Deutsches Geodätisches Forschungsinstitut (DGFI-TUM)
Technische Universität München

Machine Learning Ensemble Approach for Ionosphere and Space Weather Forecasting with Uncertainty Quantification

Randa Natras[1], Benedikt Soja[2], Michael Schmidt[1]

[1]Deutsches Geodätisches Forschungsinstitut der Technischen Universität München (DGFI-TUM),
School of Engineering and Design, Technical University of Munich, Munich, Germany
[2]Institute of Geodesy and Photogrammetry, ETH Zurich, Switzerland

randa.natras@tum.de

URSI combined Atlantic / Asia-Pacific Radio Science Conference 2022 (URSI AT-AP-RASC 2022)
Gran Canaria, May 29 – June 3, 2022

Content

- Motivation and objectives
- Learning algorithms
- Data
- Uncertainty quantification
- Results: Space weather events of September 6-10, 2017
- Conclusion

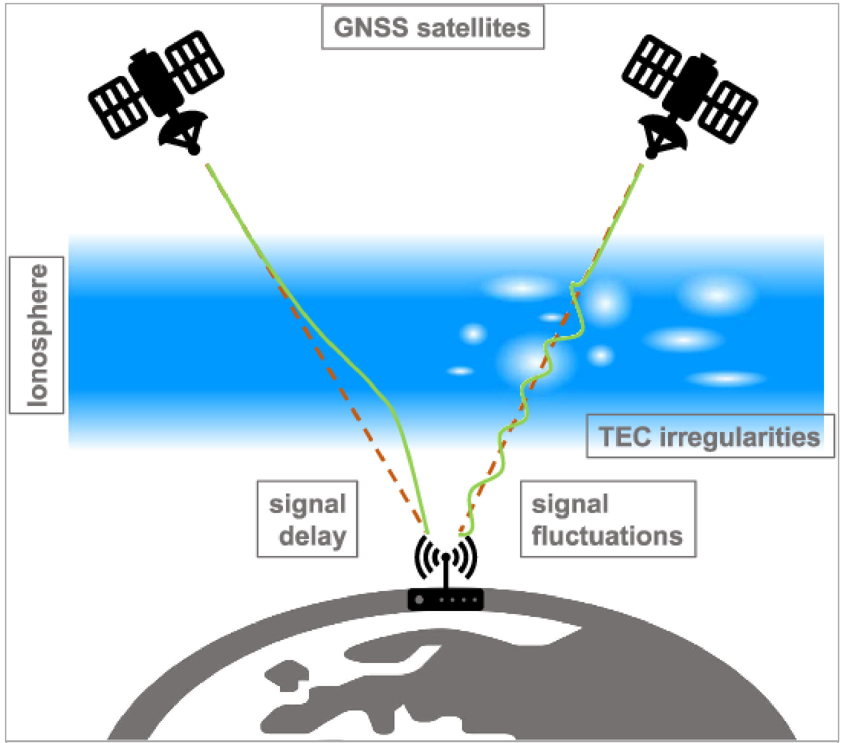
Motivation and objectives

✓ Machine learning (ML):

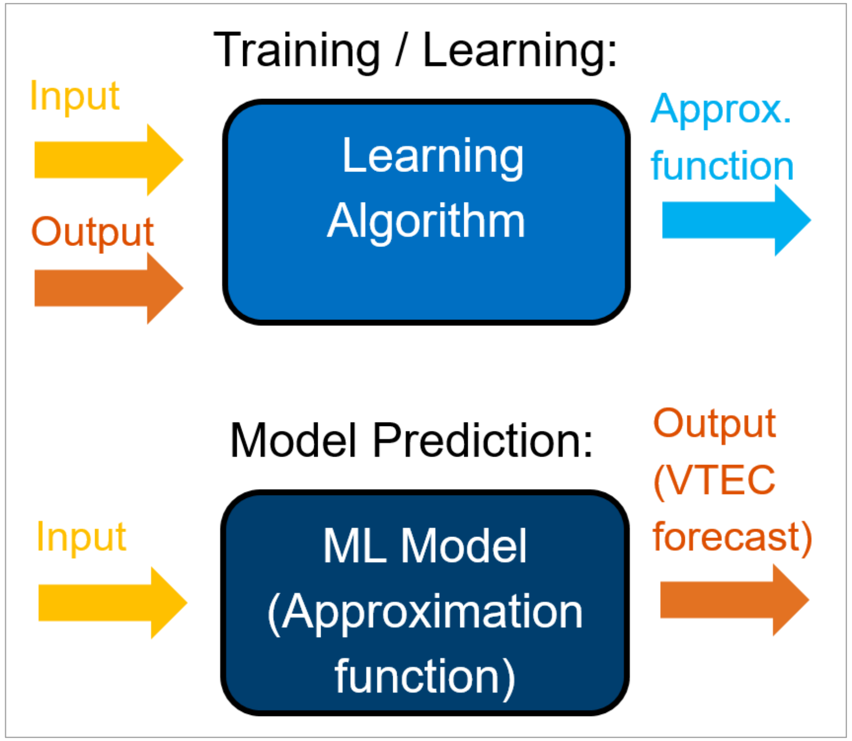
- ➢ “Learning” from the data
- ➢ Approximating nonlinear functions

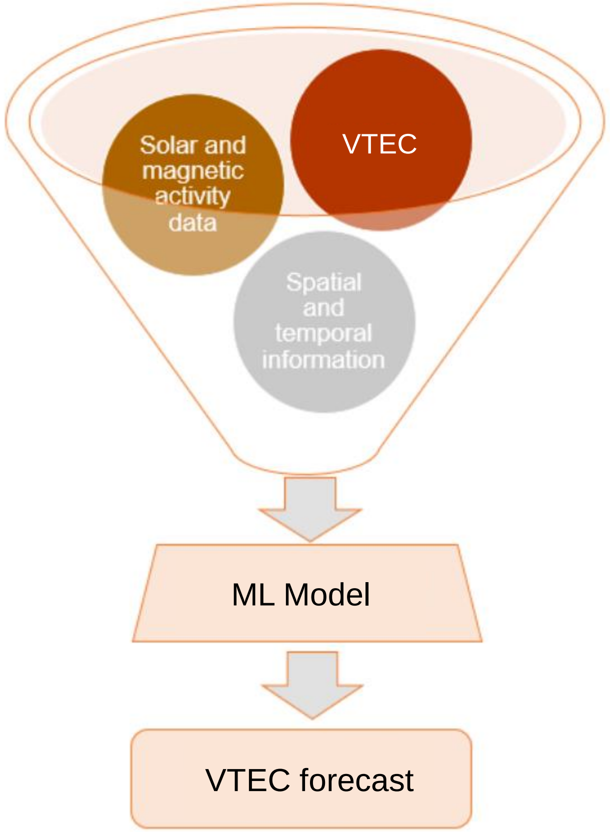
✓ Objectives:

- ➢ Forecasting/Nowcasting of Vertical Total Electron Content (VTEC) including the space weather impact

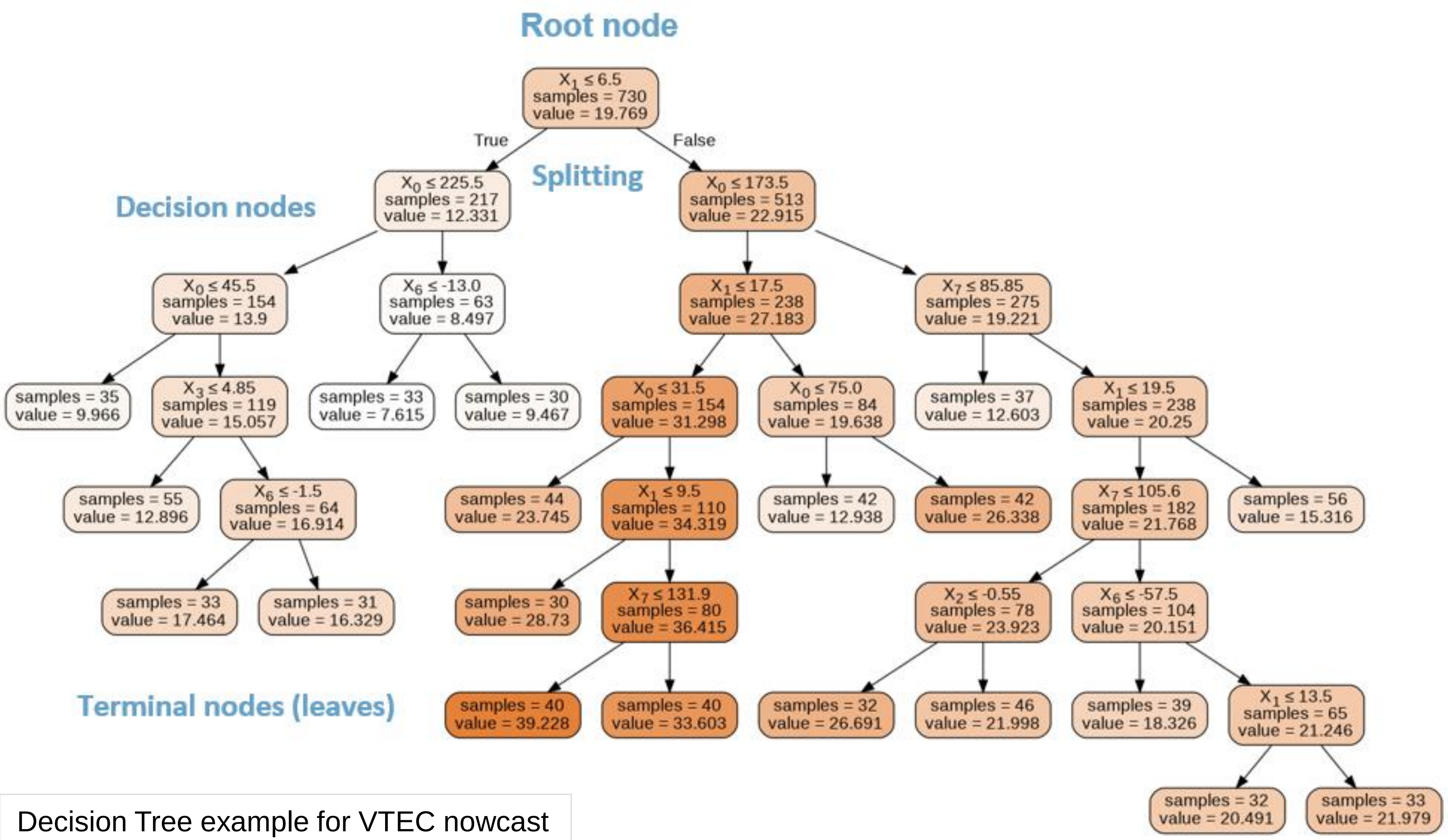


Source: https://www.semanticscholar.org/paper/Detection-of-GNSS-Ionospheric-Scintillations-Based-Linty-Farasin/3bc53da7342d4cdcd1a8bacfdc92651aeb62d5dc





Learning algorithm: Decision Tree

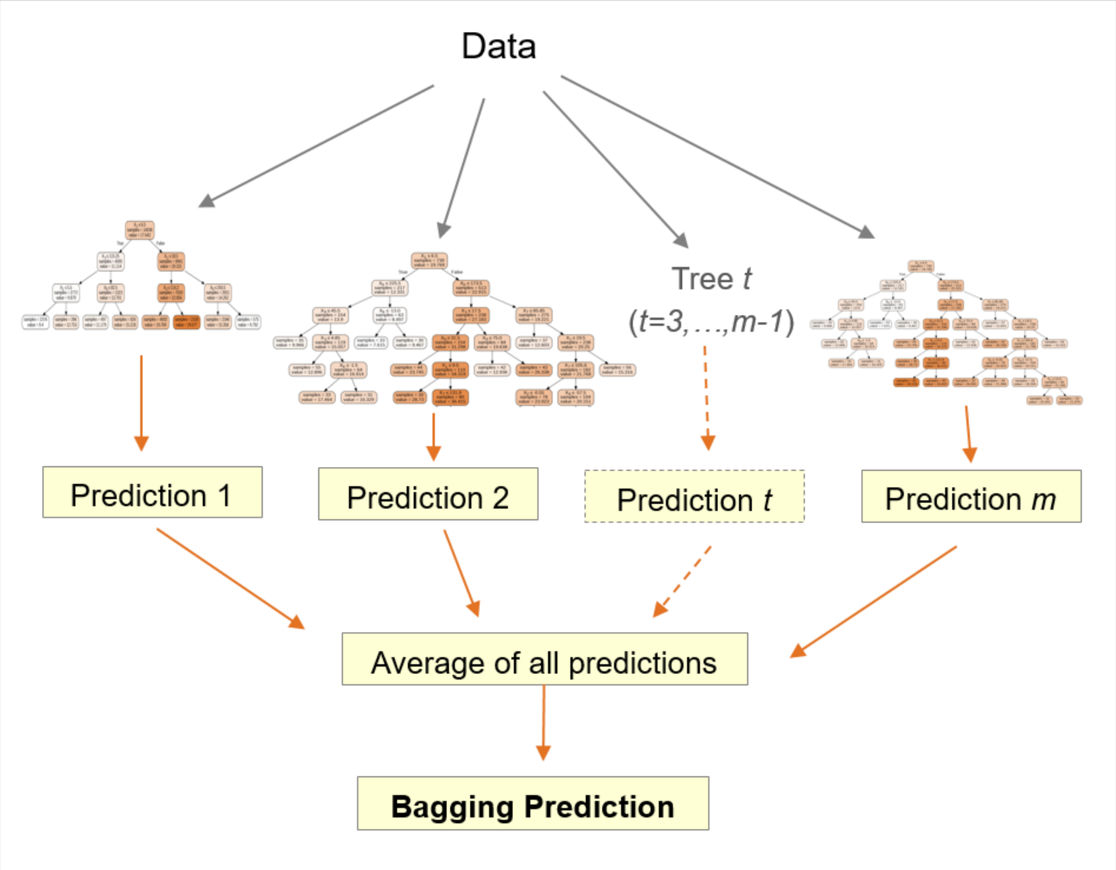


Decision Tree example for VTEC nowcast

- **Root node:** contains the entire dataset and further is divided into two subnodes
- **Splitting:** process of dividing a node into two subnodes by calculating reduction in variance
- **Decision node:** a subnode, which splits into further sub-nodes
- **Leaf / Terminal node:** nodes that do not split

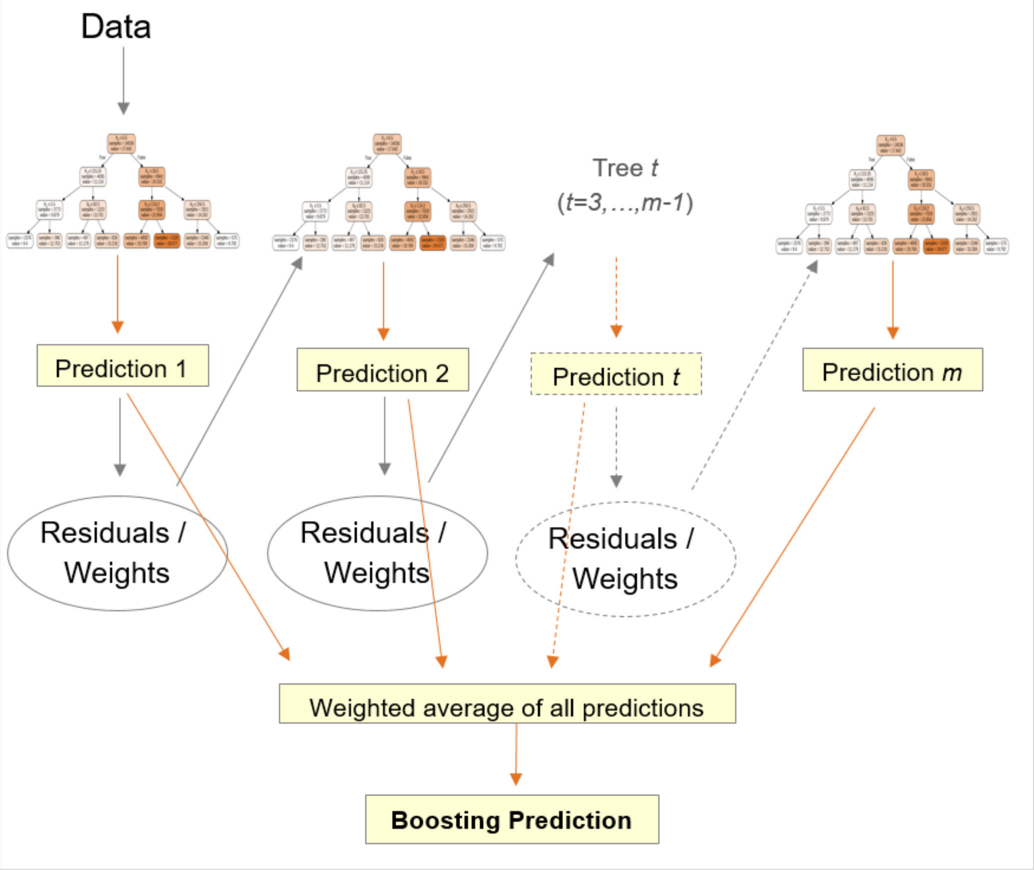
Final outcome: the average VTEC value in the particular leaf node.

Tree-based learning



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

- **Boosting** (Sequential Ensemble 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: t

- 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

Output data:

- VTEC
  - 10E 70N,
  - 10E 40N,
  - 10E 10N

Time: t+24h

Data split:

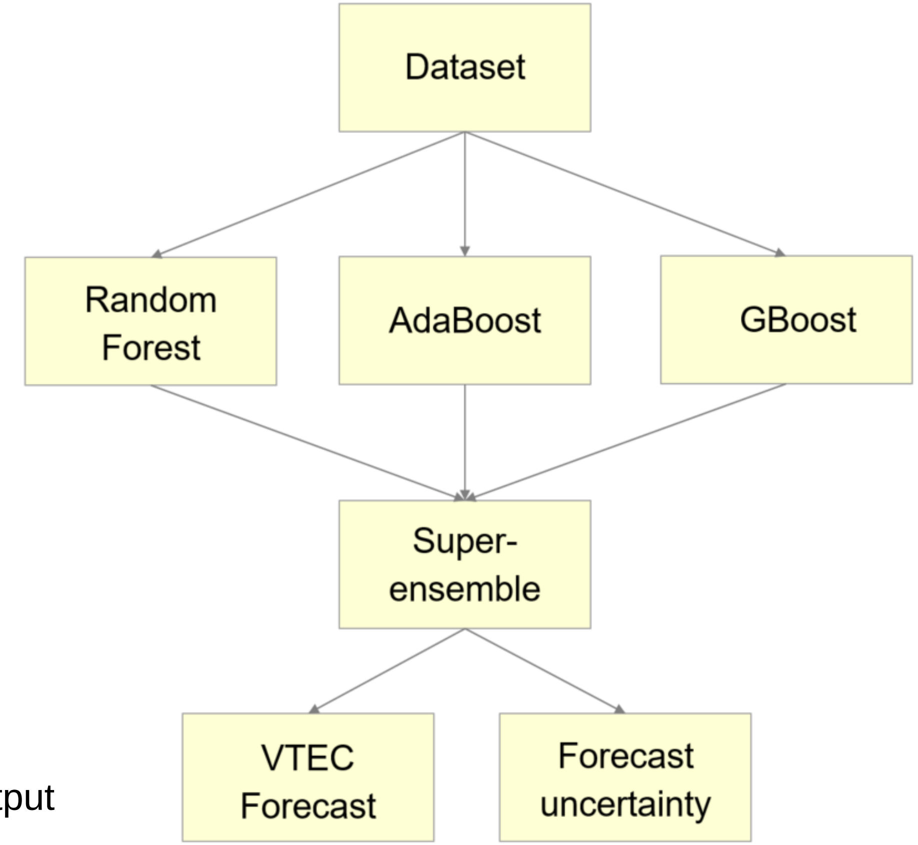
- Training & Cross-validation: 2015 - 2016
  - Rolling cross-validation
- Test: 2017

Uncertainty quantification (UQ)

➢ **Goals:**
- define the efficiency of the ML models,
- quantify the level of trust in a prediction,
- increase reliability of the VTEC predictions.

I. Multi-model and multi-data ensemble

- VTEC forecast -> ensemble mean
- Forecast uncertainty -> ensemble spread (2σ)
- 3 datasets*:
  1. Original data in input and output
  2. Daily differences in input and output
  3. Original data + daily differences in input, original data in output



II. Confidence interval

- Quantile objective loss function
- Applied for GBoost and 3rd dataset
- Quantiles: upper bound $\alpha = 0.95$, lower bound $\alpha = 0.05$

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

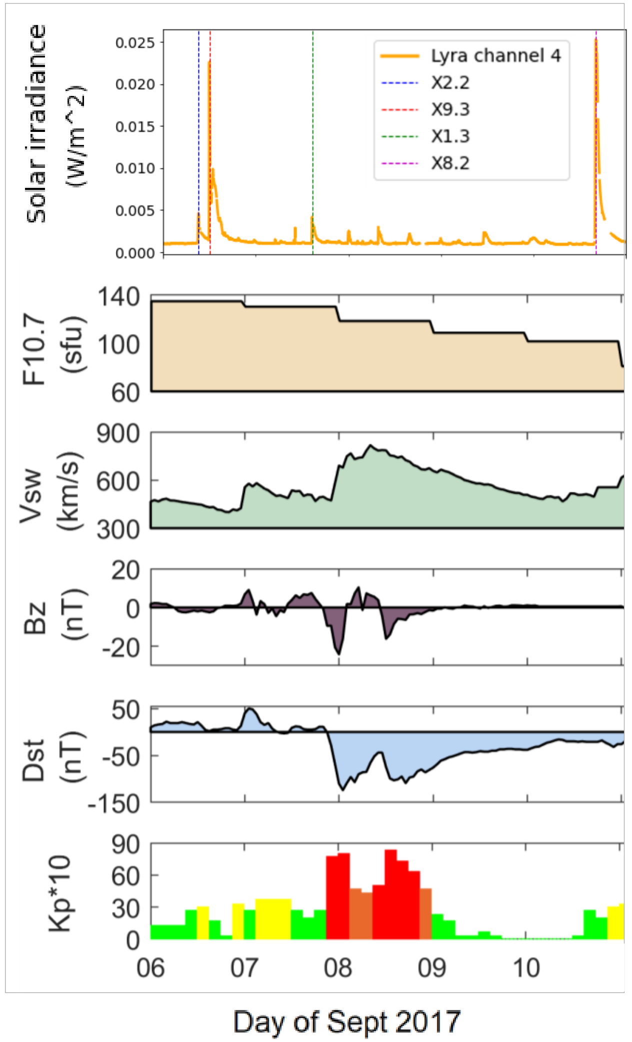
$$e_i = y_i - f(\mathbf{x}_i)$$

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

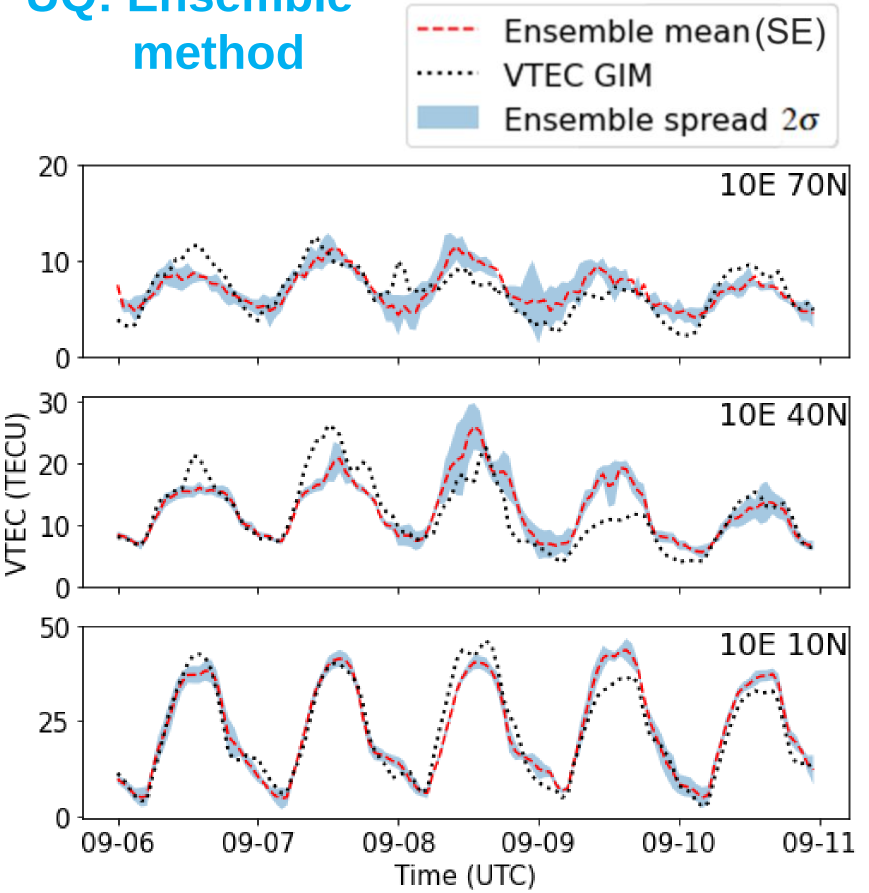
*Observations were preprocessed / cleaned before training.

Results: September 2017 space weather events

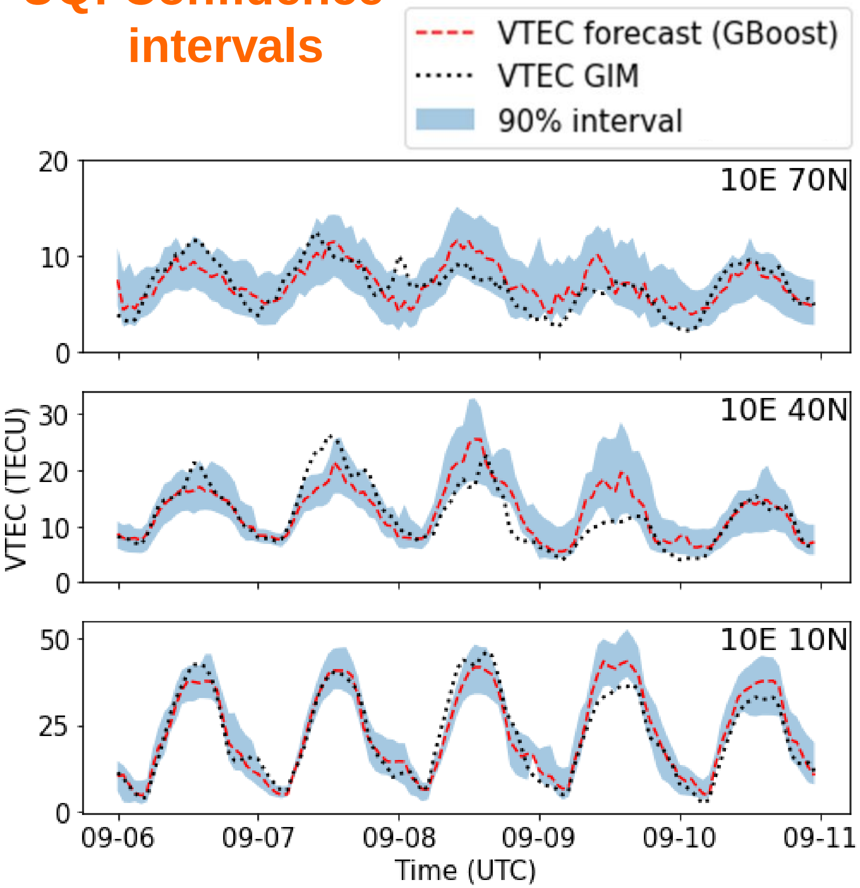
Space weather overview



UQ: Ensemble method



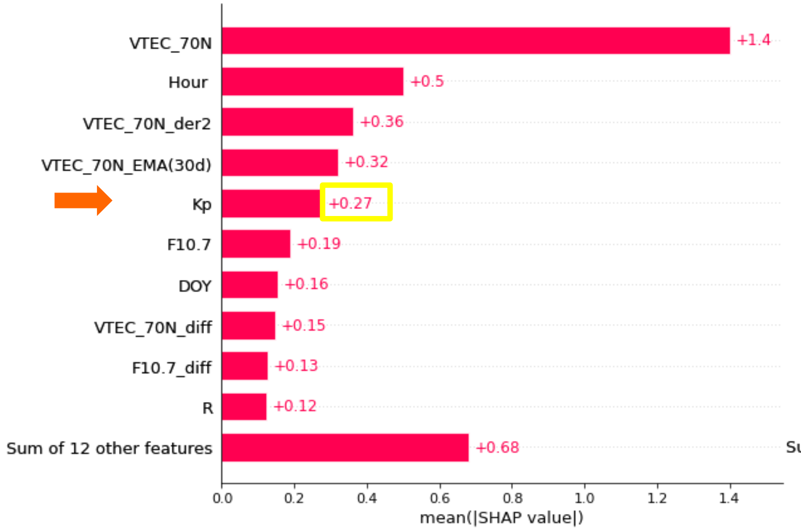
UQ: Confidence intervals

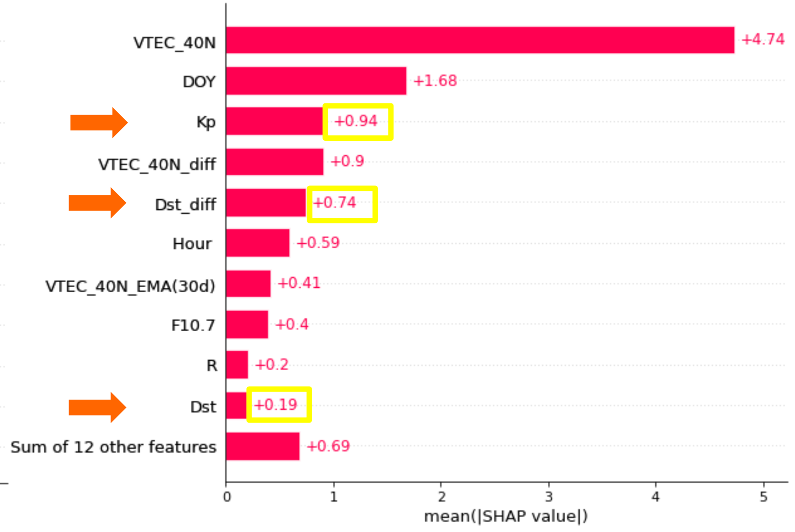


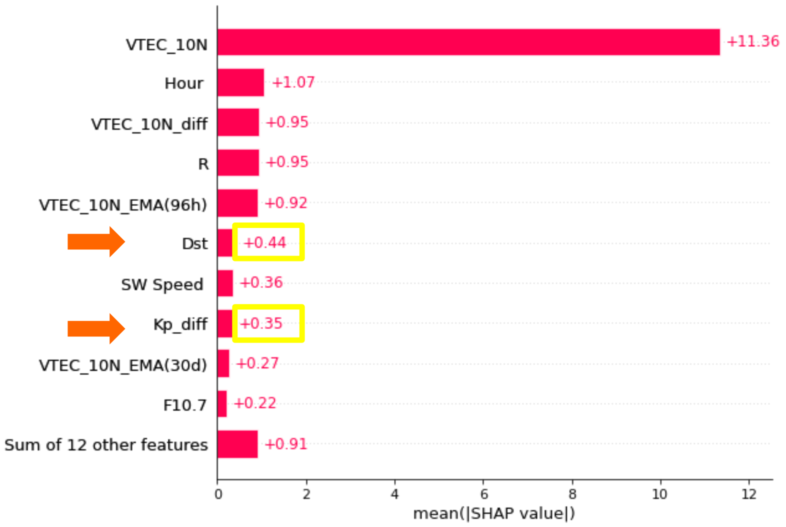
Features average impact on model output magnitude

Sept 8-9, 2017

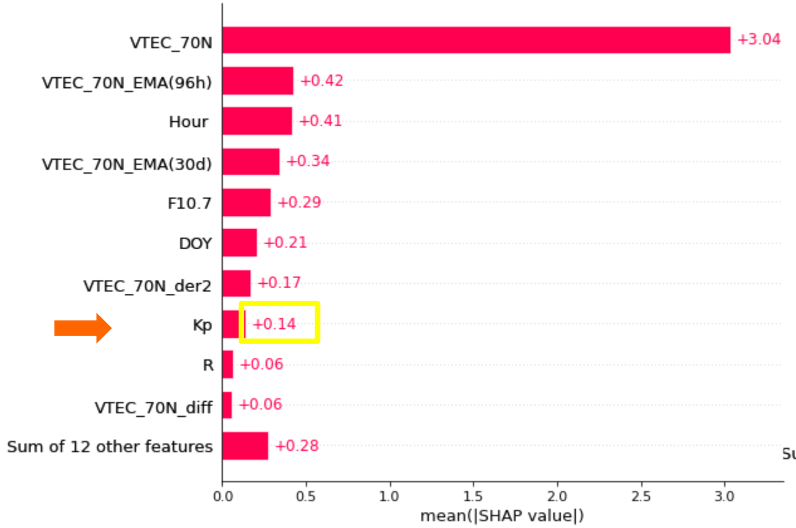
GBoost model

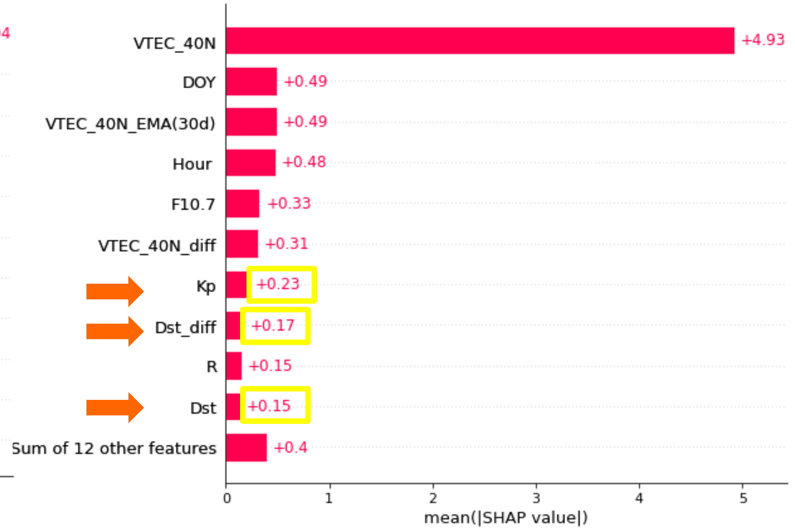


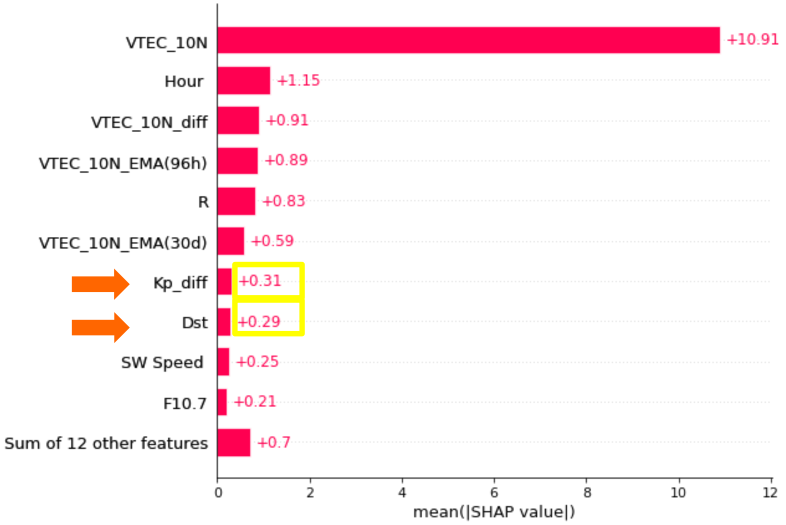




2017

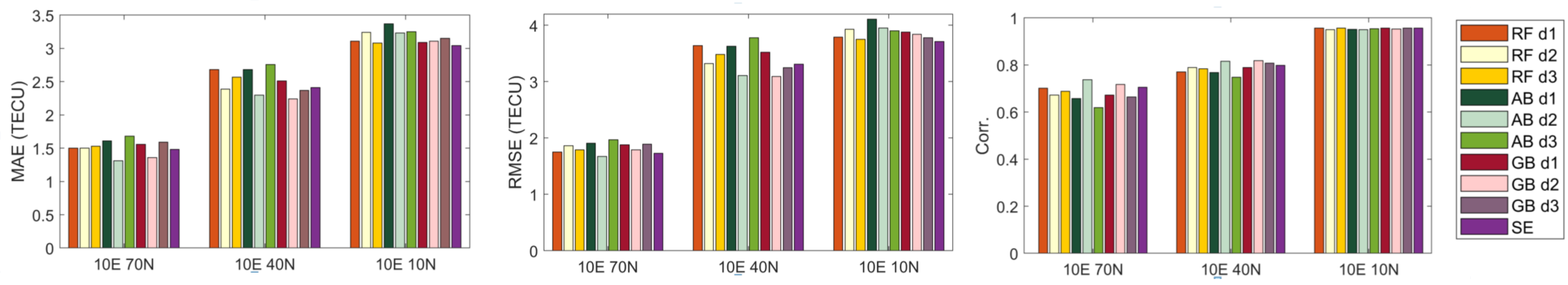




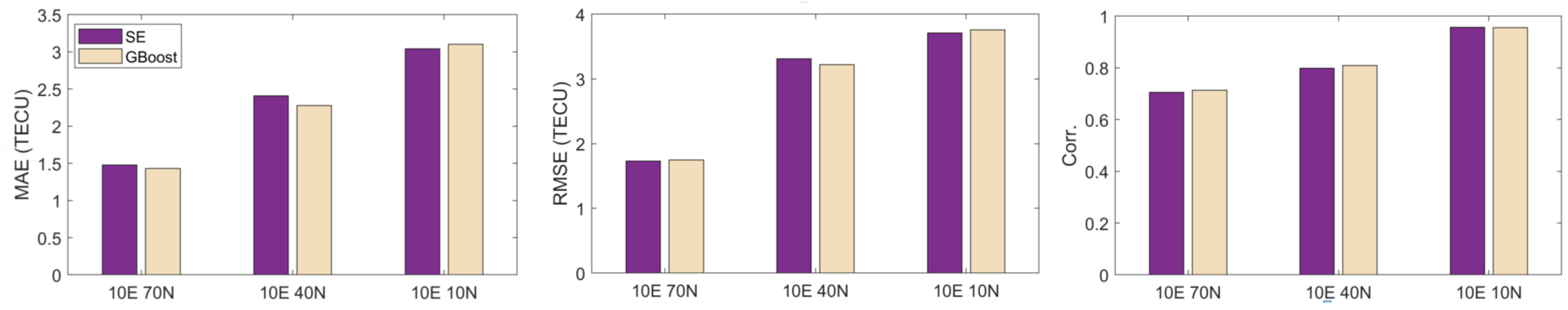


Results: September 2017 space weather events

Ensemble members within the super-ensemble model (SE)



Models for quantifying uncertainties of VTEC prediction: SE and GBoost



Conclusion

- The uncertainty information defines the reliability and precision of VTEC predictions.
- Uncertainty quantification allows to assess the trustworthiness of predictions.
- Ground-truth VTEC mostly within predicted confidence intervals for space weather events.
- Higher contribution from geomagnetic-related input features to VTEC prediction during storm.

Thank you for your attention!

randa.natras@tum.de