

# Machine Learning Approach for Forecasting Space Weather Effects in the Ionosphere with Uncertainty Quantification

Randa Natras<sup>1</sup>, Benedikt Soja<sup>2</sup>, Michael Schmidt<sup>1</sup>, Marie Dominique<sup>3</sup>, Ayse Türkmen<sup>1</sup>

<sup>1</sup>Deutsches Geodätisches Forschungsinstitut der Technischen Universität München (DGFI-TUM),  
School of Engineering and Design, Technical University of Munich, Munich, Germany

<sup>2</sup>Institute of Geodesy and Photogrammetry, ETH Zurich, Switzerland

<sup>3</sup>Solar-Terrestrial Center of Excellence, SIDC, Royal Observatory of Belgium, Brussels, Belgium

[randa.natras@tum.de](mailto:randa.natras@tum.de)

European Geosciences Union General Assembly 2022 (EGU22)

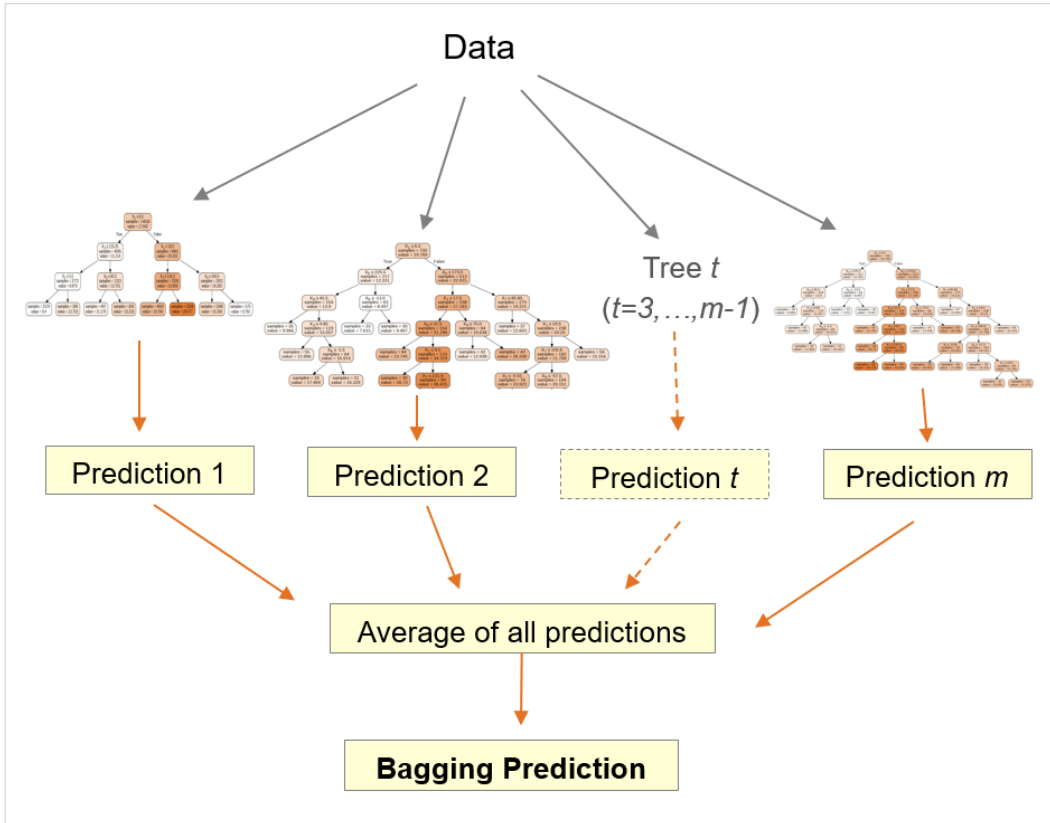
Vienna, May 27, 2022

This presentation participates in OSPP



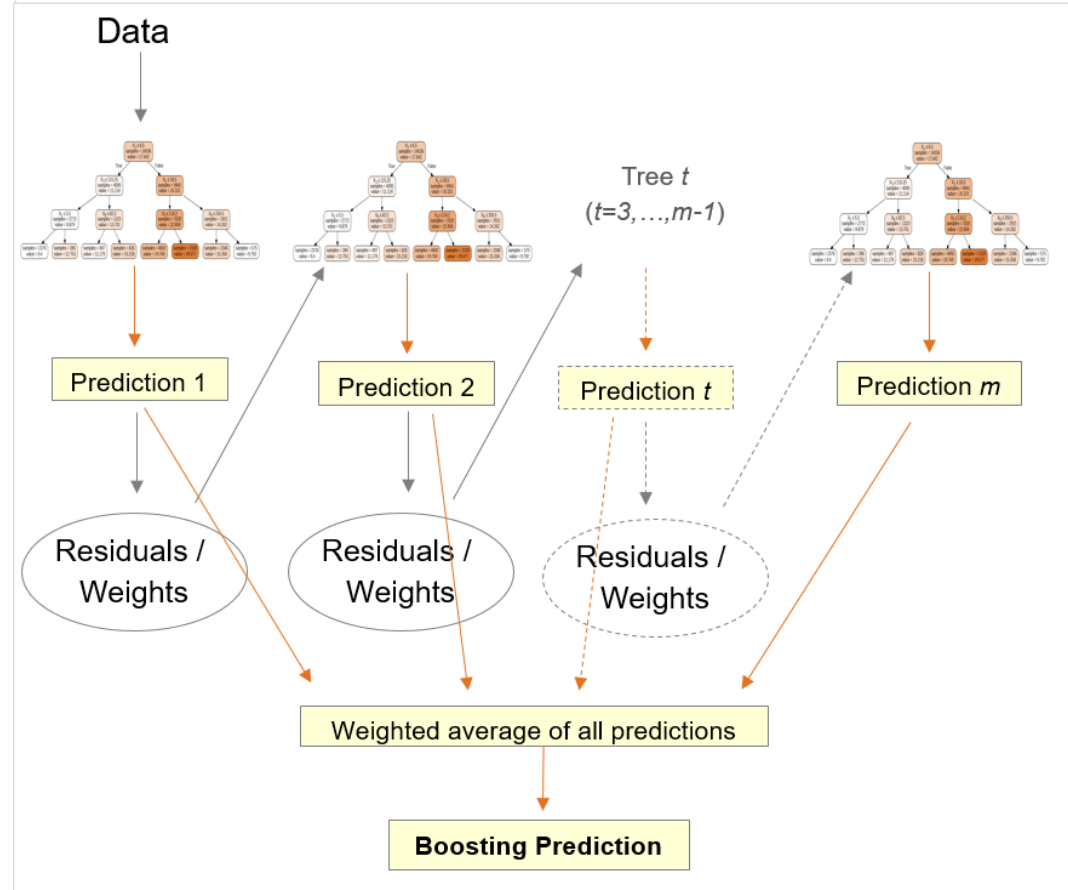
Outstanding Student & PhD  
candidate Presentation contest

# Tree-based learning



- **Bagging** (Parallel learning):
  - Random Forest (multiple randomized trees)

- **Boosting** (Sequential learning):
  - Adaptive Boosting - AdaBoost (training with weighted observations)
  - Gradient Boosting - GBoost (training with residuals)



# 1-day VTEC Forecasting, Data (time sampling 1h)

## Input data:

- Time: Hour of day and Day of year (DOY)
- Sunspot number R (daily)
- Solar radio flux F10.7 (daily)
- Solar wind plasma speed (hourly)
- Bz index (hourly)
- AE index (hourly)
- Dst index (hourly)
- Kp index (3-hour)
- VTEC from GIM CODE (hourly)
  - 10E 70N, 10E 40N, 10E 10N
- Exponential moving average of VTEC over previous 4 days and 30 days
- First time derivative of VTEC
- Second time derivative of VTEC

Time:  
t

## Output data:

- VTEC
    - 10E 70N,
    - 10E 40N,
    - 10E 10N
- Time:  
t+24h

## Data split:

- Training & Cross-validation: 2015 - 2016
  - Time series cross-validation
- Test: 2017

# VTEC Nowcasting, Data (time sampling 1h)

## Input data:

- Time:  $t$ 
  - Time: Hour of day and Day of year (DOY)
  - Sunspot number R (daily)
  - PROBA2 Lyra (1 min)
  - Solar wind plasma speed (hourly)
  - Bz index (hourly)
  - AE index (hourly)
  - Dst index (hourly)
  - Kp index (3-hour)
- Time:  $t-1h$ 
  - VTEC from GIM CODE (hourly)
    - 10E 70N, 10E 40N, 10E 10N
  - Exponential moving average of VTEC over previous 4 days and 30 days
- Time:  $t-1h$ 
  - First time derivative of VTEC
  - Second time derivative of VTEC

## Output data:

- VTEC
    - 10E 70N,
    - 10E 40N,
    - 10E 10N
- Time:  $t$

## Data split:

- Training & Cross-validation: 2015 - 2016
  - Time series cross-validation
- Test: 2017

# Uncertainty quantification (UQ)

## ➤ Goals:

- define the **accuracy and precision** of VTEC prediction,
- quantify the level of **trust** in VTEC prediction,
- assess the **reliability** of VTEC prediction.

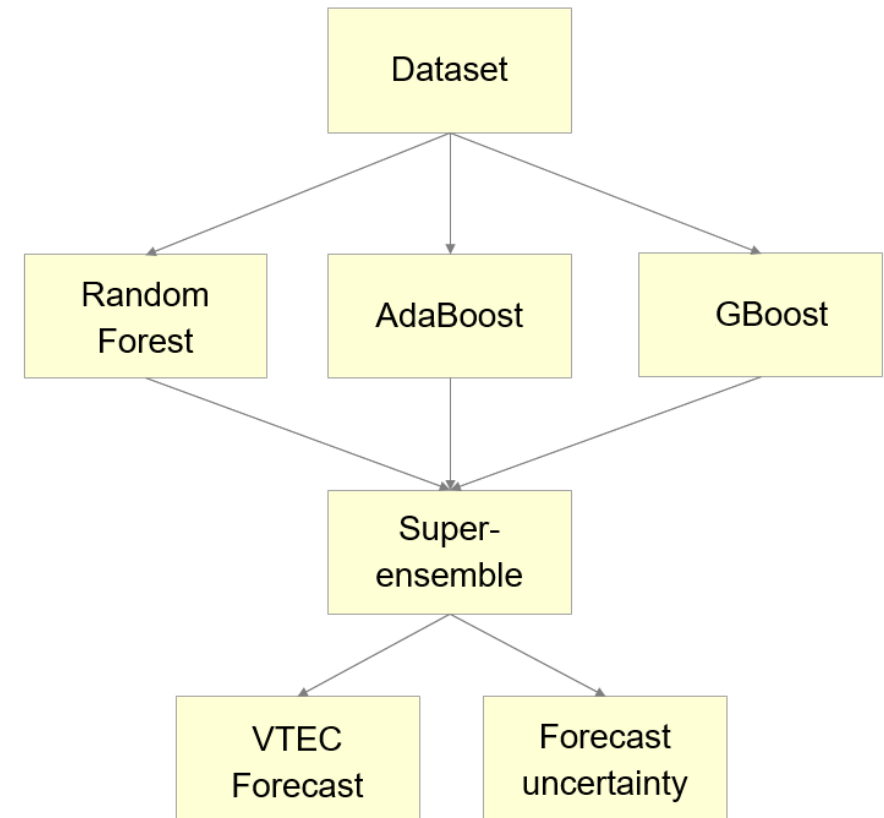
## I. Multi-model and multi-data ensemble

- VTEC forecast -> ensemble mean
- Uncertainty -> ensemble spread ( $2\sigma$ )
- 3 datasets<sup>\*</sup>:
  1. Original data as input and output,
  2. Daily differences as input and output,
  3. Original data + daily differences as input, original data as output.

## II. Confidence intervals

- Quantile objective loss function
- Applied for GBoost and 3. dataset
- Selected quantile  $\alpha = 0.95$

<sup>\*</sup> Observations were preprocessed / cleaned before training.



$$\mathcal{L}(e_i|\alpha) = \begin{cases} \alpha \cdot e_i & \text{if } e_i \geq 0, \\ (\alpha - 1) \cdot e_i & \text{if } e_i < 0 \end{cases}$$

$$e_i = y_i - f(\mathbf{x}_i)$$

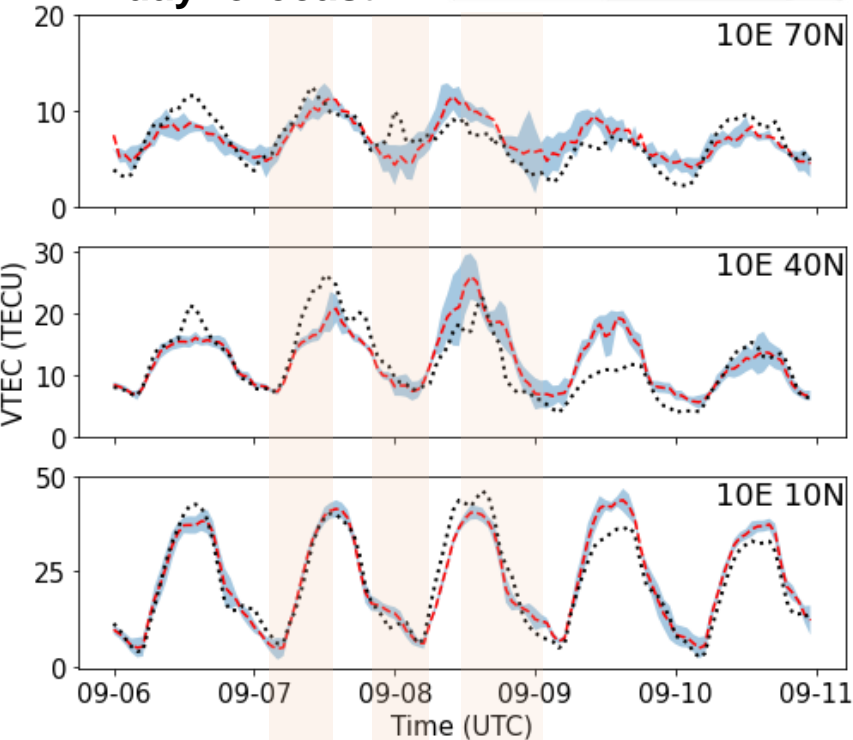
$$\mathcal{L}(\mathbf{e}|\alpha) = \frac{1}{N} \sum_{i=1}^N \mathcal{L}(e_i|\alpha)$$

# Results for space weather events in September 2017

## UQ: Ensemble

- - - Ensemble mean
- · - · - VTEC GIM
- Ensemble spread  $2\sigma$

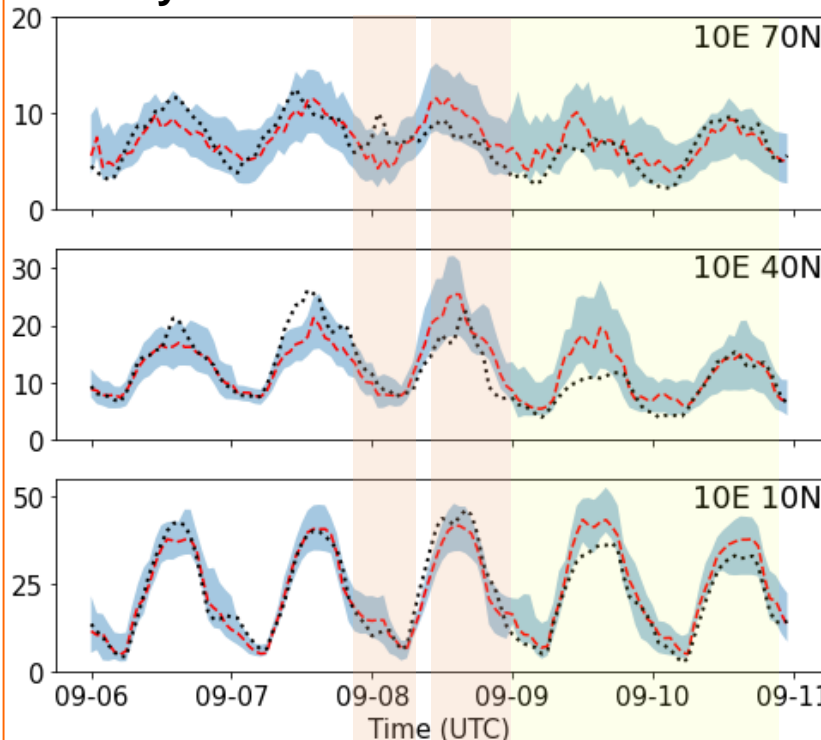
### 1-day forecast



## UQ: Confidence intervals

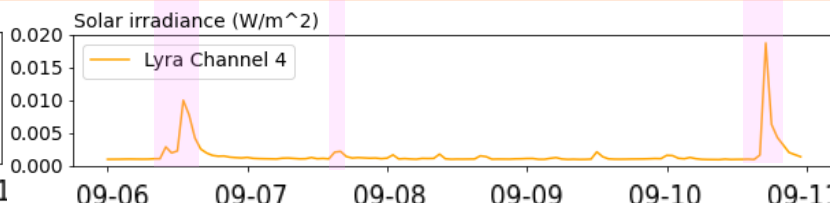
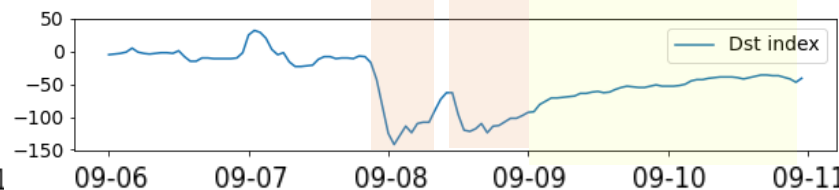
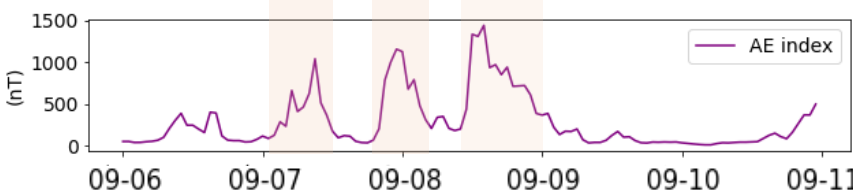
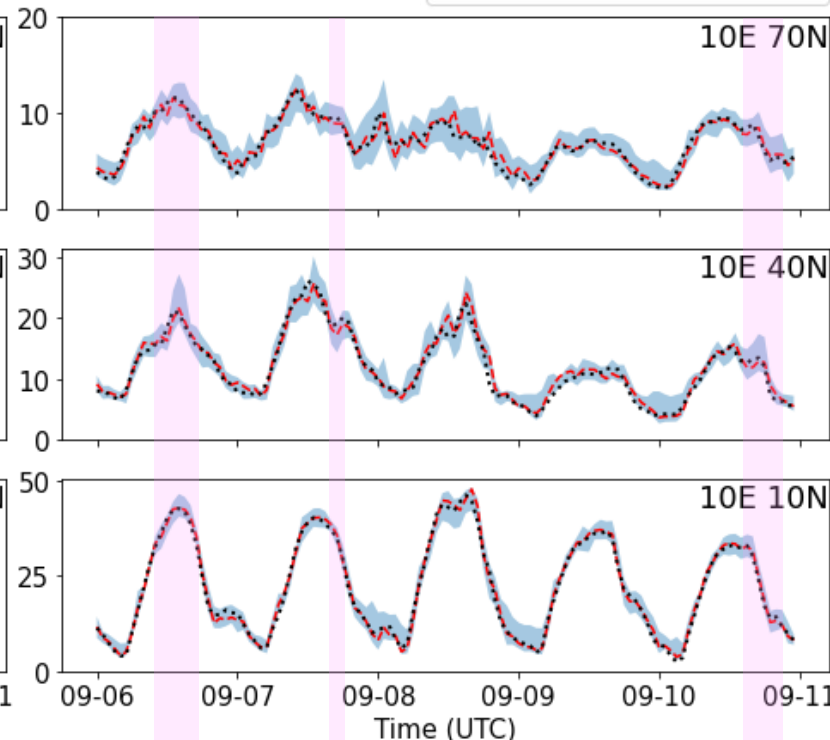
- Geomagnetic storm
- Recovery phase
- X solar flares

### 1-day forecast



### Nowcast

- - - VTEC forecast (GBoost)
- · - · - VTEC GIM
- 90% interval ( $\alpha = 0.95$ )



- The **uncertainty** information defines the **reliability** and **precision** of VTEC predictions.
- UQ allows to assess the **trustworthiness of predictions**.
- Ground-truth VTEC mostly within predicted confidence intervals for space weather events.

**Thank you for  
your attention!**

randa.natras@tum.de



Scan for the abstract and display material

