ITSC-25

Prototype for bias-correction of microwave radiance observations using ML methods

Alice Abramowicz
Isabel Monteiro
Kirien Whan
Irene Garcia Marti

Thanks to Erik Dedding

EUMETSAT

Koninklijk Nederlands Meteorologisch Instituut
Ministerie van Infrastructuur en Waterstaat

INTERNATIONAL ATOVS WORKING GROUP

Introduction

HARMONIE-AROME (H-A)

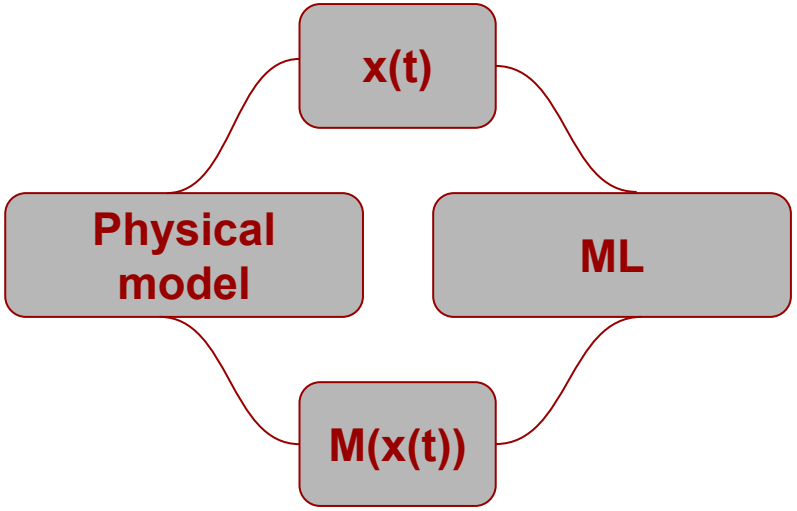
Observation biases

Variational Bias Correction (VarBC) linear model:

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

VarBC updates Betas every few hours inside NWP

→ Predict $\boldsymbol{\beta}$'s



Research questions

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

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 forecast quality compare between a regular H-A 4D-Var VS an ML-based H-A 4D-Var?

Metadata

- Date time
- Cycle

Time

- Satellite
- Sensor
- Channel

Data group

- Amount of observations
- Amount of bias predictors
- Bias predictor mean
- Bias predictor covariances

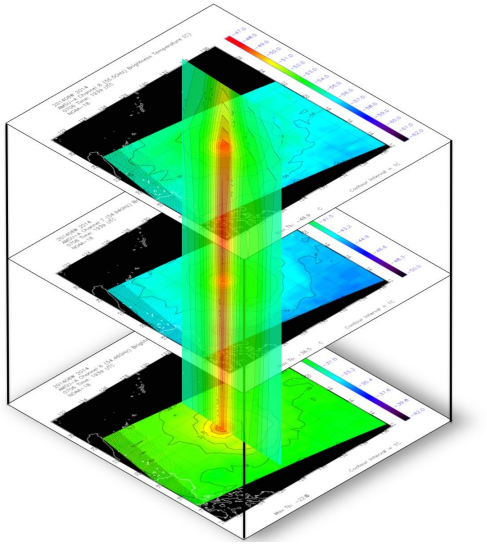
VarBC cycles

- First-guess departures
- Analysis departures
- Bias-corrected first-guess
- Observation value

CCMA

- Betas

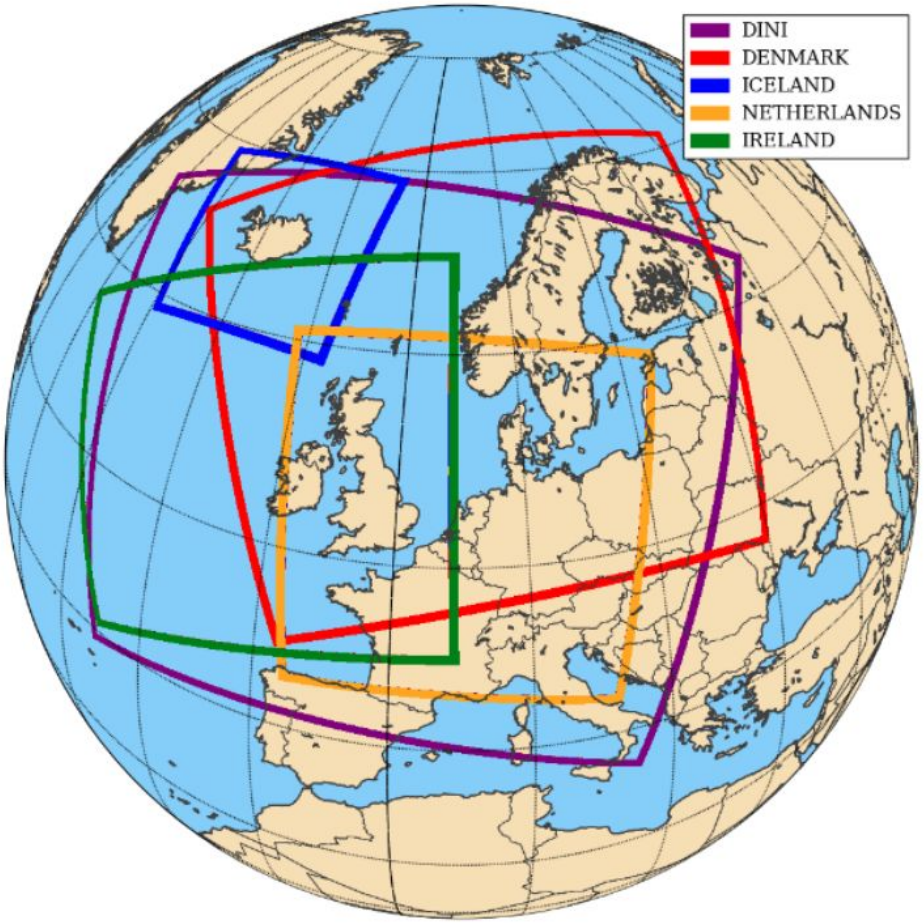
$$b(x,\beta) = \beta_0 + \beta_1 p_1(x) + \beta_2 p_2(x) + ... + \beta_n p_n(x)$$



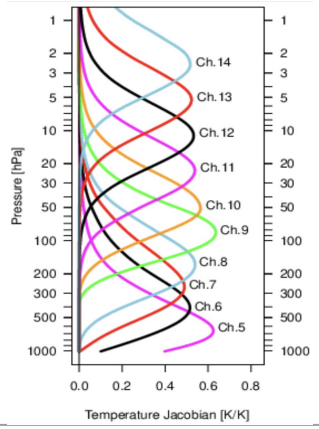
Research design

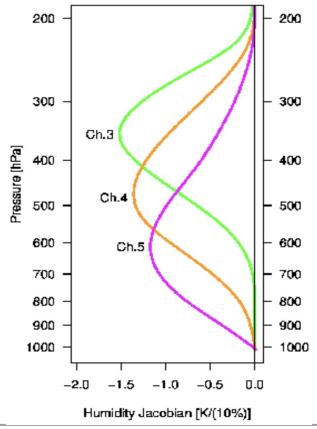
DINI domain (2023) for training

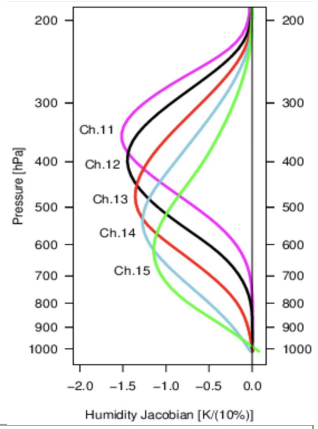
Dutch domain (2021) for testing



Data groups







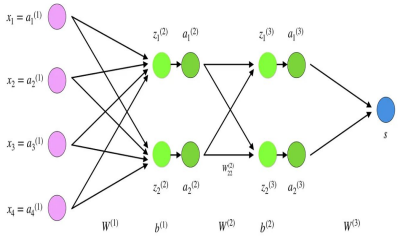
Sensors/ Satellites	**AMSU-A**	**MHS**	**MWHS2**
Metop-B	channels 7, 8, 9	channels 3, 4, 5	-
Metop-C	channels 6, 7, 8, 9	channels 3, 4, 5	-
NOAA18	channels 6, 7, 8	-	-
NOAA19	channels 6, 9	channels 4, 5	-
FY-3D	-	-	channels 4, 5, 6, 11, 12, 13, 14, 15

Figures from ECMWF observation monitoring package

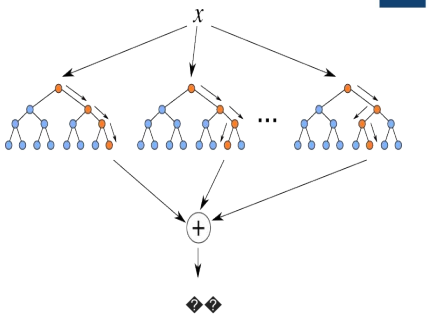
ML models

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

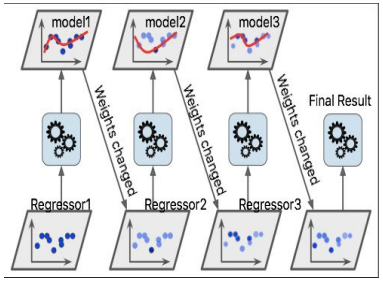
Results of ML models



NN

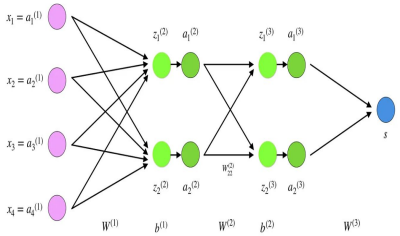


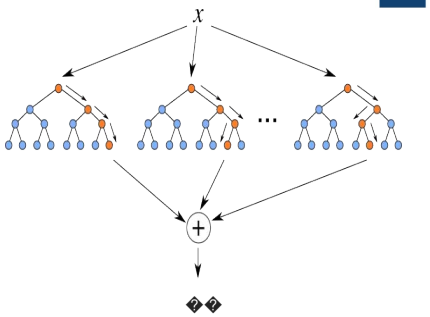
RF

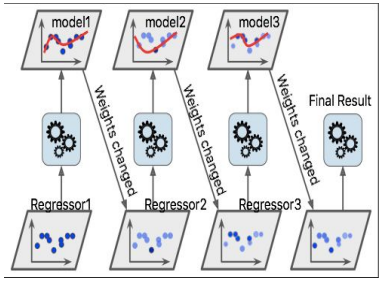


AB

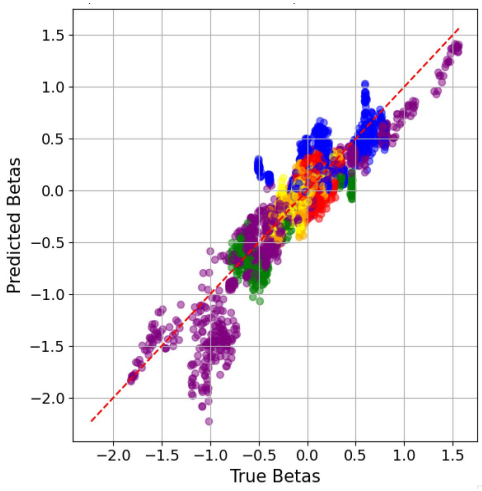
Results of ML models

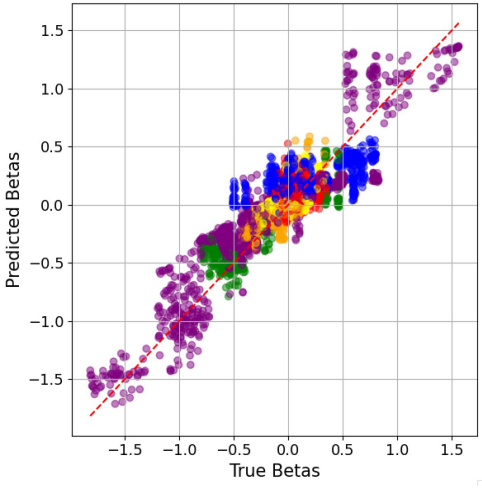


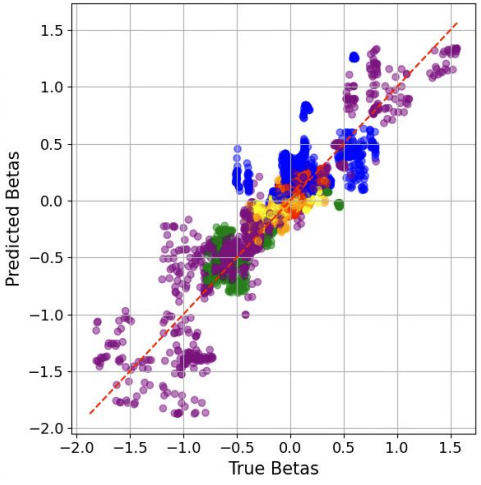


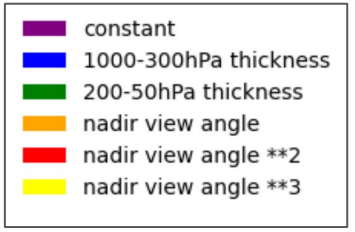


NN **RF** **AB**









Scoring metrics/Models	MSE	R^2
NN	0.0250	0.757
RF	0.0252	0.754
AB	0.0267	0.740

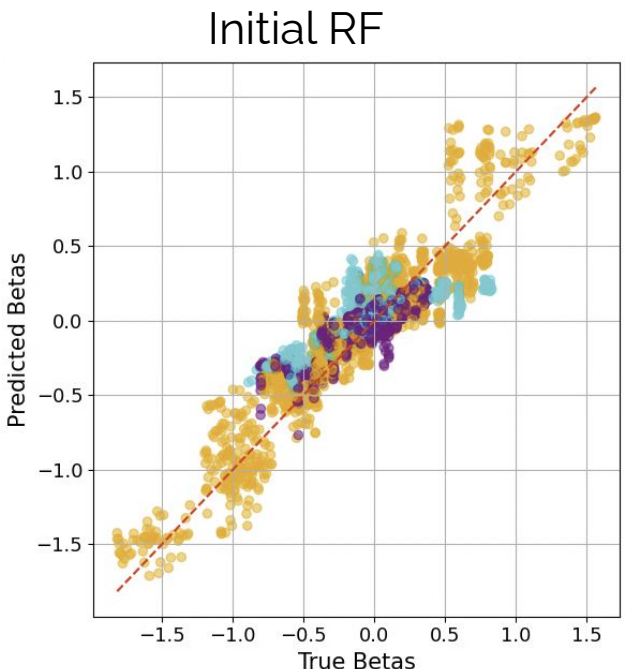
	Training time
NN	3min
RF	9min
AB	2min

On-demand sub-km domains?

Random Forest: Sensitivity analysis

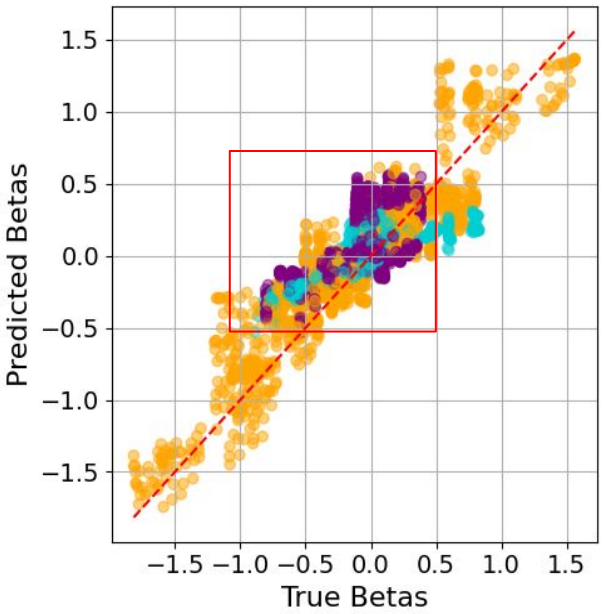
Is knowledge transferable from one instrument to another?

Remove a sensor from training

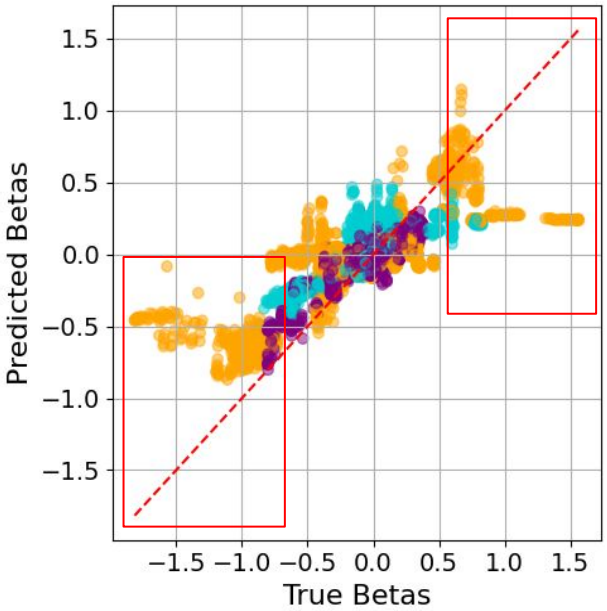


Remove a sensor from training

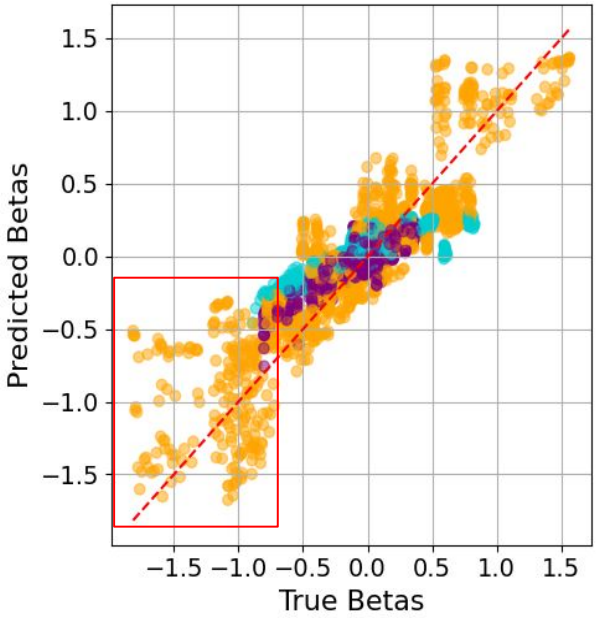
Remove AMSU-A



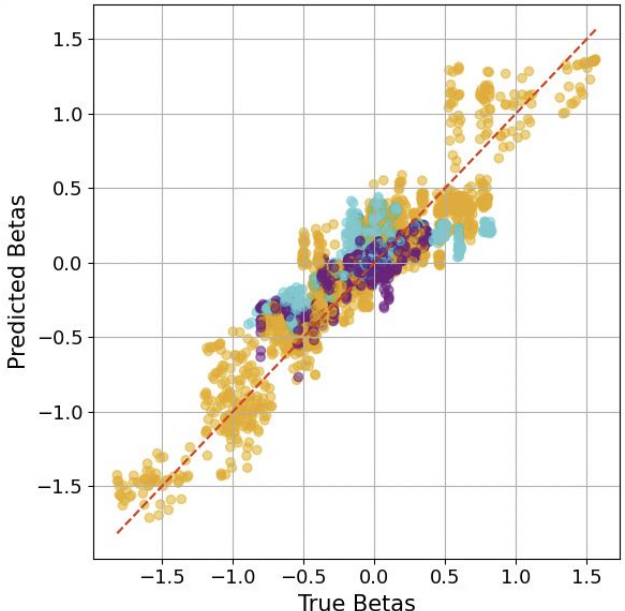
Remove MWHS2



Remove MHS



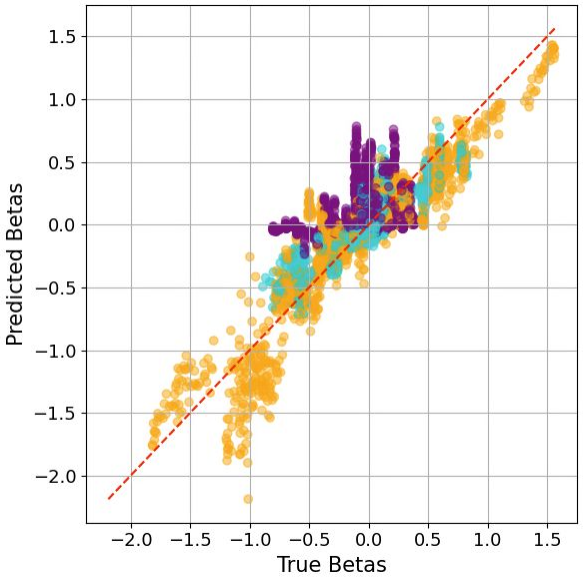
Initial RF



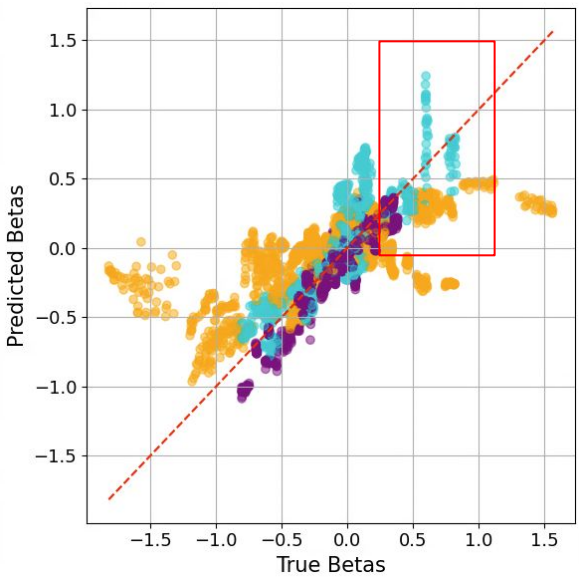
AMSU-A
MHS
MWHS2

Remove a sensor from training

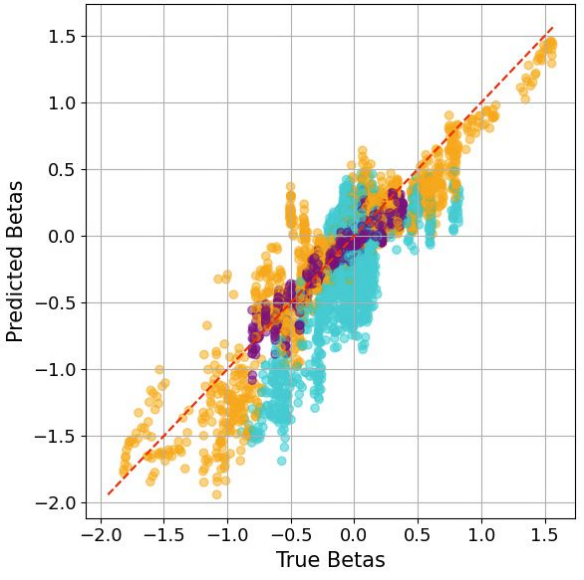
Remove AMSU-A



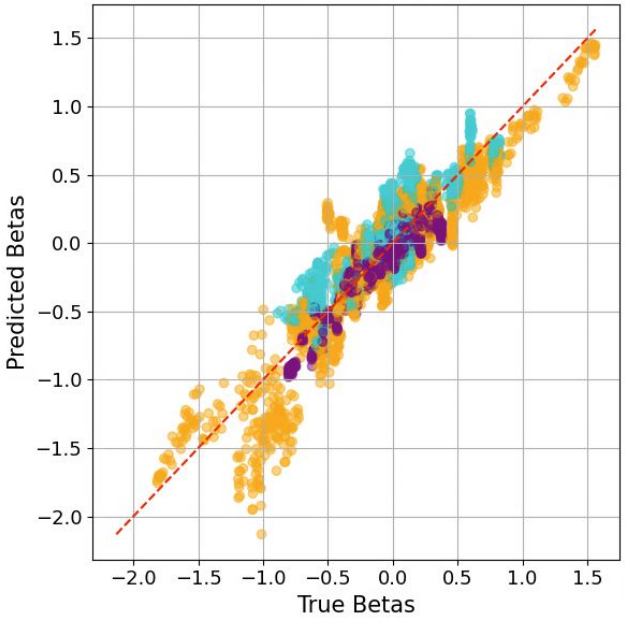
Remove MWHS2

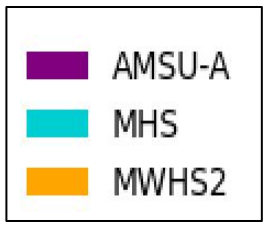


Remove MHS



Initial NN



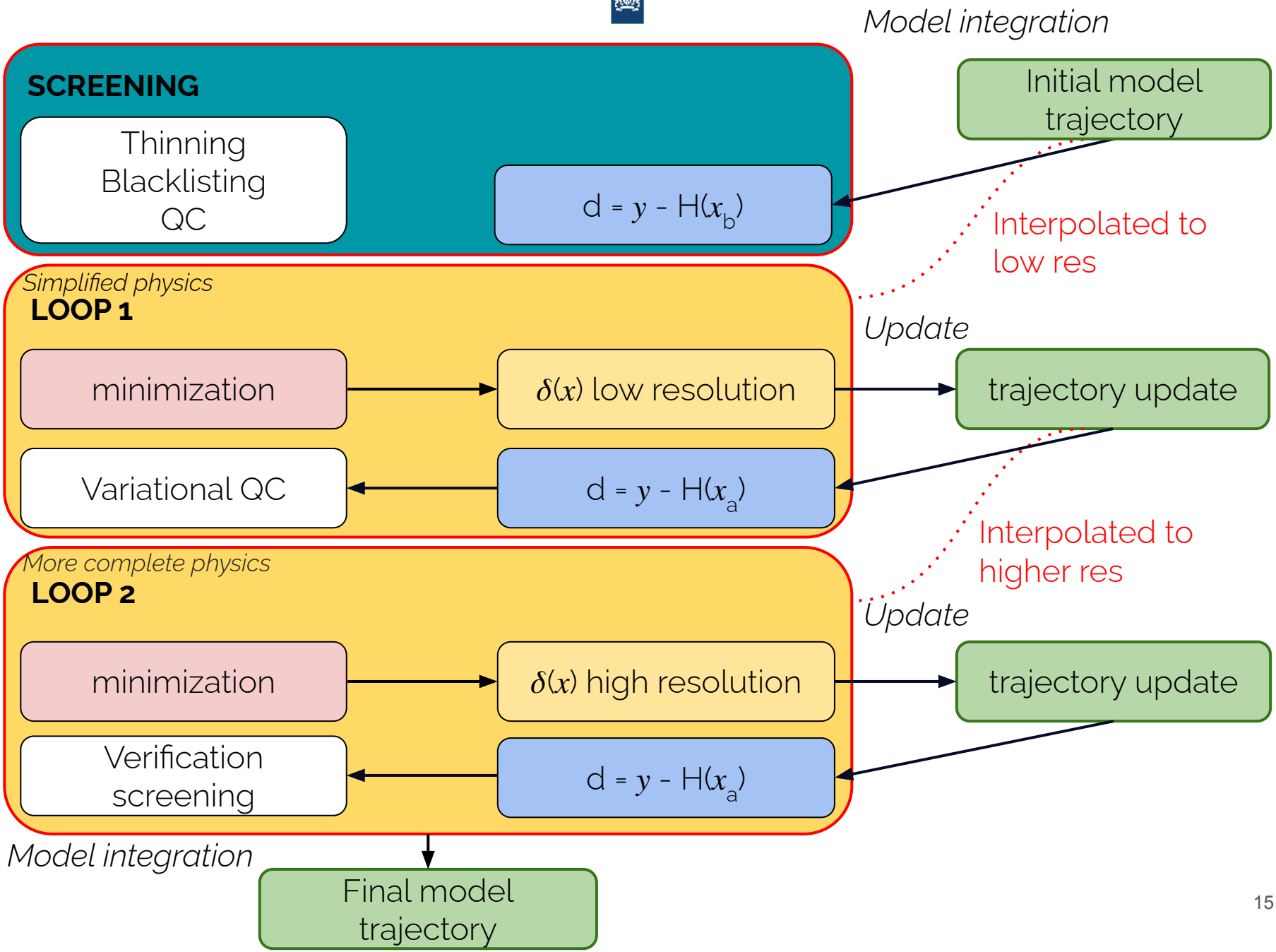


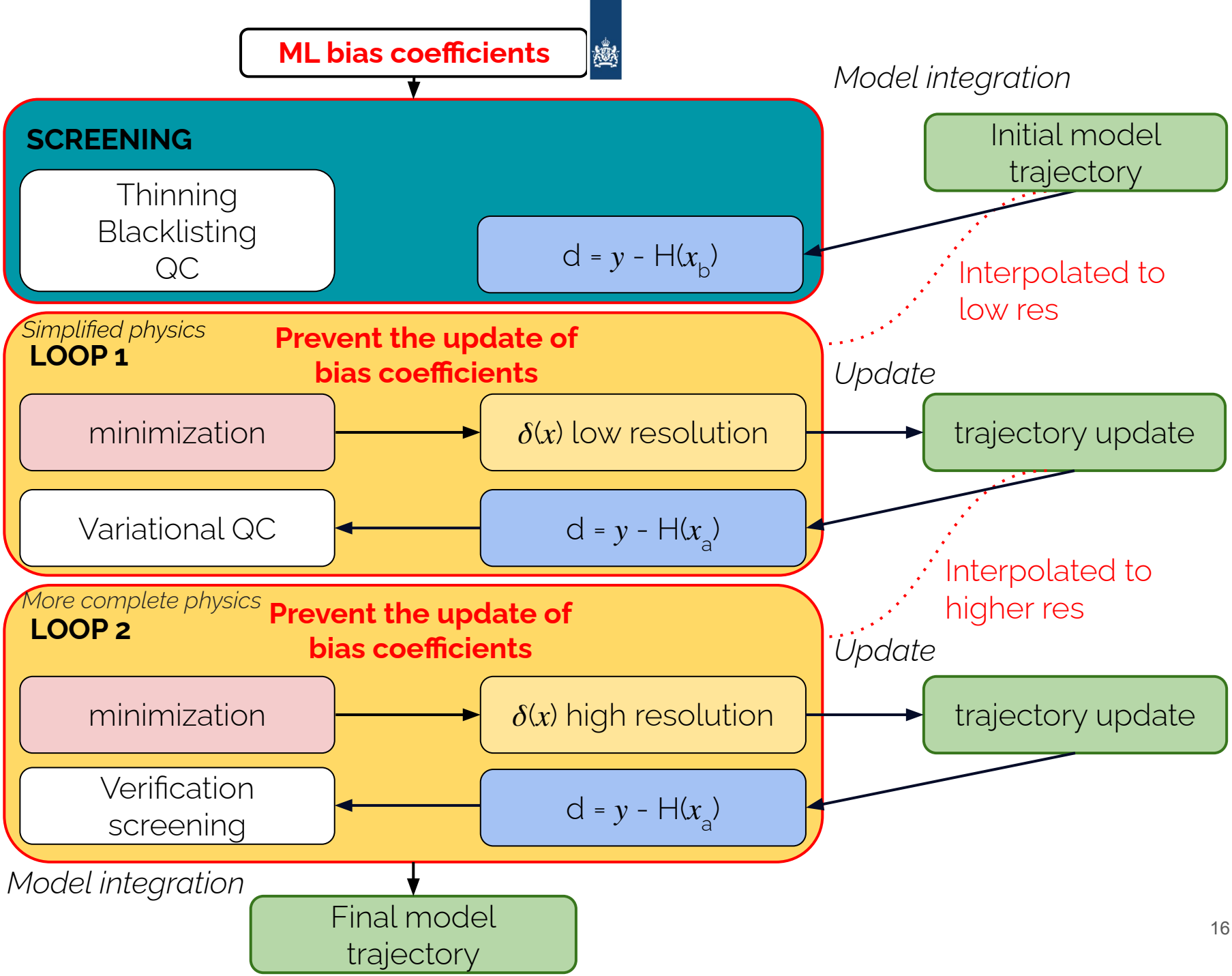
Next generation of instruments?

HARMONIE-AROME (H-A)

How does the forecast quality compare between a regular H-A 4D-Var VS an ML-based H-A 4D-Var?

4D-Var in H-A

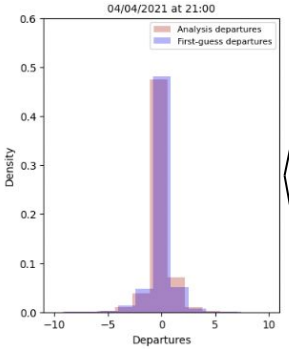

Model integration
SCREENING
Thinning
Blacklisting
QC
d = y - H(x_b)
Initial model trajectory
Interpolated to low res
Simplified physics
LOOP 1
Update
minimization
δ(x) low resolution
trajectory update
Variational QC
d = y - H(x_a)
Interpolated to higher res
More complete physics
LOOP 2
Update
minimization
δ(x) high resolution
trajectory update
Verification screening
d = y - H(x_a)
Model integration
Final model trajectory

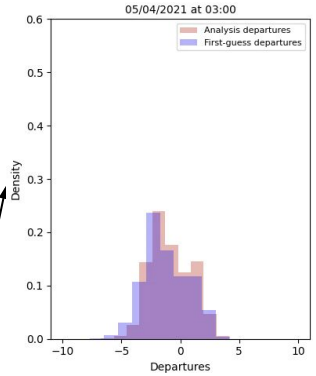

ML bias coefficients
Model integration
SCREENING
Thinning
Blacklisting
QC
d = y - H(x_b)
Initial model trajectory
Interpolated to low res
Simplified physics
LOOP 1
Prevent the update of bias coefficients
Update
minimization
δ(x) low resolution
trajectory update
Variational QC
d = y - H(x_a)
Interpolated to higher res
More complete physics
LOOP 2
Prevent the update of bias coefficients
Update
minimization
δ(x) high resolution
trajectory update
Verification screening
d = y - H(x_a)
Model integration
Final model trajectory

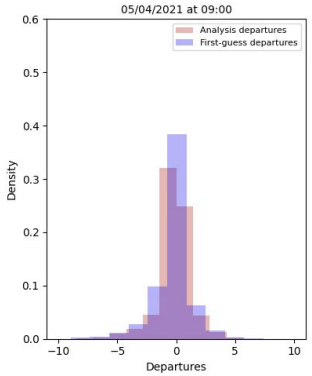
ODB preliminary results

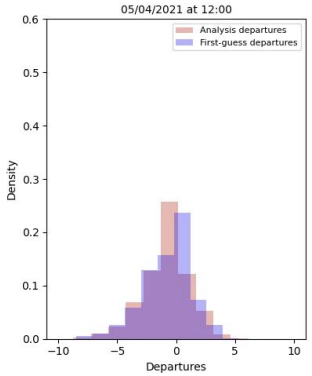
Active traditional exp

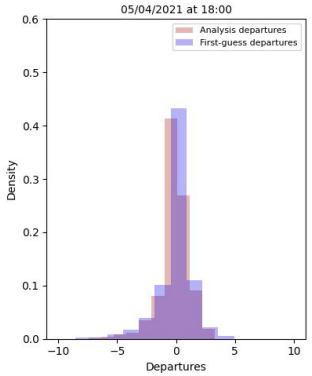
Passive warm start

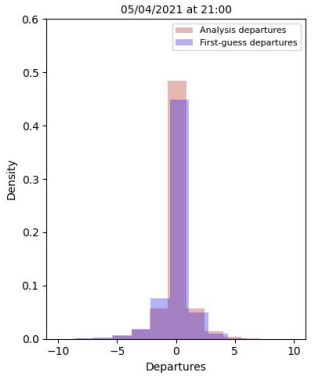




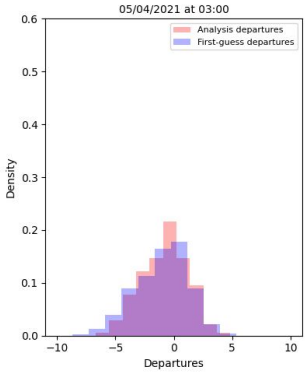


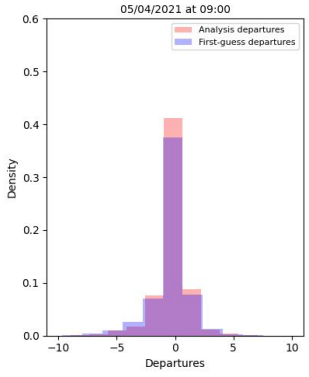


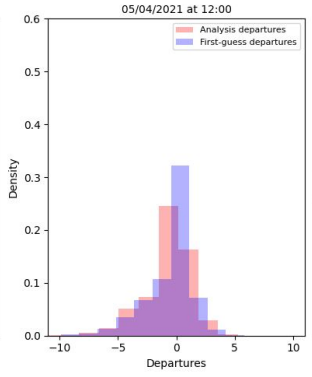


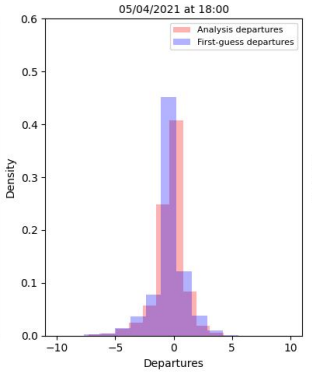


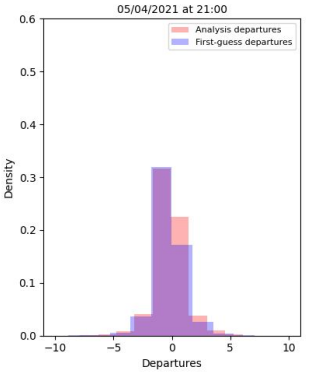
Active ML exp











MWHS2

AMSU-A MHS

AMSU-A MHS

AMSU-A MHS

AMSU-A MHS

Data Assimilation in KNMI

Erik Dedding (17p.01): Towards a full exploitation of satellite radiance information using transformed retrievals in HARMONIE-AROME 4D-Var

Key takeaways & Looking ahead

Discussion

- Emulate VarBC
- Dependency between instruments
- Relevance:
  - on demand sub-km domains
  - new generations of instruments
- Next steps:
  - Assess the skill difference between the traditional VS ML forecast
  - Different training data
  - Offline → online?

Thank you!