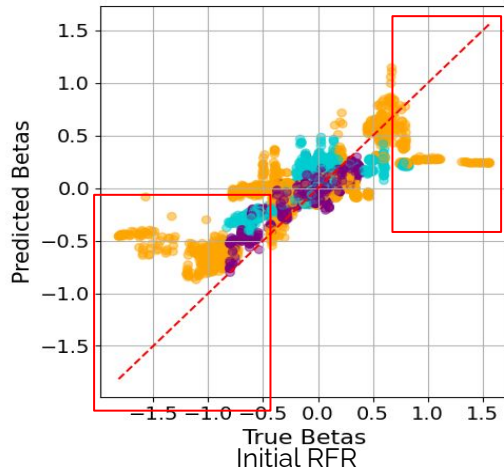
Is knowledge transferable from one instrument to another?

Remove a sensor from training

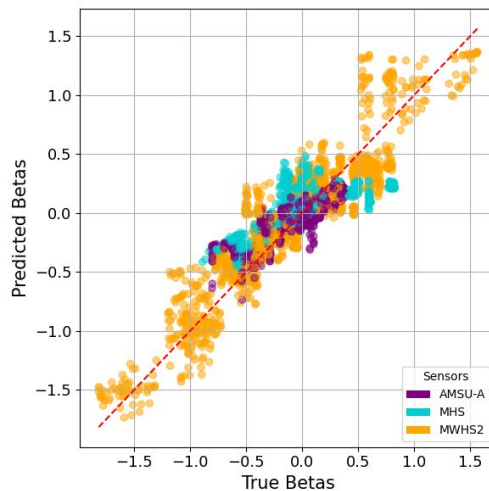
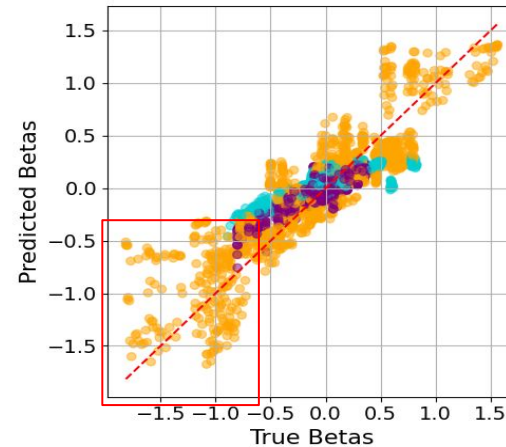
Remove AMSU-A



Remove MWHS2



Remove MHS



Next generation of instruments?

Computational costs

Computational costs of ML models

	Training time
Artificial Neural Network	3min
Random Forest	9min
AdaBoost	2min
Quantile Random Forest	30min



	Spin-up phase
VarBC	several days/a week



Next steps

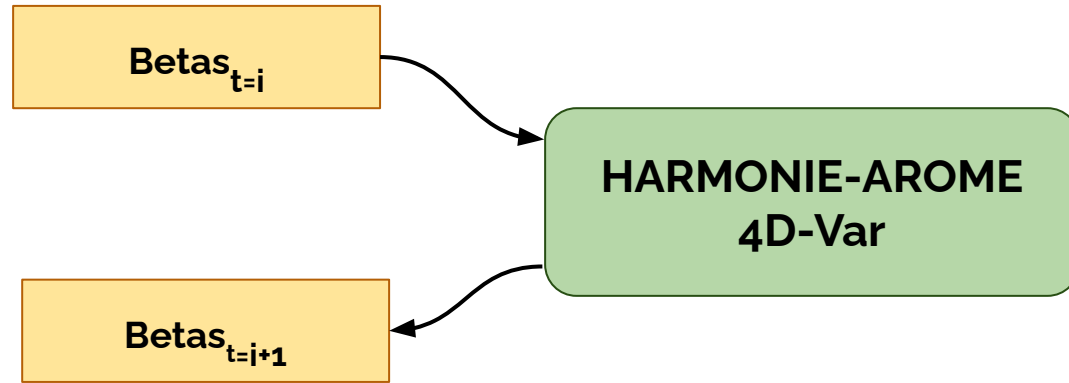
How does the traditional HARMONIE-AROME 4D-Var compare with the ML-based
HARMONIE-AROME 4D-Var?

HARMONIE-AROME 4D-Var

VarBC updates Betas inside NWP model:

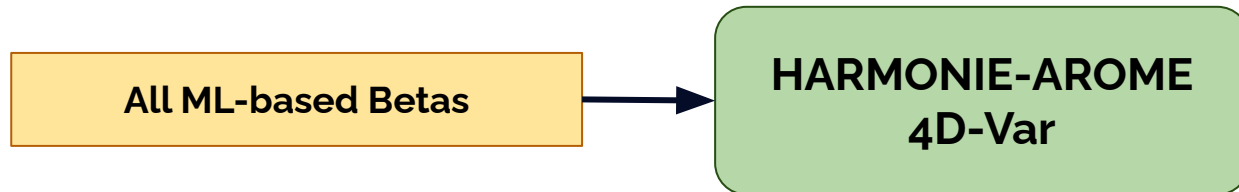
$$\text{VarBC at } t=i: b_i(x, \beta) = \beta_{0,i} + \beta_{1,i}p_{1,i}(x) + \beta_{2,i}p_{2,i}(x) \dots + \beta_{n,i}p_{n,i}(x)$$

Traditional:



$$\text{VarBC at } t=i+1: b_{i+1}(x, \beta) = \beta_{0,i+1} + \beta_{1,i+1}p_{1,i+1}(x) + \beta_{2,i+1}p_{2,i+1}(x) \dots + \beta_{n,i+1}p_{n,i+1}(x)$$

ML-based:



Discussion

Key takeaways & Looking ahead

Key takeaways & looking ahead

- Successfully emulate VarBC
- Dependency between instruments
- Relevance:
 - sub-km domains
 - new generations of instruments
- Next steps:
 - Assess the skill difference between the traditional VS ML H-A forecast

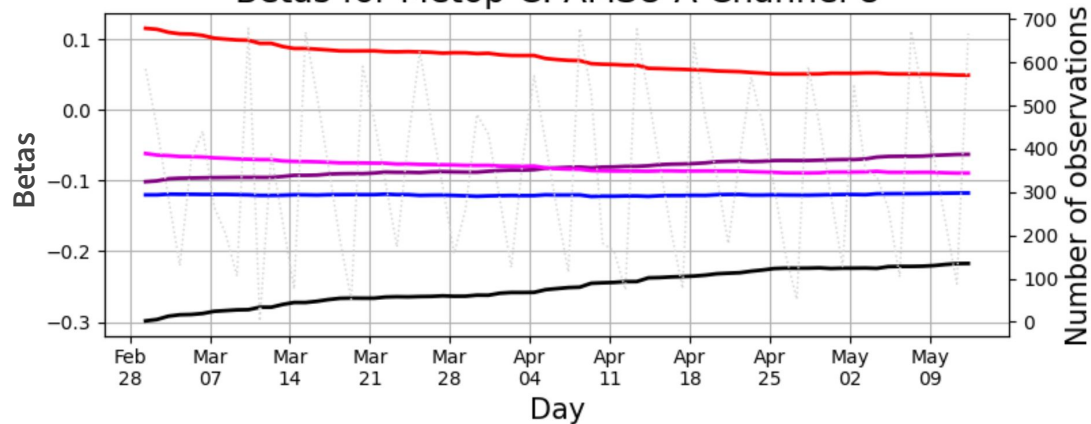


Thank you!

Outcome variable

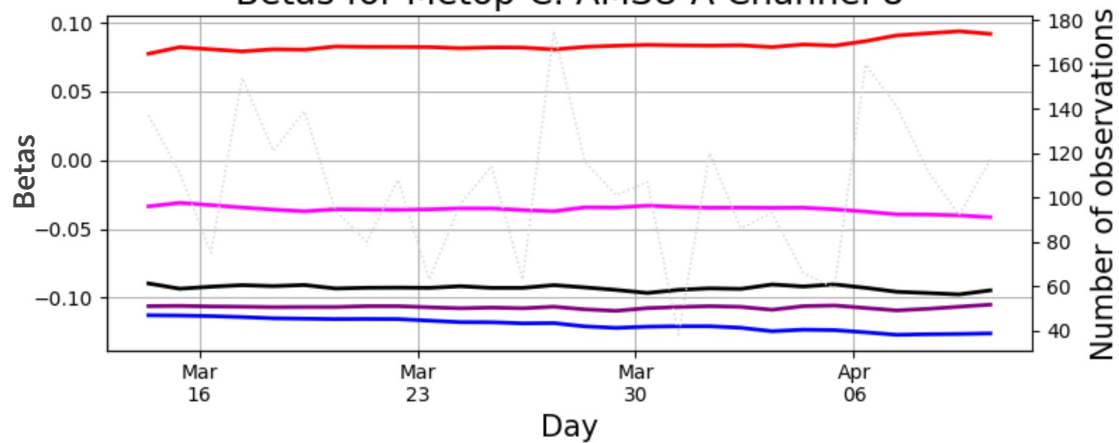
DINI domain (2023) 09:00 UTC

Betas for Metop-C: AMSU-A Channel 8



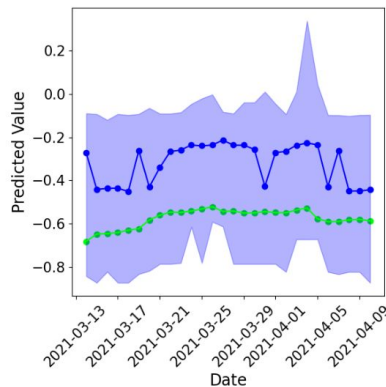
Dutch domain (2021) 09:00 UTC

Betas for Metop-C: AMSU-A Channel 8

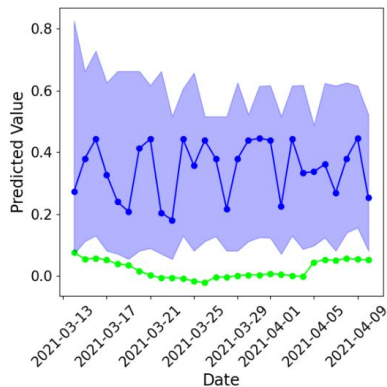


- constant
- 1000-300hPa thickness
- nadir view angle
- nadir view angle **2
- nadir view angle **3

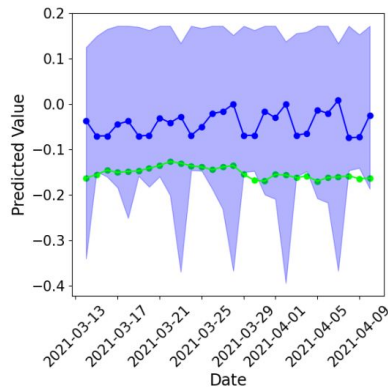
QRF time series: MetOp-B MHS Channel 5 around 09:00 UTC



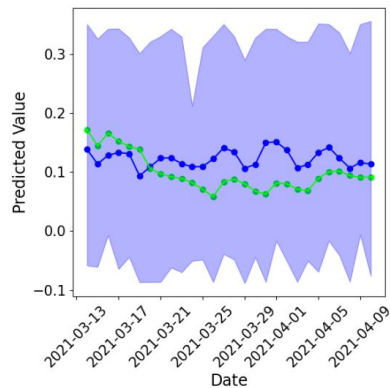
(a) Constant bias predictor



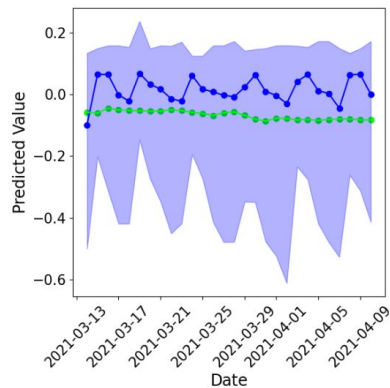
(b) 1000-300hPa thickness



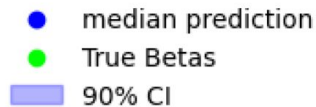
(c) Nadir view angle



(d) Nadir view angle**2



(e) Nadir view angle**3



→ no spin-up phase

4D-Var HARMONIE-AROME experiment



- **InitRun:** Setup the necessary directories and copy the system files needed
- **Build:** The build system generator CMake Compiles the binaries we need to run our experiment
- **MakeCycleInput:** Prepares boundaries and observations
- **Date:** Run assimilation and forecasts
 - **StartData**
 - **Analysis**
 - **Addsurf**
 - **AnUA**
 - **Fetch Assim_data**
 - **4DVscreen:** screening for 3h window
 - **4DVprolog:** prepares the initial conditions at the appropriate resolution for the forecasts used in minimization
 - **4DVloop1 and 4DVloop2**
 - **4DVminim:** produces the low resolution at the beginning of the observation window which is then transformed to the field at forecast resolution
 - **Blendhr:** complements 4DVminim's output with surface fields and the resulting field to act as the initial condition for the trajectory run 4DVtraj
 - **4DVtraj:** 2h forecast which will be used for main forecast
 - **AnSFC2:**
 - **Interpol_sst:** interpolate Sea Surface Temperature from the boundaries
 - **Addsurfcan**
 - **Canari:** Analysis software to produce quality control and analysis based on optimum interpolation (→ produces analysis and/or updated ODB)
 - **PostAnalysis:**
 - **Forecasting:** produces the forecast
- **PostProcessing:** Post process and archive the result