Utilisation du Machine learning en sciences du climat :

Améliorer la connaissance des ondes internes de gravité

Rencontres annuelles du GDR

Défis théoriques pour les sciences du climat

5 juin 2023

Aurélie FISCHER
(LPSM, Université Paris Cité)

Laboratoire de probabilités statistique & modélisation

Collaboration with :

Milena CORCOS (LMD)

Sothea HAS (LMD & LPSM)

François LOTT (LMD)

Riwal PLOUGONVEN (LMD)

Albert HERTZOG (LMD)

Aurélien PODGLAJEN (LMD)

Raj RANI (LMD)

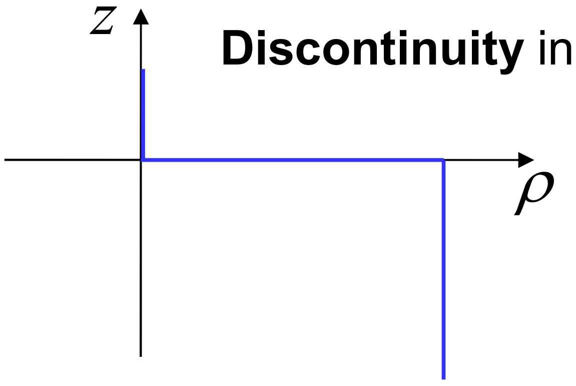
IMPT project (Institut des Mathématiques pour la Planète Terre, CNRS)

DataWave : Collaborative Gravity Wave Research

VESRI project (Virtual Earth System Research Institute, Schmidt Futures)

Gravity waves

Waves due to gravity and to a contrast in density ρ in the vertical (denser fluid below...)

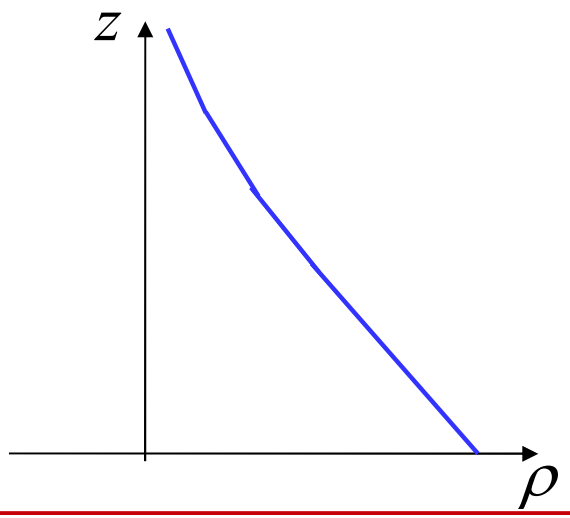


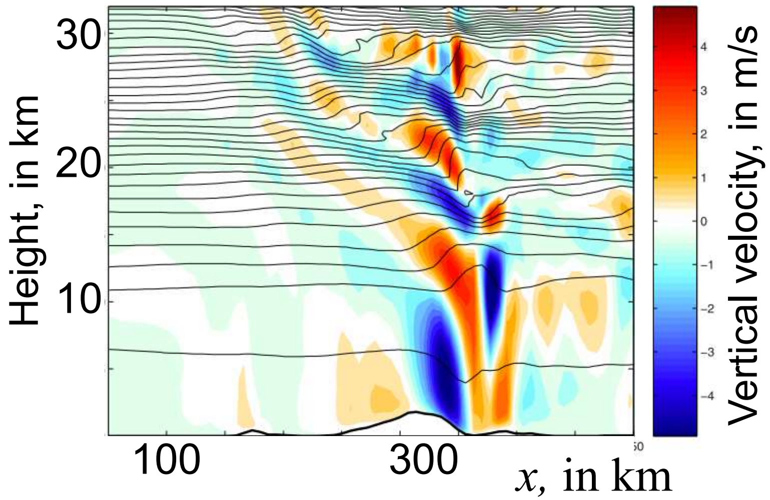
Discontinuity in density →

Periods starting from a few seconds

Continuous decrease in density →

Internal gravity waves





Periods from 5min to ~1-2 days

Gravity waves

- Air displaced in the vertical :
  - due to mountains (orographic waves)
  - by jet streaks, fronts, convection

- Impact for the general circulation : vertical transfer of momentum from the troposphere to the stratosphere and mesosphere.

- Important role in daily weather + long-term climate fluctuations.

→ One of the wave families forcing the Quasi-Biennal Oscillation (QBO)

- Quantity of interest = GW momentum fluxes

accelerate or decelerate flow higher up = change in air momentum

Need of parameterizations

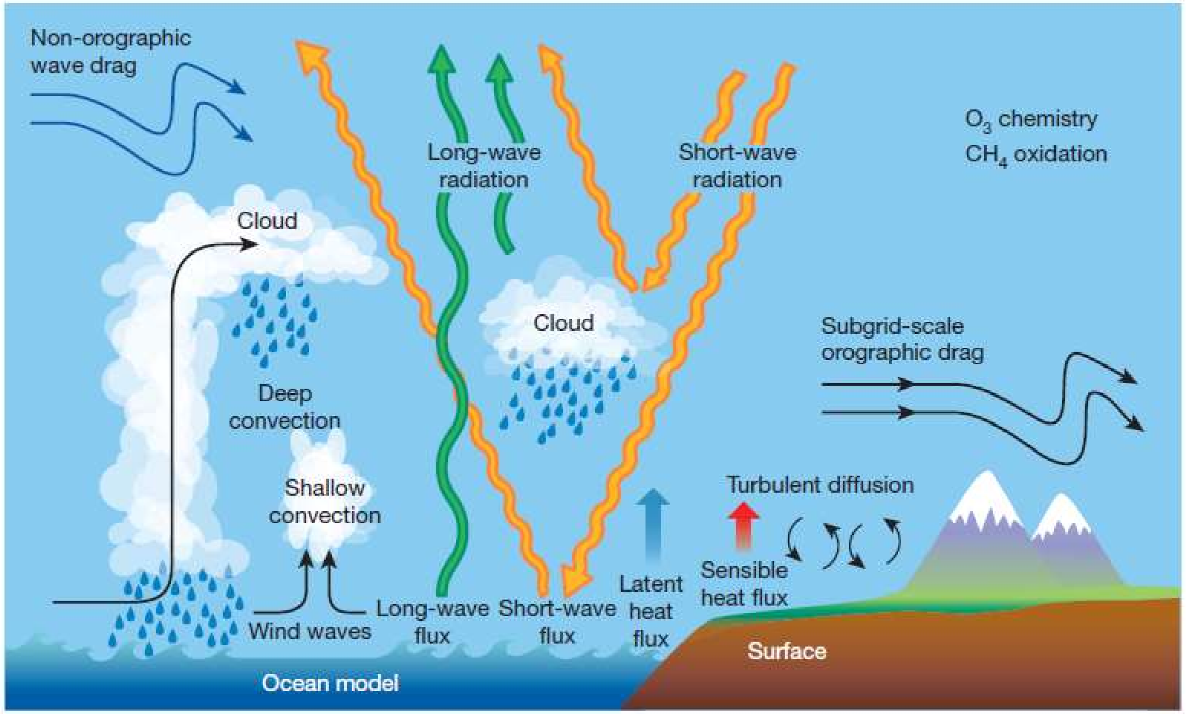


Figure 2 | Physical processes of importance to weather prediction. These are not explicitly resolved in current NWP models but they are represented via parameterizations describing their contributions to the resolved scales in terms of mass, momentum and heat transfers.

Bauer et al 2015

Need of parameterizations

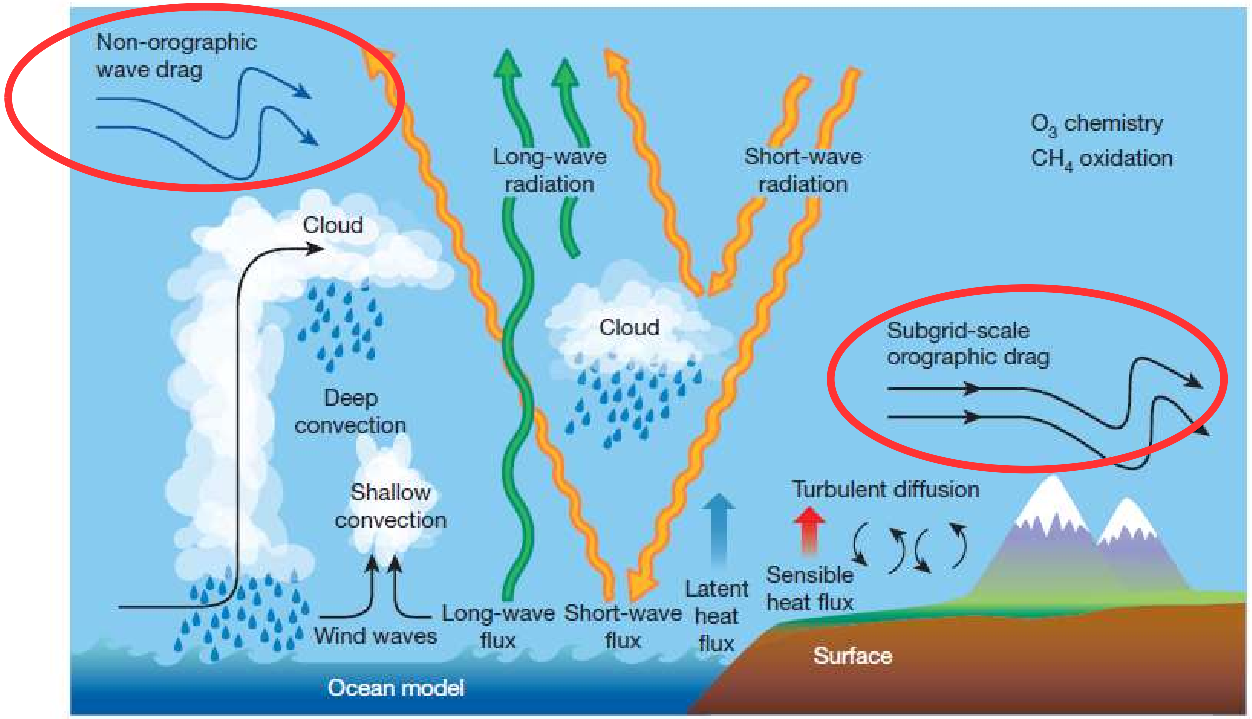


Figure 2 | Physical processes of importance to weather prediction. These are not explicitly resolved in current NWP models but they are represented via parameterizations describing their contributions to the resolved scales in terms of mass, momentum and heat transfers.

Bauer et al 2015

GW are subgrid scale, unresolved processes
→ necessary to parameterize GW

Parameterizations

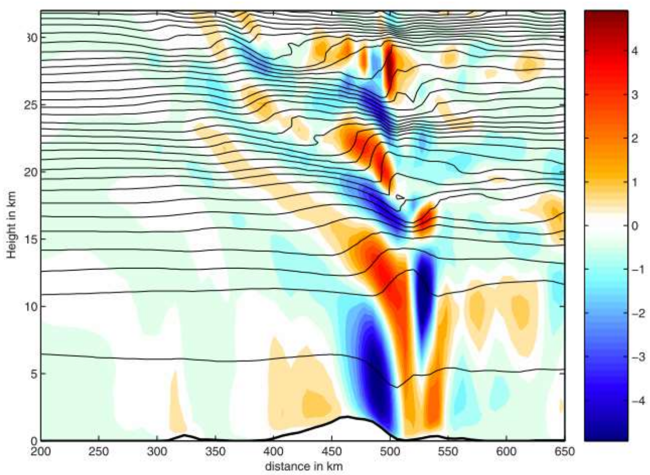
In climate and weather models, workaround to represent subgrid-scale = unresolved processes.

→ Even if unable to include GW in the model, using the knowledge of their actions, represent their impacts on the resolved flow.

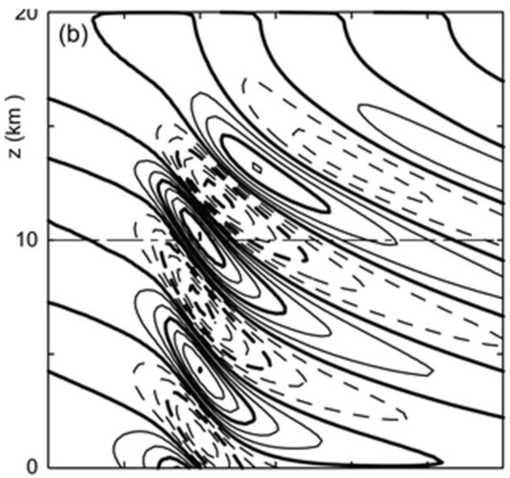
- Universal : in any location, relies on resolved physical variables, not location-specific.
- Physics-based : ideally, should be based on physical laws, as the equations of motions are for the resolved flow.

Example : orographic waves

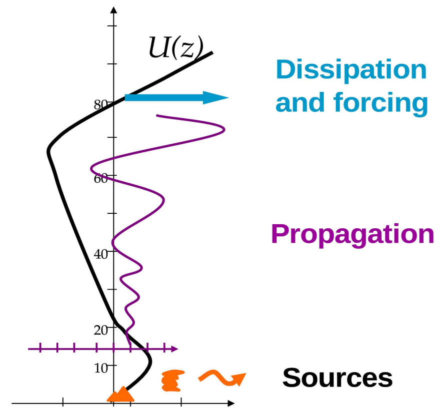
Real / realistic



Idealized, analytic



Parameterization



Parameterizations

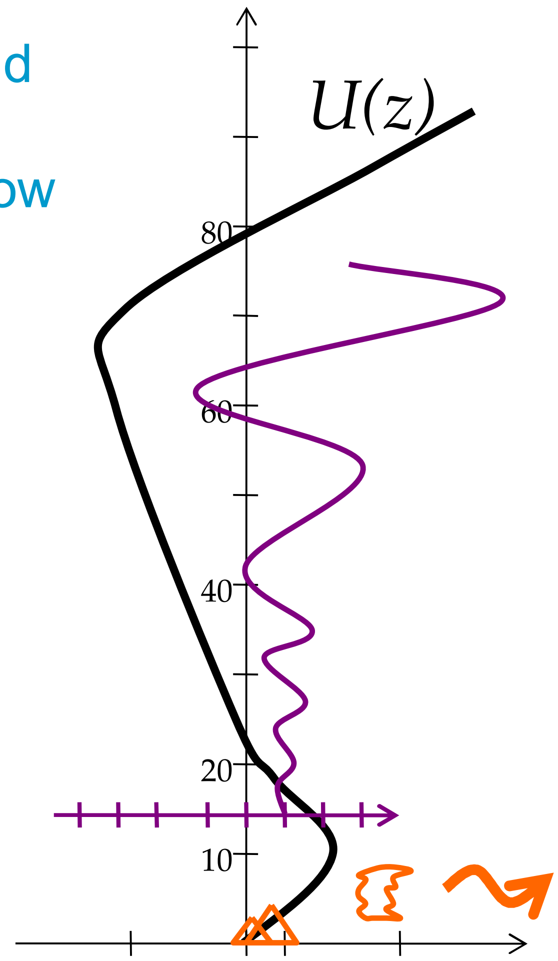
GW dynamics simplified to minimum :
- source specification
- vertical propagation
- dissipation and forcing of the flow

Much research targeting sources.
- Fairly arbitrary, poorly constrained.
-Parameters conveniently tuned.
→ Errors, uncertainty

Dissipation and forcing of the background flow

Propagation

Sources



Some examples of machine learning applications

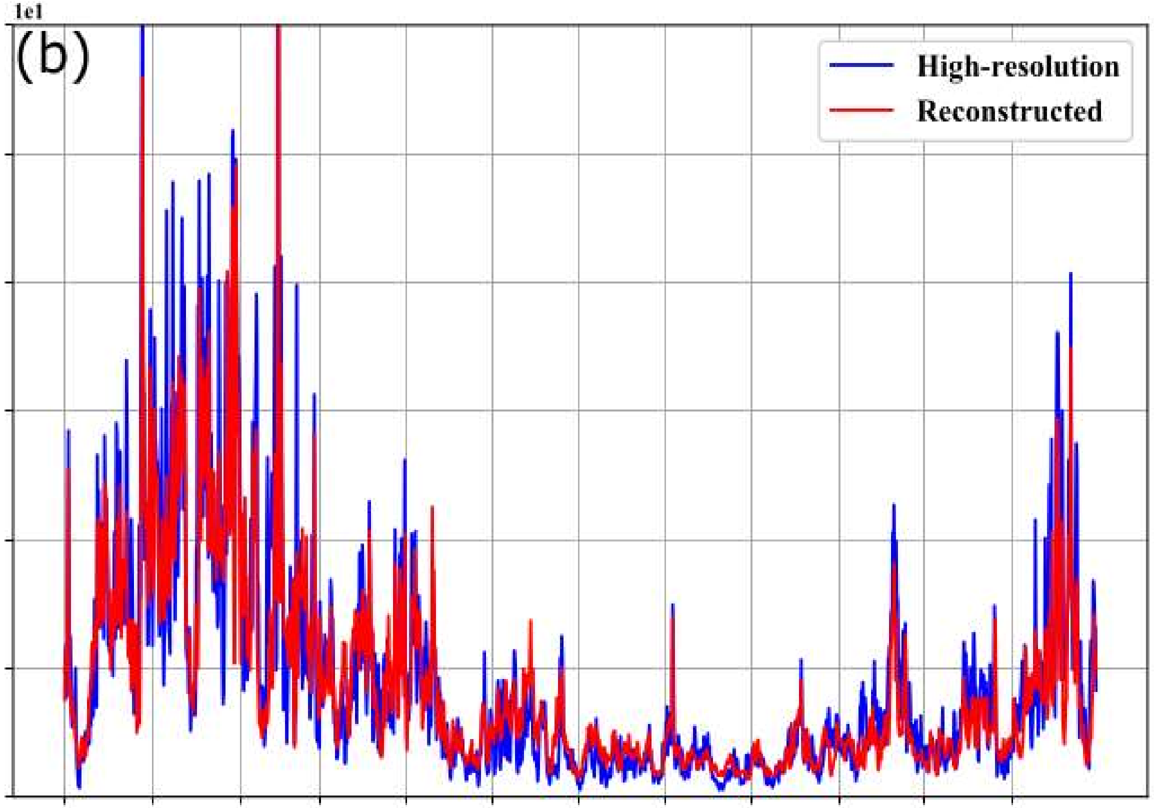
- Emulate parameterizations to save computing time
- « Metamodel » from higher resolution simulations
- Data-driven parameterizations built using machine learning
- Relate large scale flow to local observations (gravity waves momentum fluxes)

Metamodeling

Amiramjadi et al, 2020

ECMWF moderate resolution → GW information (target = model information)
ECMWF low resolution → Large-scale flow

Random Forests to reconstruct GWs



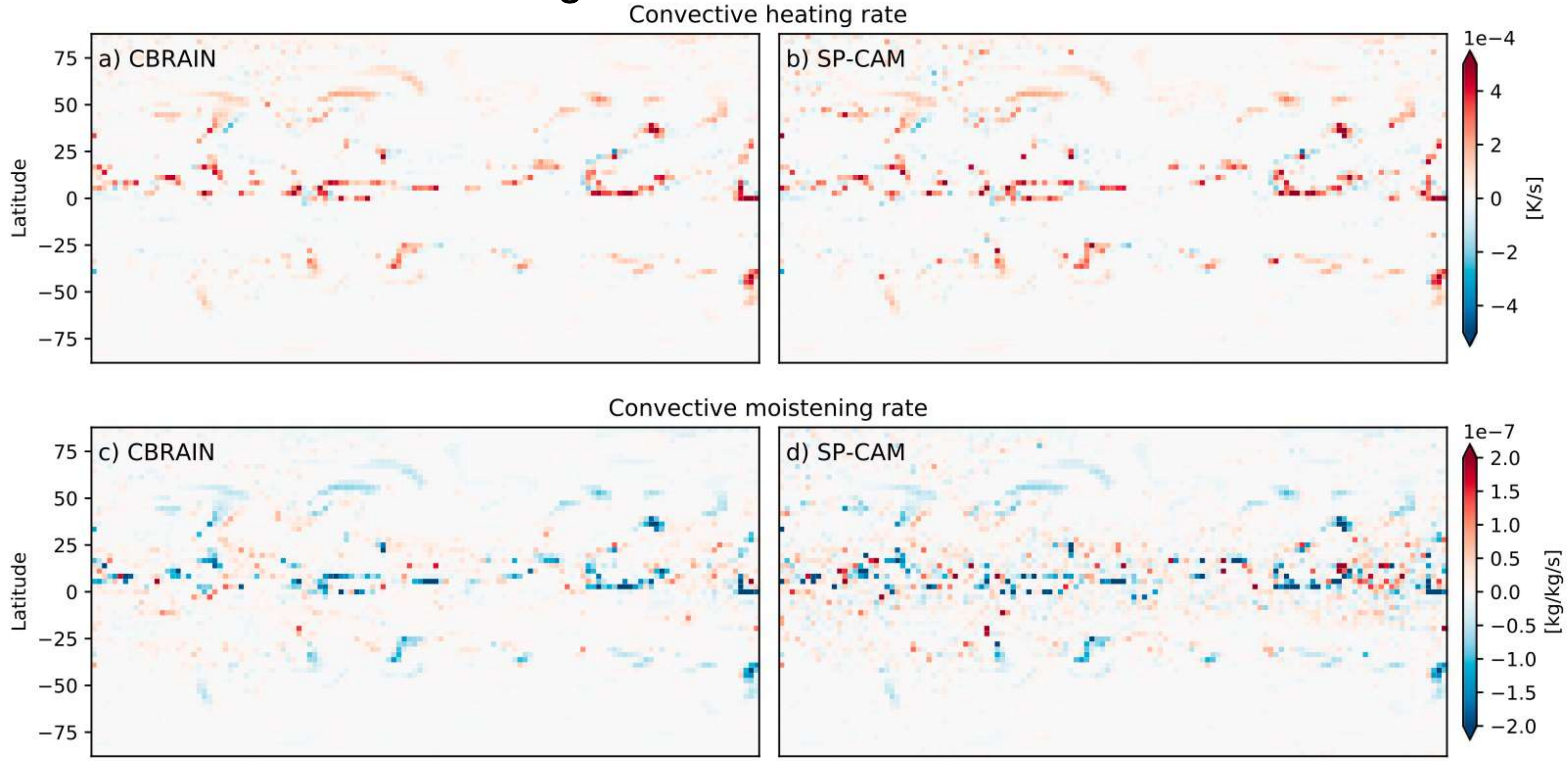
Data-driven parameterizations using ML

For processes that are *resolvable* (gravity waves, clouds), short high-resolution simulations provide information.

→ Capture relationship between resolved and unresolved processes

NN after training

Hi-res simulations



Example for convection (clouds)

Gentine et al 2018

ECMWF data → Information on the large-scale flow

Stratospheric balloons → Accurate observations on gravity waves

ML to reconstruct observed GW momentum fluxes from large-scale.

Observations : stratospheric balloons

Superpressure balloons, 11 and 13 m in diameter

Flight levels : ~18 and ~20 km

Lifetime : 2 to 3 months

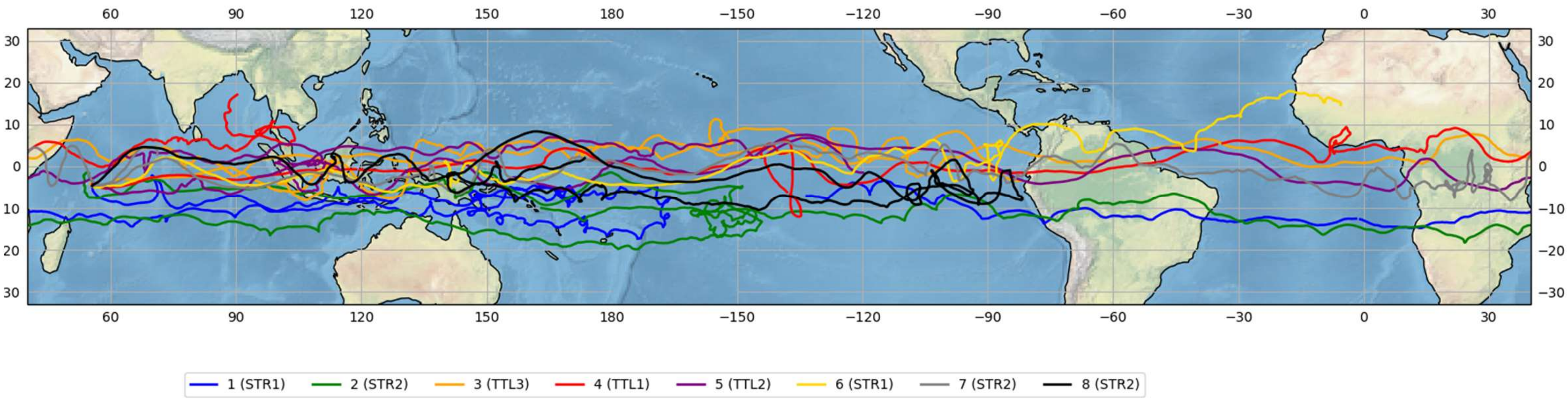
Observations : stratospheric balloons

Stratéole 2

French-US project

2019 campaign C0 :
8 balloons,
November 2019 to February 2020, along the tropics,
680 days of measurements.

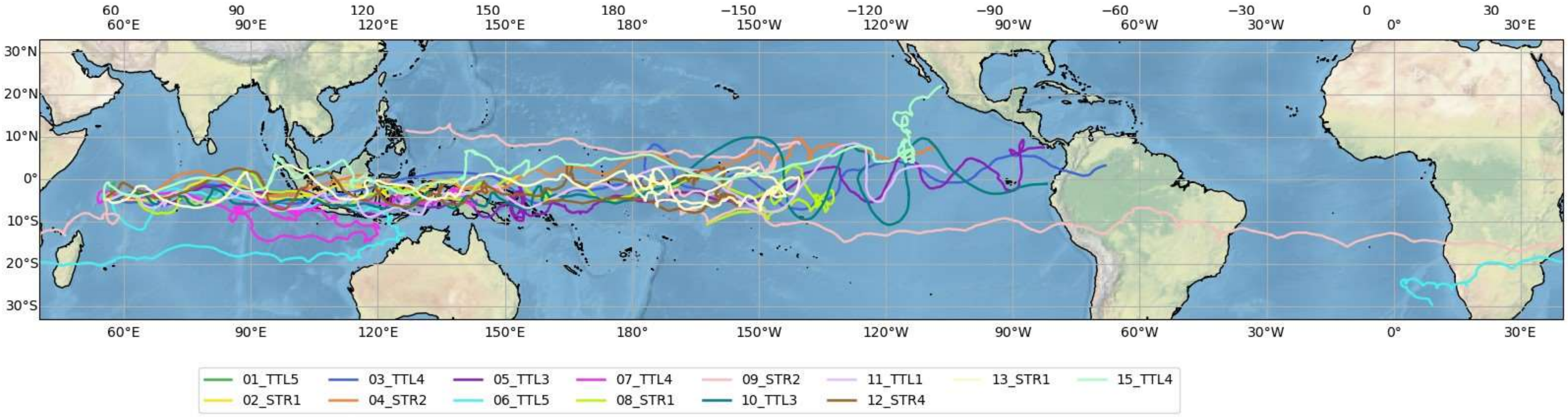
Data registered = 30s observations of position (wind), in-situ air pressure + temperature.



Observations : stratospheric balloons

Stratéole 2

2021 campaign C1 :
17 balloons
October 2021 to January 2022.



Observations : stratospheric balloons

Unique and valuable source of information on GW :

Quasi-Lagrangian behavior.
→ Direct access to the intrinsic frequency of the GW, thanks to in situ measurements
(not remote sensing, from temperature data : incertainty).

→ accurate estimate of key quantities, using wavelet analysis : momentum fluxes (Hertzog et al., 2012).

→ Large spatial cover since the balloons drift.

Some remarks on the balloons

- Direction :

Surface wind near the Equator has direction East→West
At balloon altitude, winds alternate between westerlies and easterlies, period of ~28 months (QBO) → East
+ reversal
+ 2 balloons → West (further from the Equator, in South hemisphere)

- Oscillation 3min : high frequency GW → period 15min

Explanative variables from Reanalysis ERA5

5th generation of the European Reanalysis

Reanalysis : historical observations + numerical models
→ weather/climate datasets

ERA5 : state-of-the-art global atmospheric reanalysis dataset (hourly from 1 to 137 vertical levels).

Extracted variables : precipitation, pressure, wind and temperature profile (67 vertical levels) at 5 x 5 horizontal grid points of 1° x 1° (100km) resolution.

Question : Which large-scale variables are most informative about GW ?

Explanative variables from Reanalysis ERA5

Inputs :

Temperature : temp

Zonal and meridional wind : u and v

4 levels : 19, 9, 2km and surface level (0km).

log surface pressure: lnsp

Solar zenith angle : sza

Precipitation: tp, tp_{mean}, tp_{sd}

Targets : two types of absolute, eastward and westward GWMFs

- High frequency waves (HF) : period 15mn to 1h.
- Wide frequency waves (WF) : period 15mn to 1 day.

Statistical learning setting

We observe a sample $D_n = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$ from a generic random pair (X, Y) taking its values in $\mathbb{R}^d \times \mathbb{R}$.

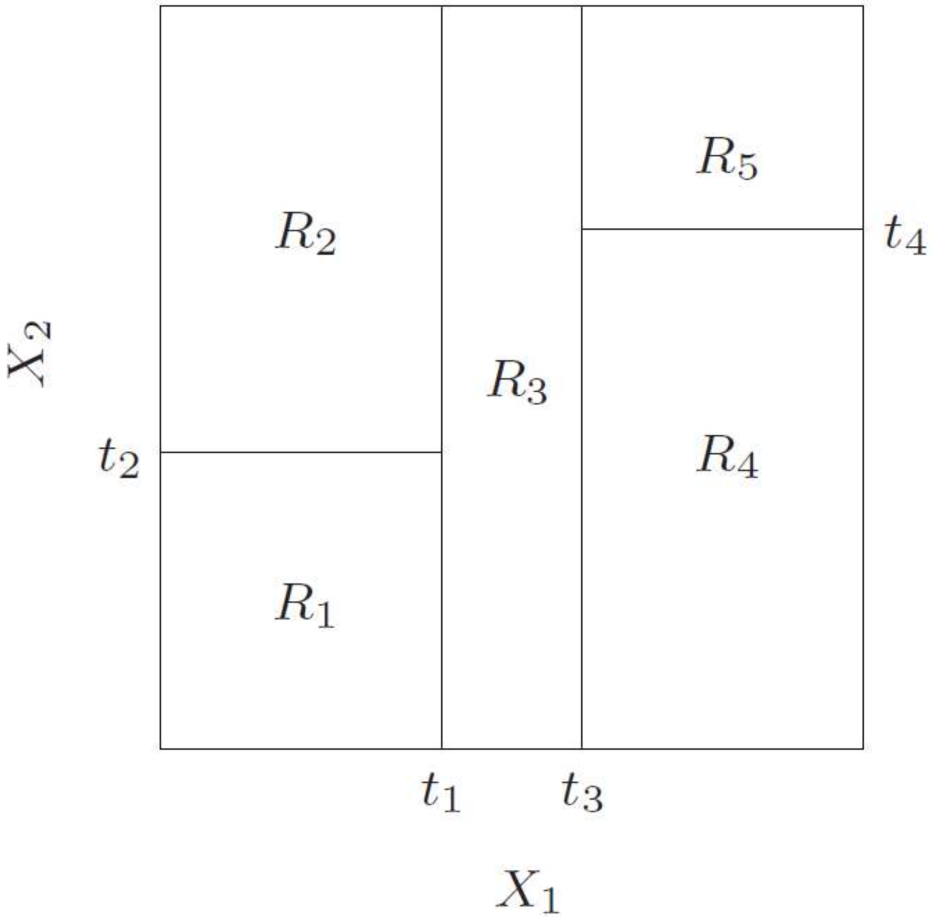
Explain the variable of interest / output Y using the different features or inputs $X = (X^1, \ldots, X^d)$.

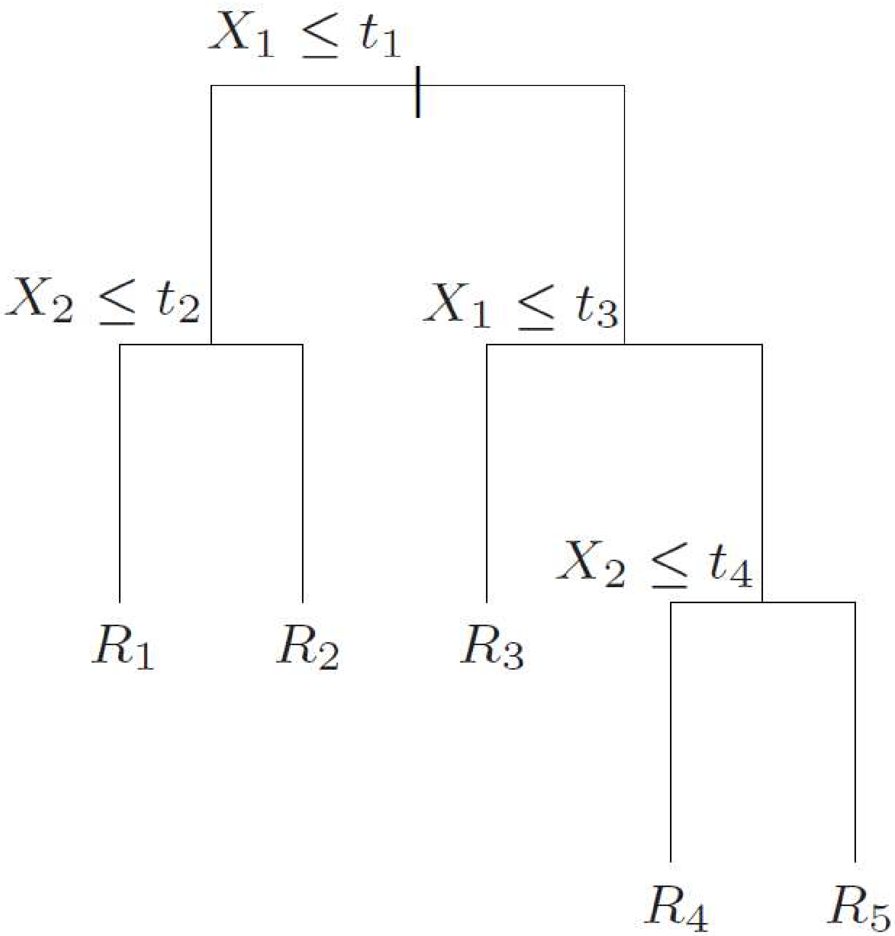
In other words, based on the data D_n, we look for some function g such that $Y = g(X)$.

For new x, predict associated y by $g(x)$.

Statistical methods

Nonparametric tree-based methods : combining several regression trees





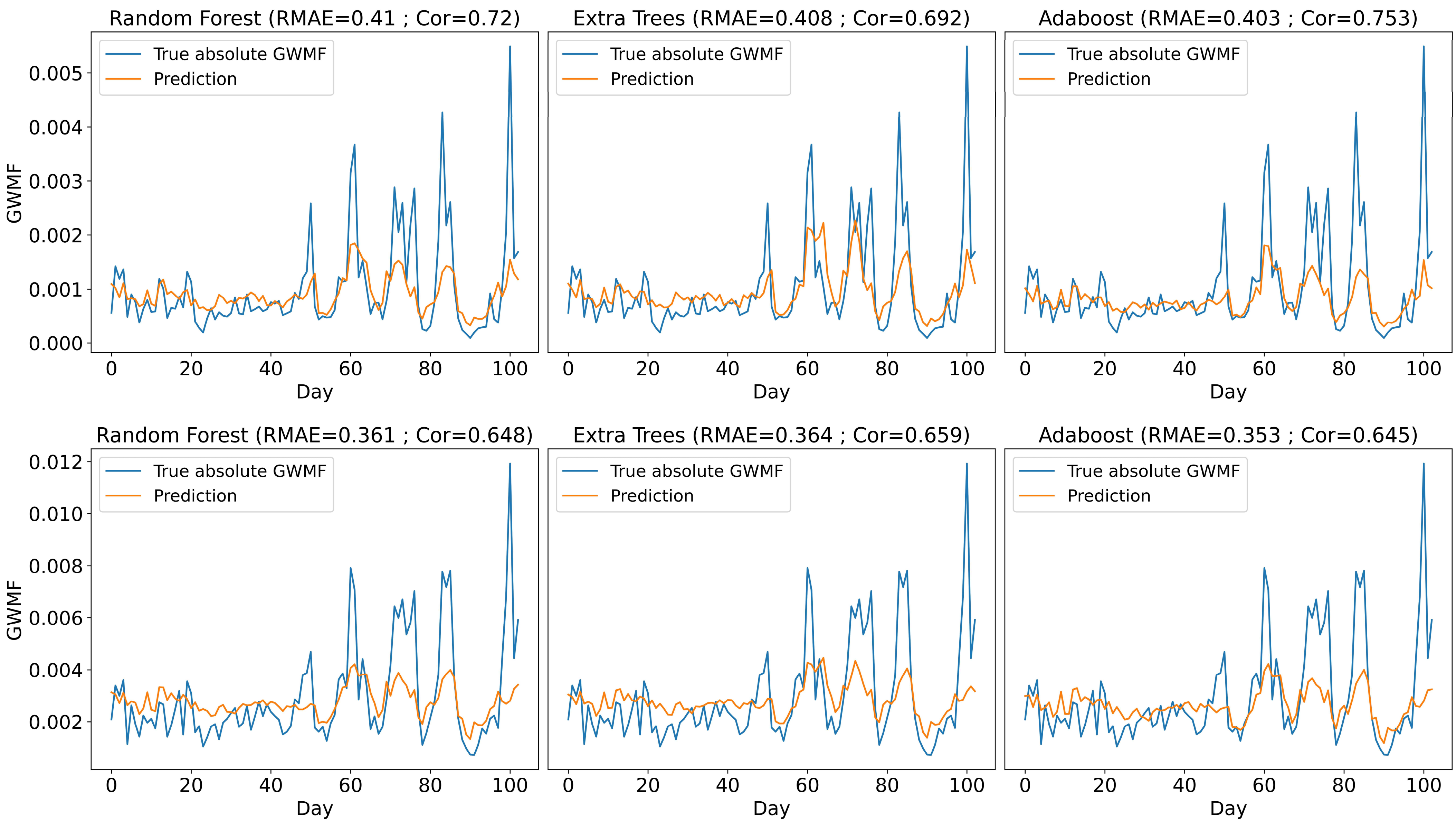
Statistical methods

- Random forests : bootstrap samples (resampling) = bagging
  + subset of variables, at random

- ExtraTrees : initial sample, subset of split thresholds, at random

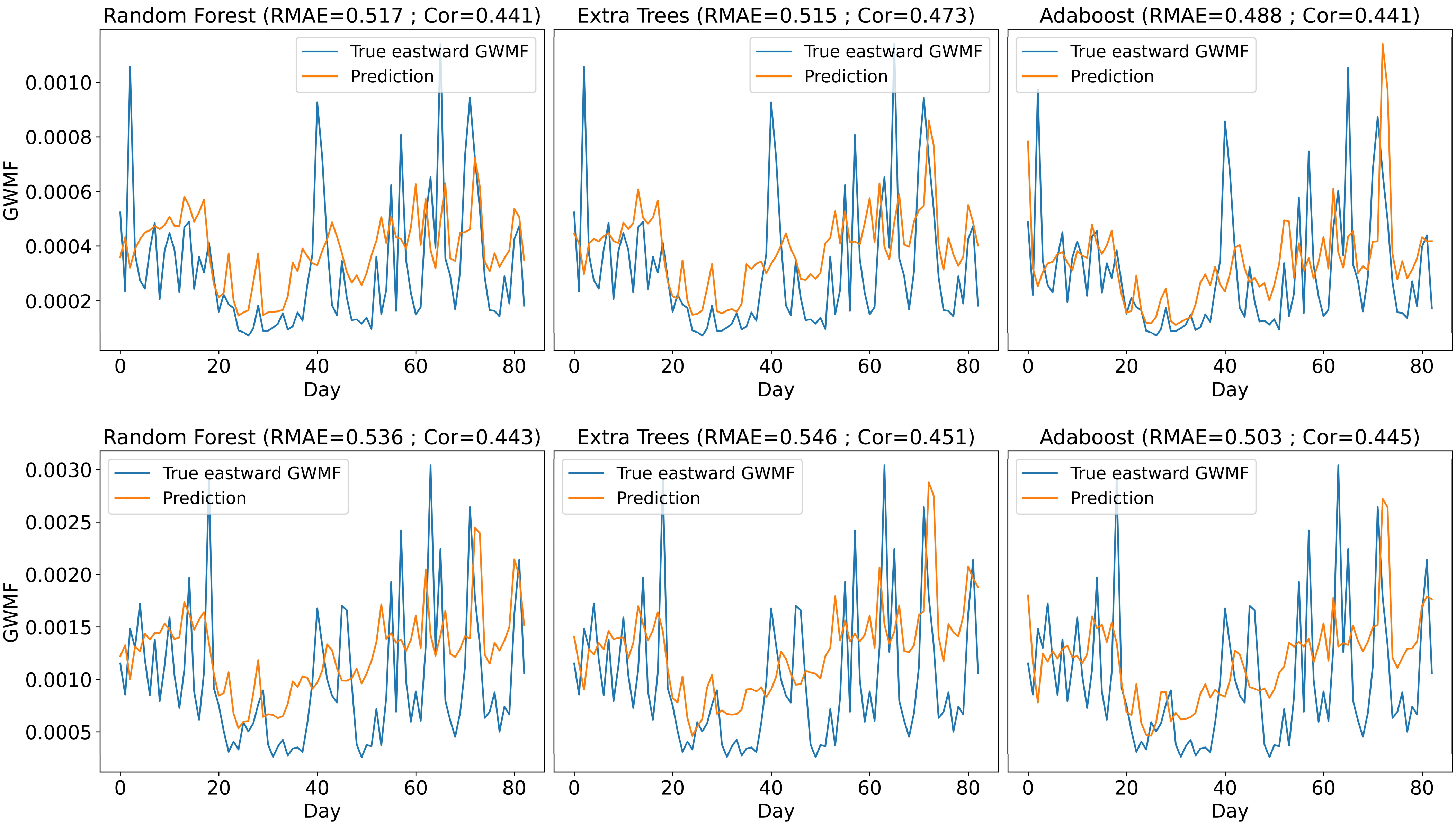
- Boosting : weak estimators, iterative, based on weights

Absolute GWMF : Balloon 2



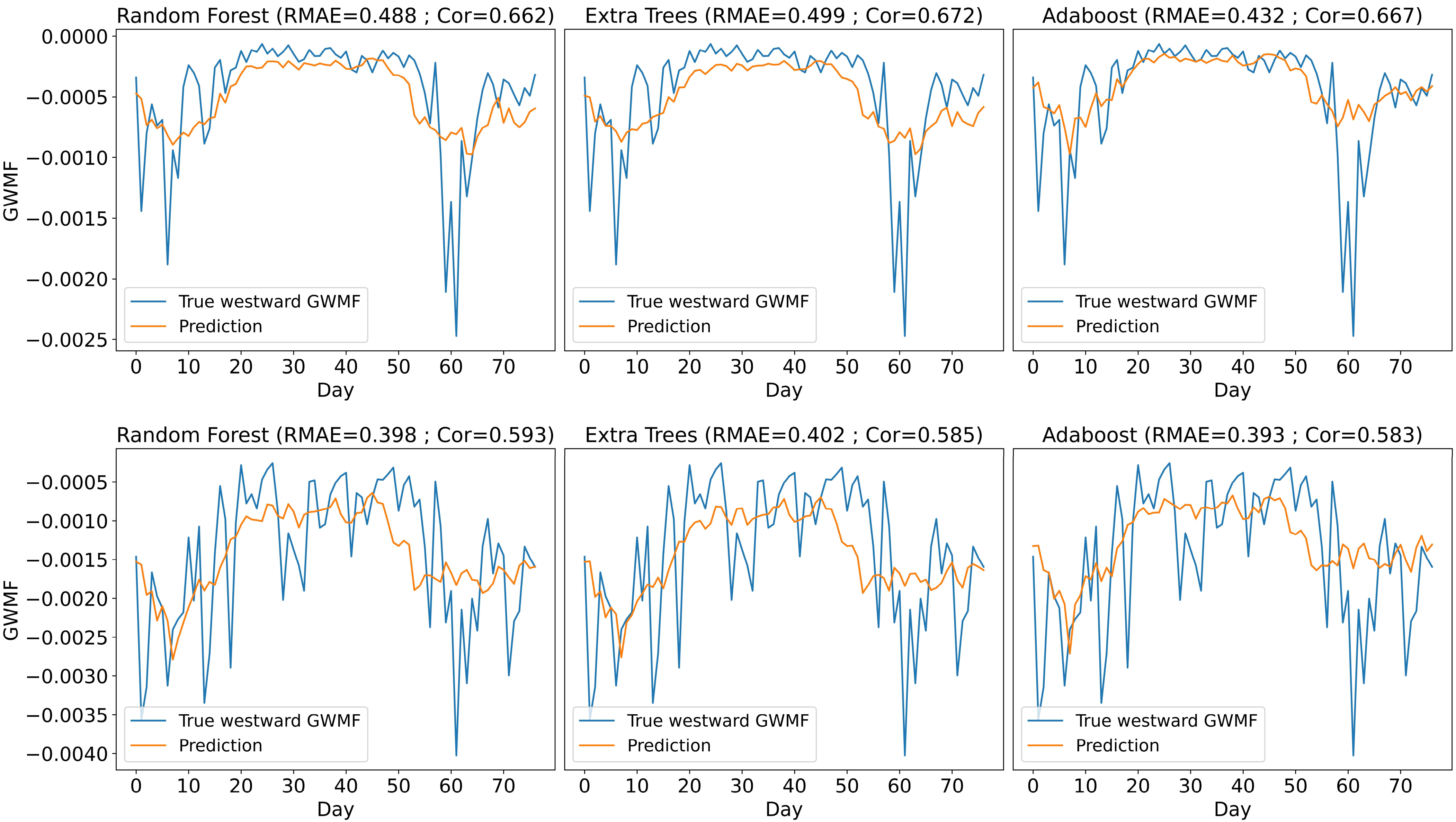
Predicted and actual absolute GWMF of HF (top) and WF (bottom) waves in 24h resolution

Eastward GWMF : Balloon 7



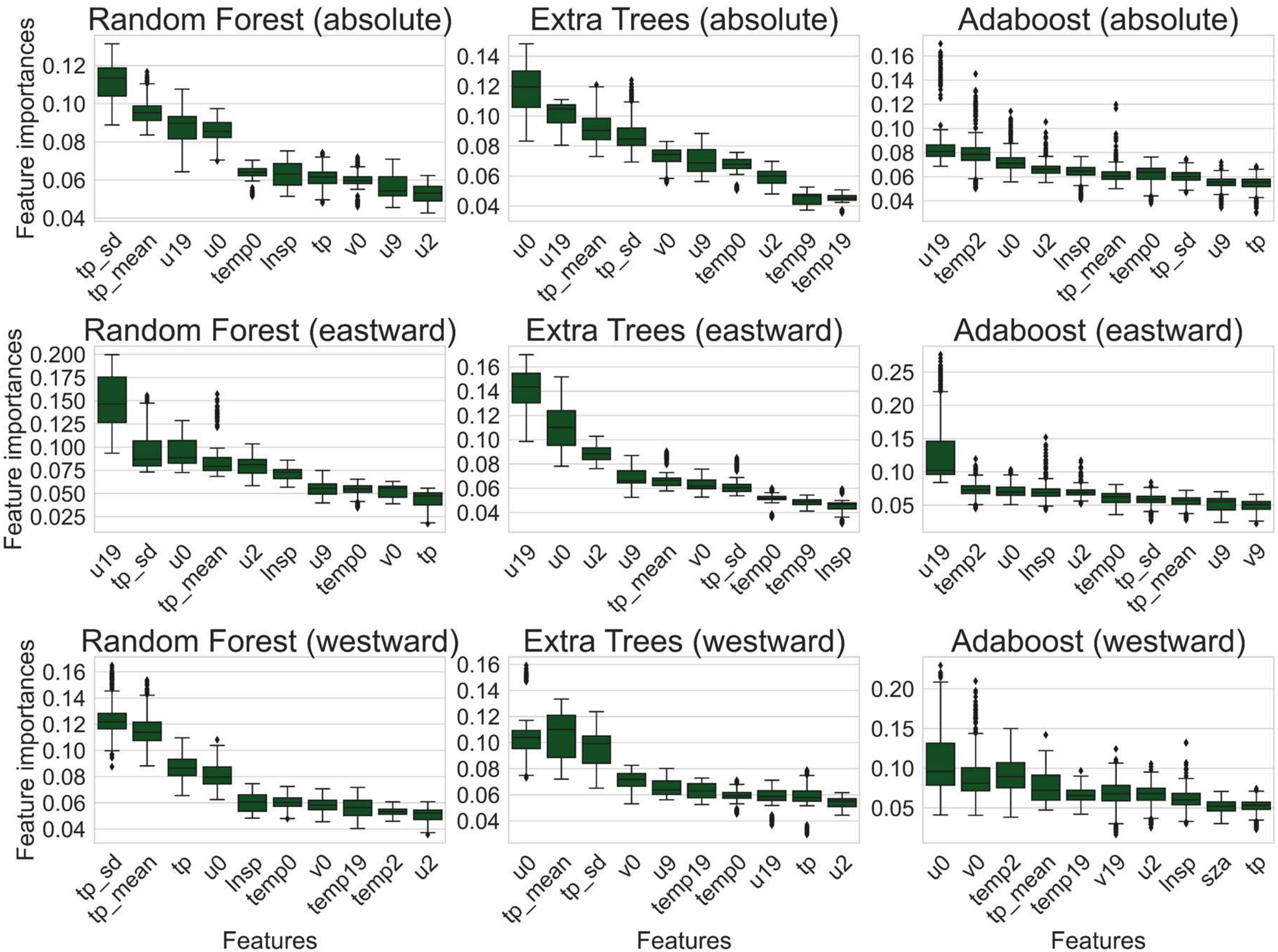
Predicted and actual eastward GWMF of HF (top) and WF (bottom) waves in 24h resolution

Westward GWMF : Balloon 8



Predicted and actual westward GWMF of HF (top) and WF (bottom) waves in 24h resolution

Feature importance : HF

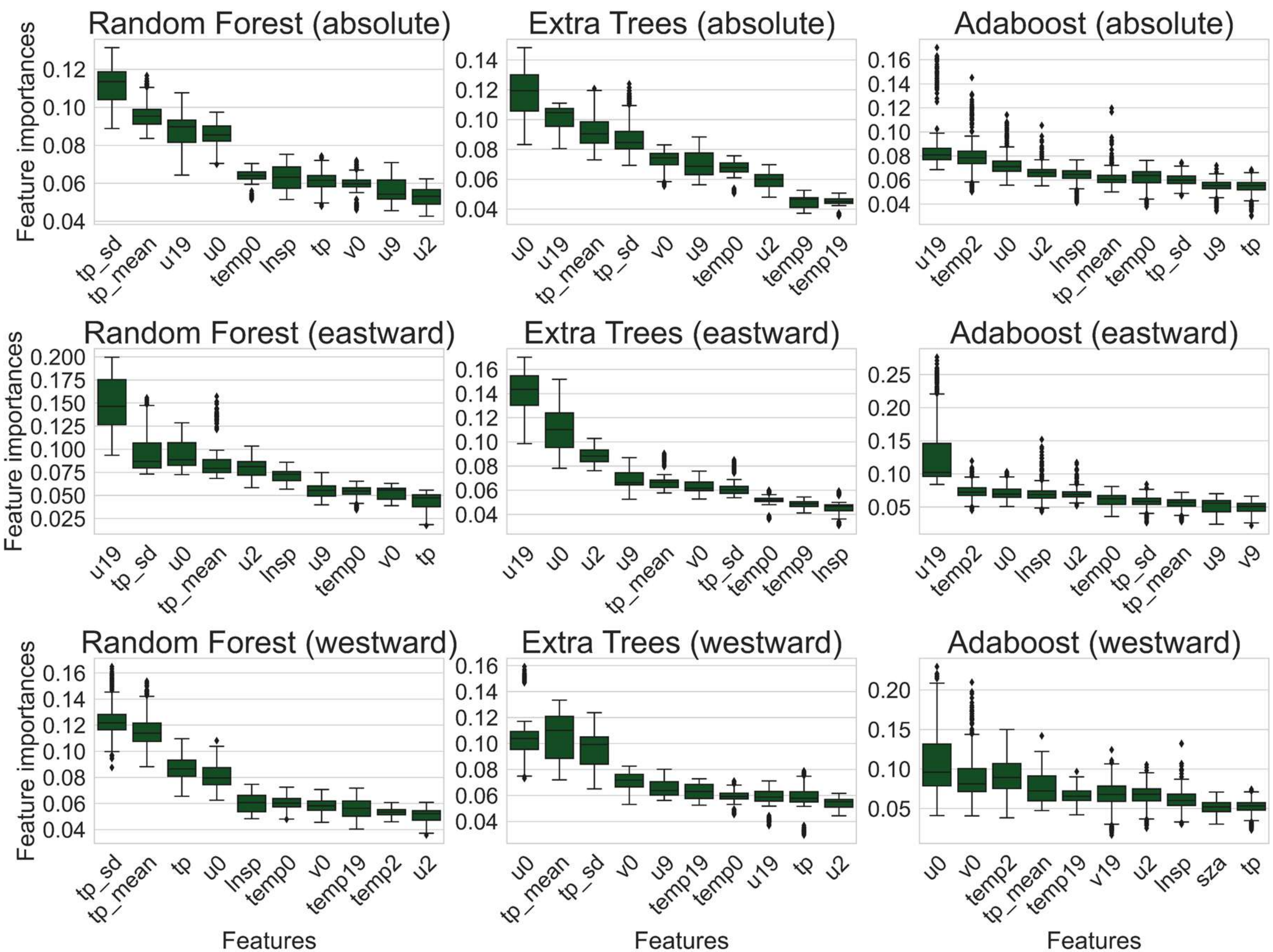


In general, precipitation and zonal wind are the most important features

Wind at balloon level u19 first in eastward case for all models

Surface wind also very informative in many cases

Feature importance : WF

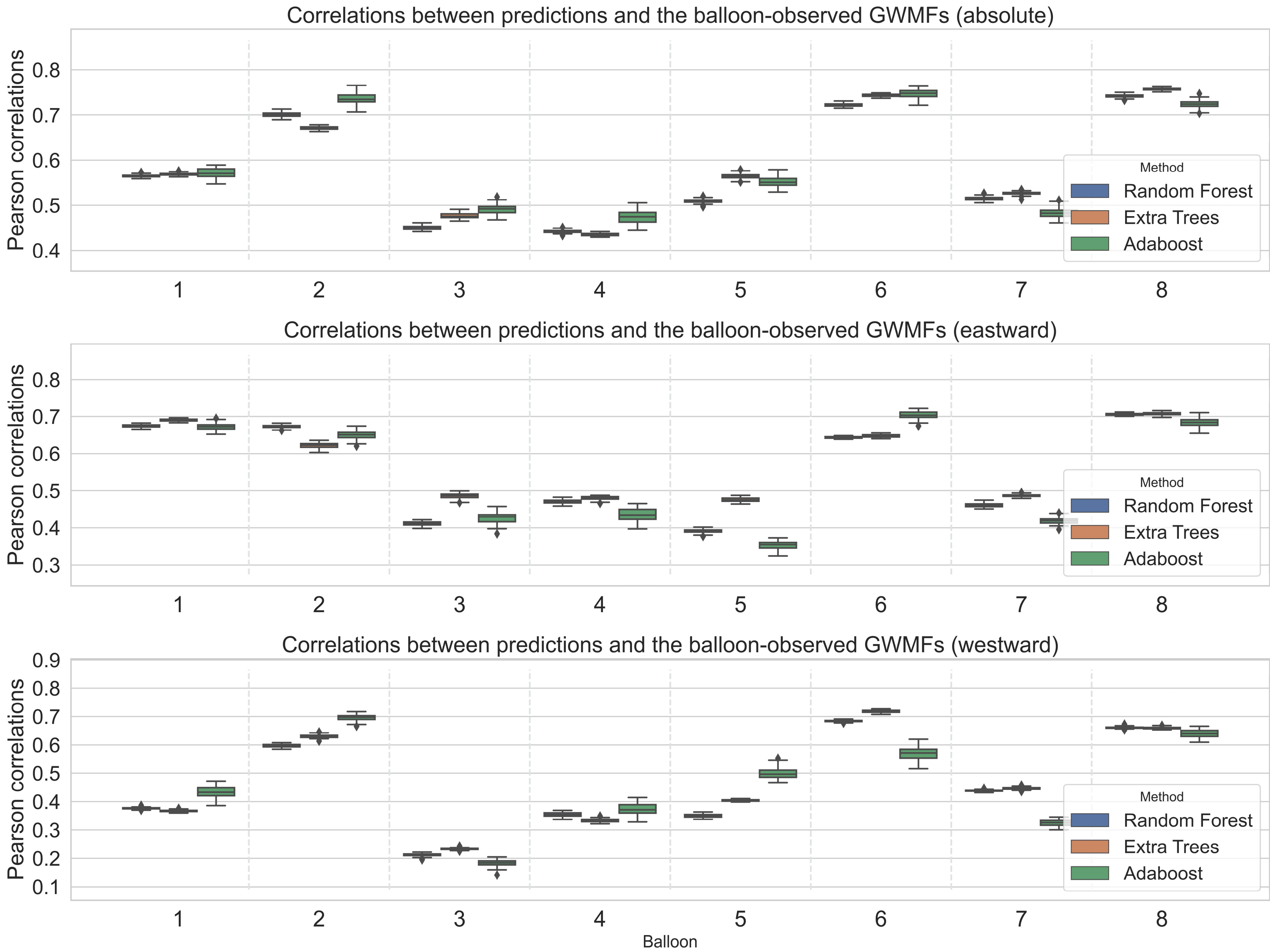


Importance of zonal wind for absolute GWMF

Wind at balloon level u19 in eastward case for all models

Precipitations more informative in westward cases

Correlations HF (50 runs)

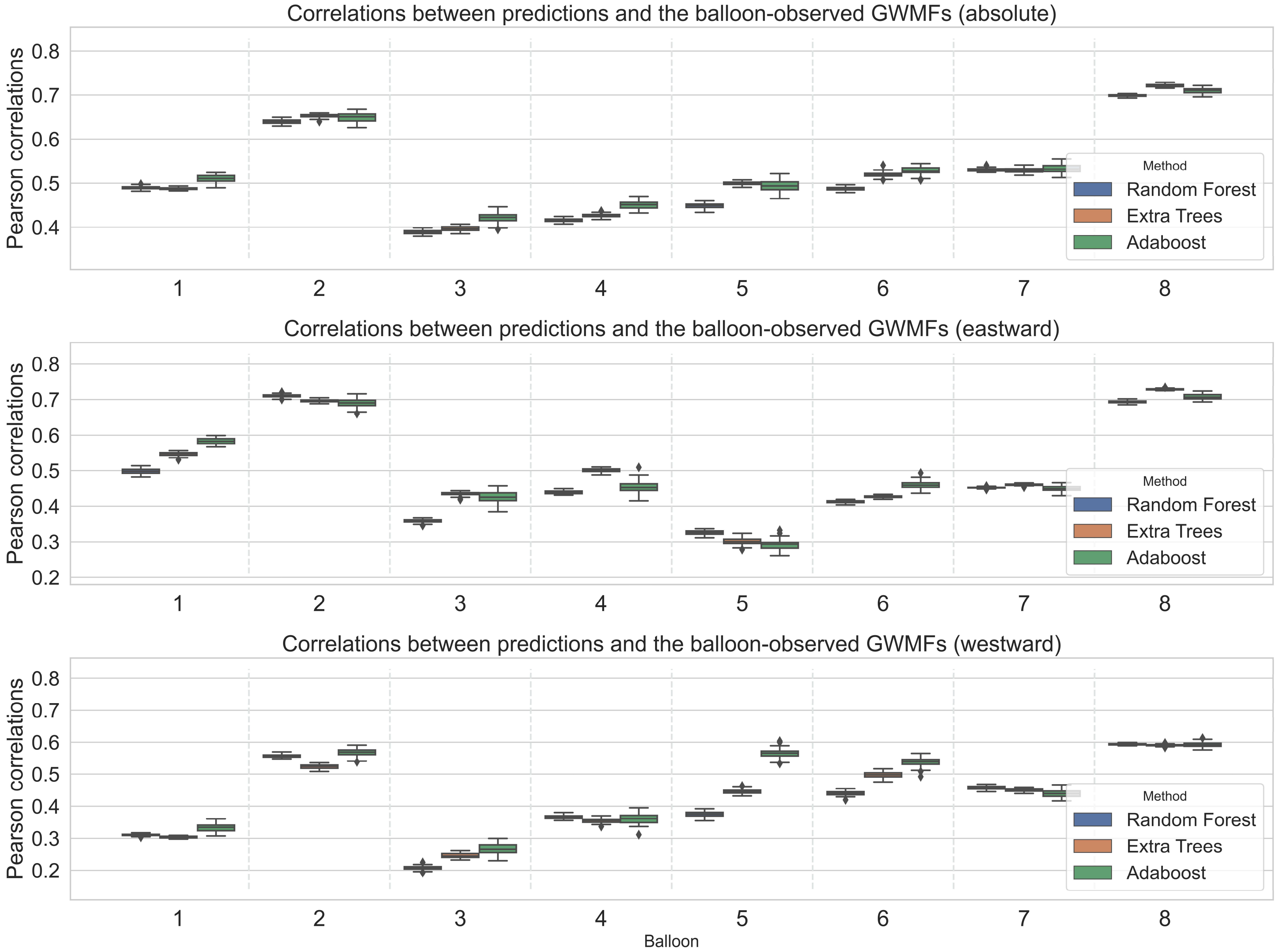


ML methods perform similarly

Balloons 2, 6, 8 well predicted (cor > 0.7)

Westward GWMF more challenging

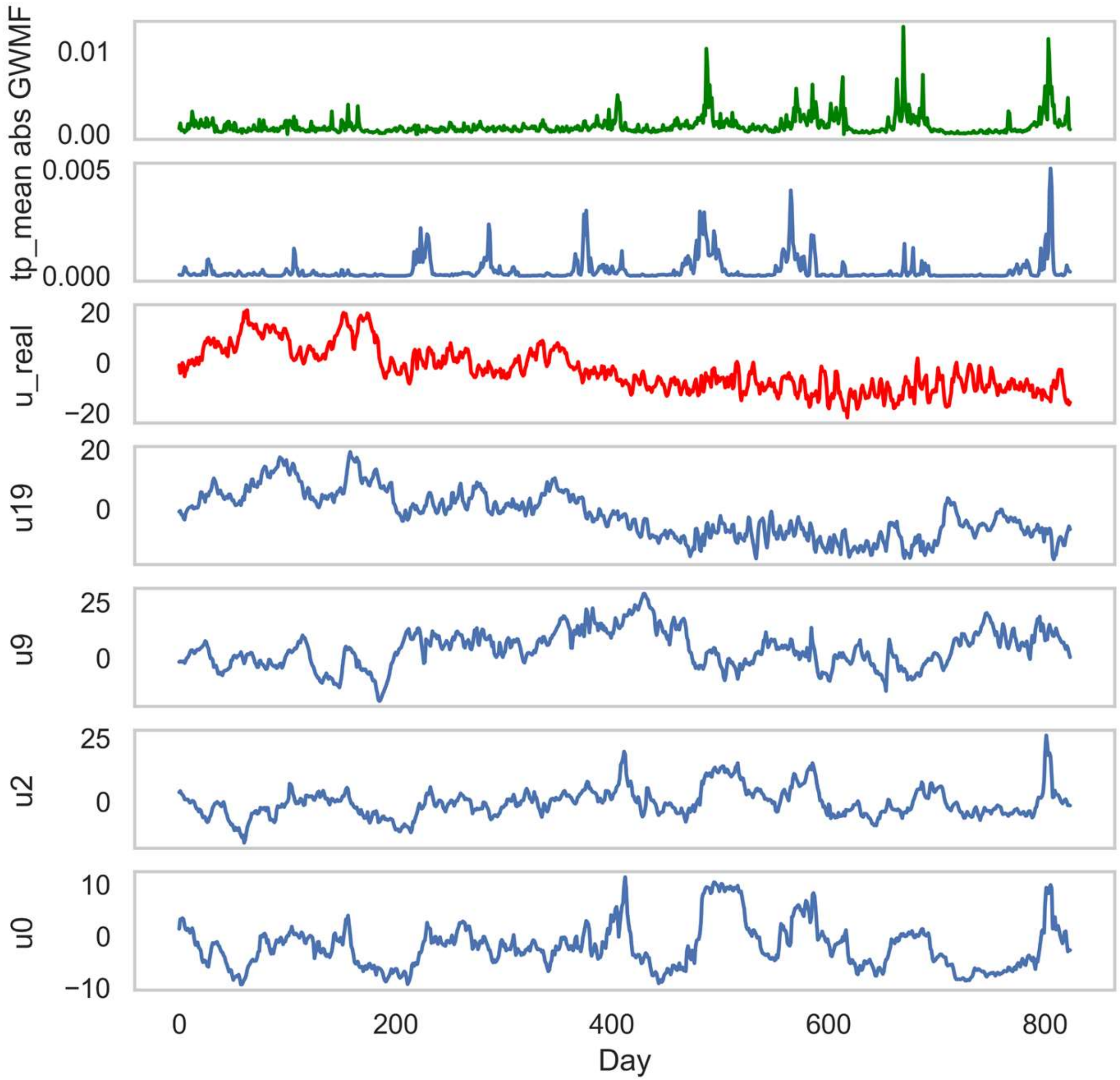
Correlations WF (50 runs)



Balloons 2 and 8 still well predicted (cor > 0.7), but not 6

Performance on WF often lower

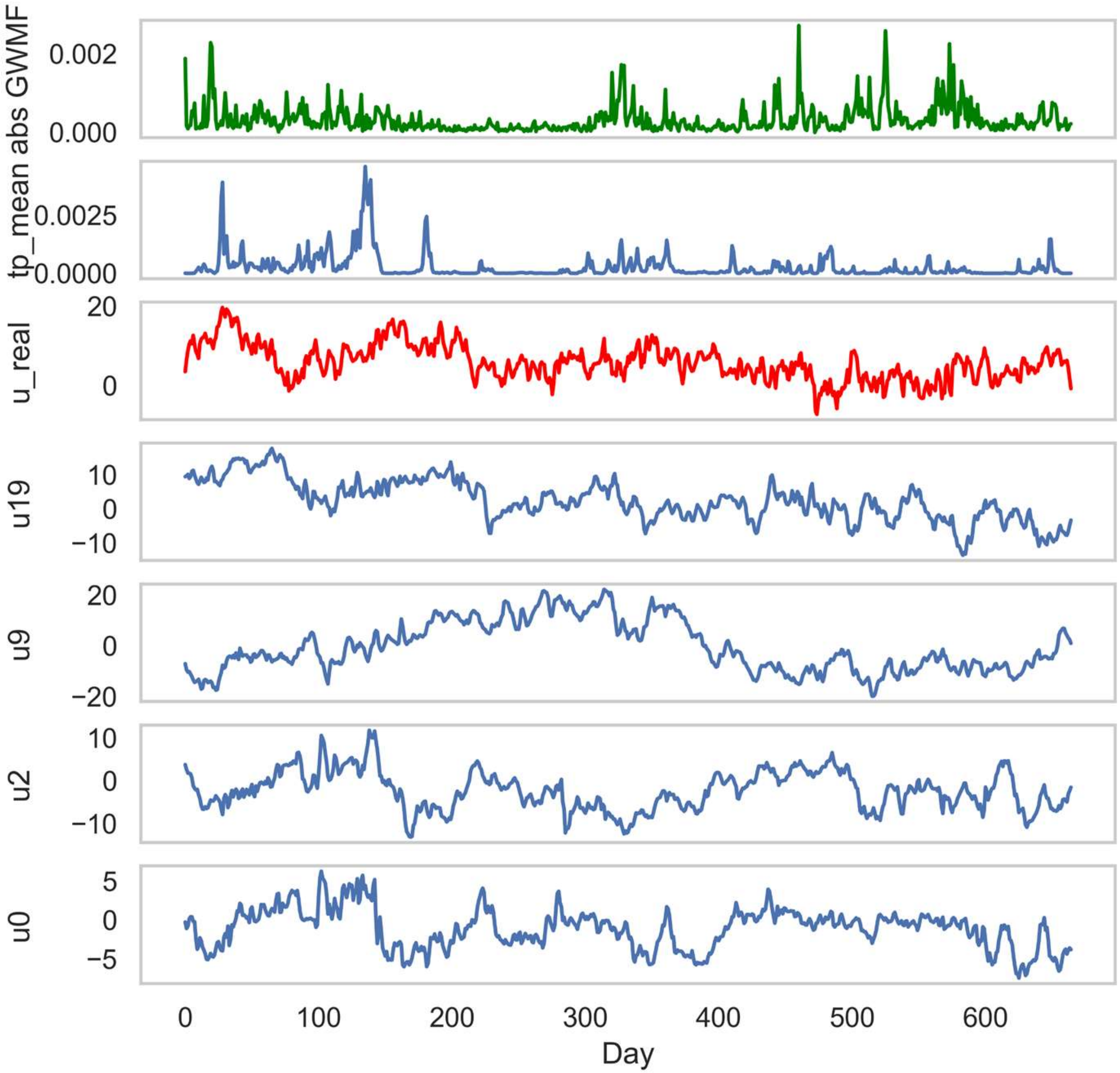
Absolute GWMF vs important variables : Balloon 2



Precipitations correspond well to GWMF

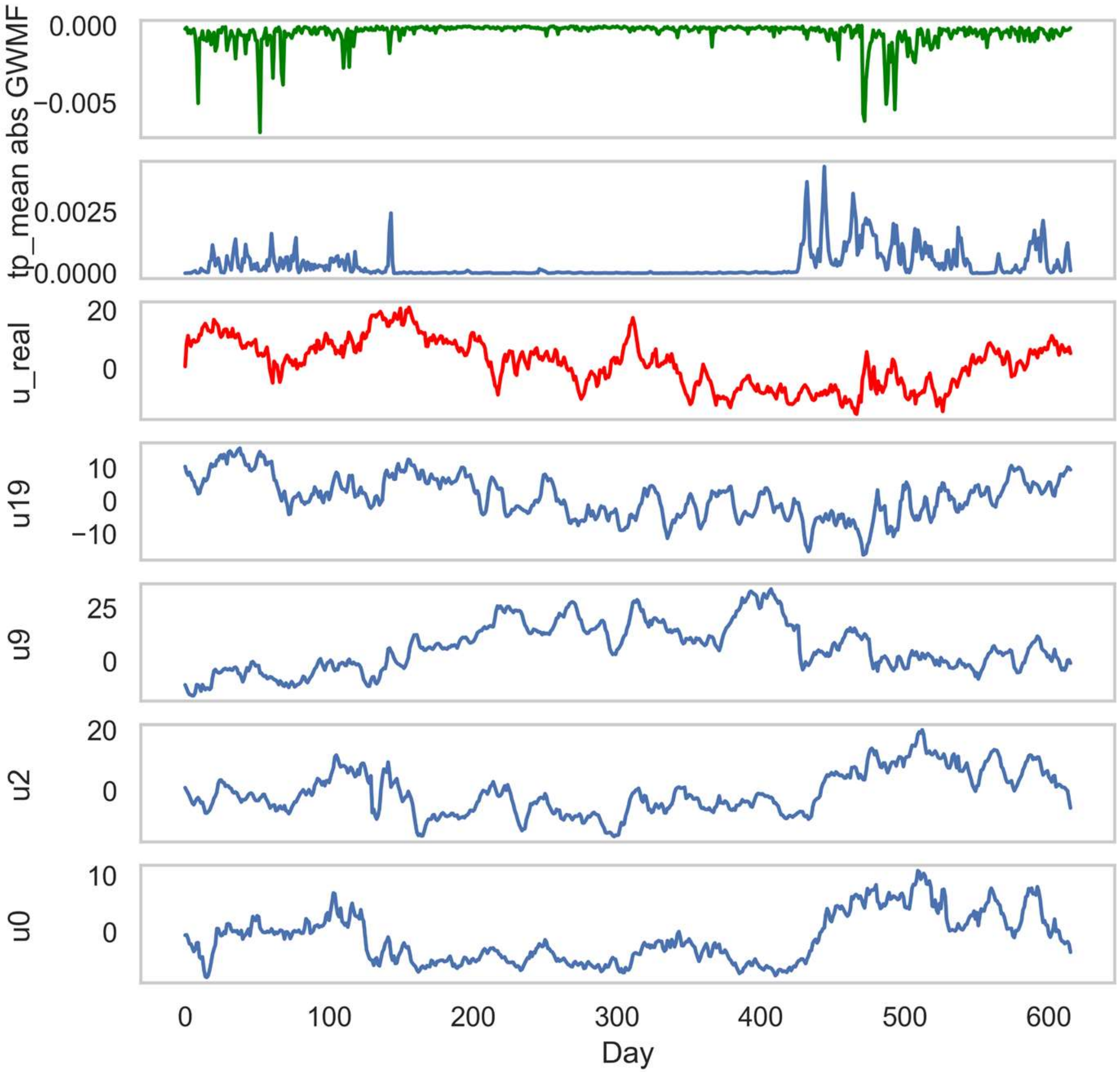
Winds seem informative as well, both at balloon level and below

Eastward GWMF vs important variables : Balloon 7



Precipitations not very informative.

Westward GWMF vs important variables : Balloon 8



Precipitations and wind seem more informative than in previous case

Remark

Differences HF / WF ?

Frequency determined by the angle of the phase lines :
HF : almost vertical (gravity effective as a restoring force)
LF : oblique, almost horizontal.

Air motion parallel to phase lines

Local information corresponds well to HF waves propagating vertically
WF background noise difficult to link to a source

Conclusion and Perspectives

Reconstruction of GWMF up to an encouraging level (correlation > 0.7) → lower bound on how much can be reconstructed fromlarge-scale flow described by reanalysis

Most informative variables : precipitations + zonal wind at and below balloon level

Ocean / land

Observations from C1, next campaign, combination with high resolution simulations

Add further informative inputs ?
For instance, idea : add brightness temperature images