APPLY NOISE FILTERS FOR BETTER FORECAST PERFORMANCE IN MACHINE LEARNING

Nhung Le*[1,2,3], Benjamin Männel[1], Randa Natras[4], Pierre Sakic[5], Zhiguo Deng[1], Harald Schuh[1,3]

[1] *GFZ German Research Centre for Geosciences, Germany (Corresponding: nhung@gfz-potsdam.de)*

[2] *Hanoi University of Natural Resources and Environment (HUNRE), Vietnam*

[3] *Technische Universität Berlin, Germany*

[4] *Deutsches Geodätisches Forschungsinstitut der Technischen Universität München (DGFI-TUM), Germany*

[5] *Institut de physique du globe de Paris, Université de Paris, Paris, France*

EGU General Assembly 2022

Machine learning & Geodetic data analytics

GFZ Helmholtz Centre Potsdam

Overview

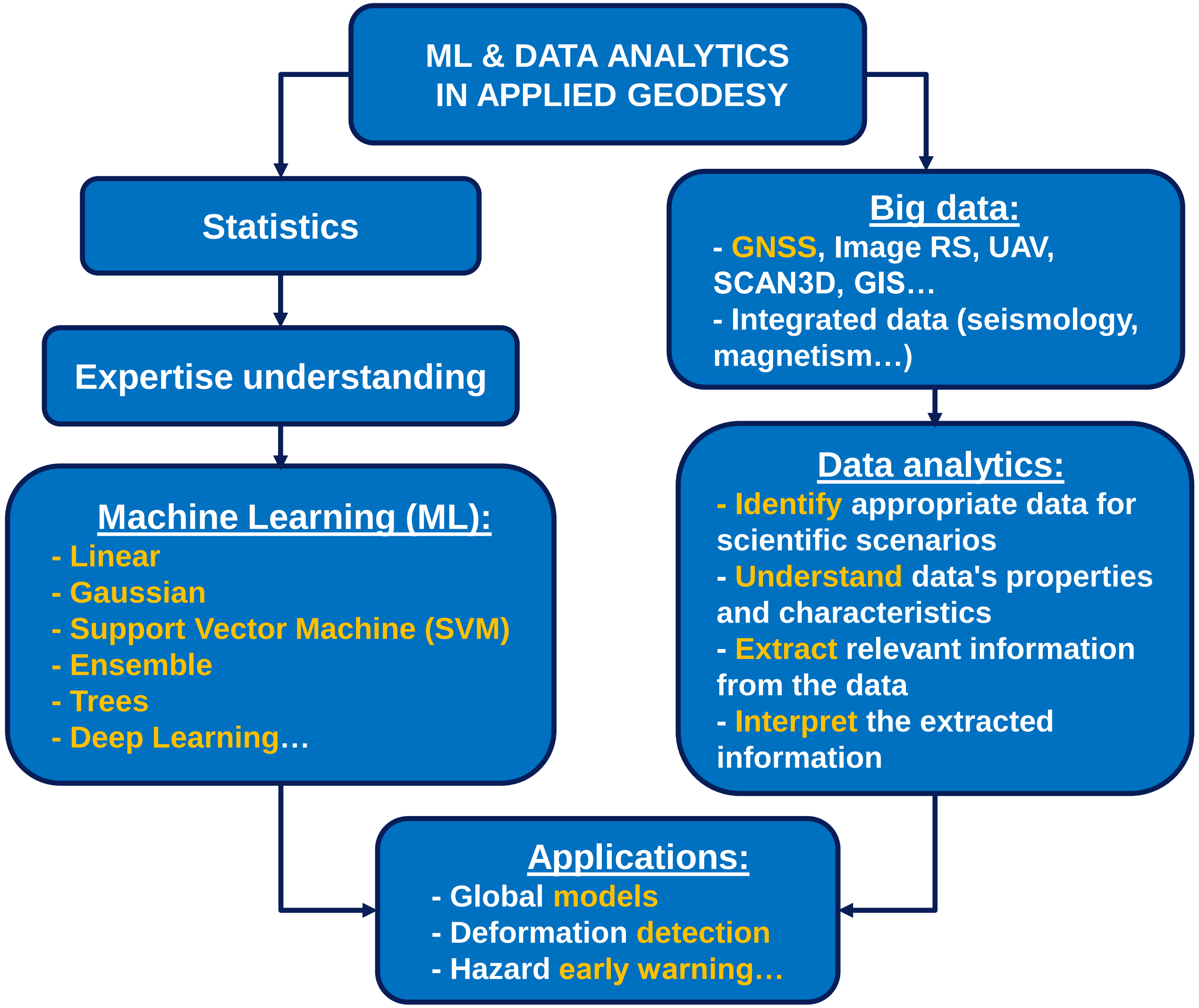

Fig 1. *Machine learning and data analytics in applied geodesy for hazard managements*

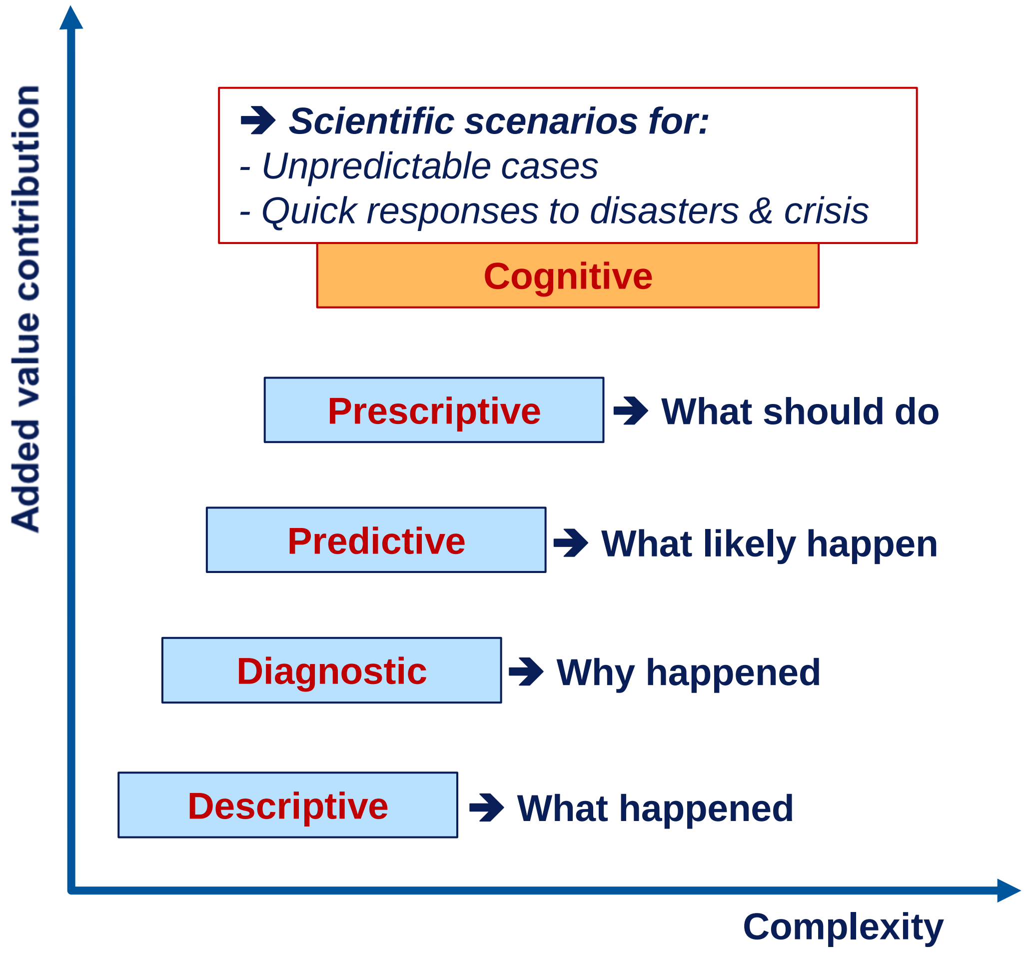

Fig 2. *Progress of data analytics in applied geodesy*

Apply noise filters for better forecast performance of ML

Aim

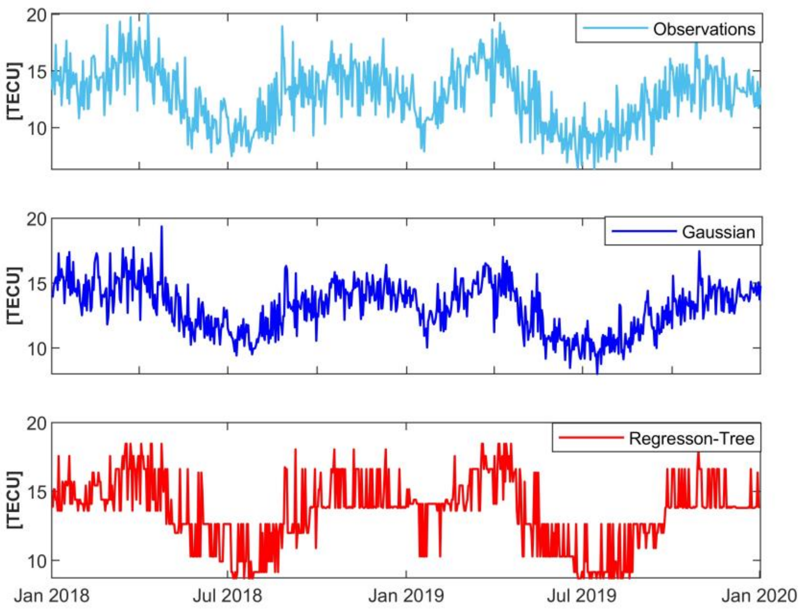

Fig 3. *One-month forecast of VTEC (daily sampling rate) at the IGS station BAKO (Indonesia) using Gaussian and Regression-Coarse Tree algorithms*

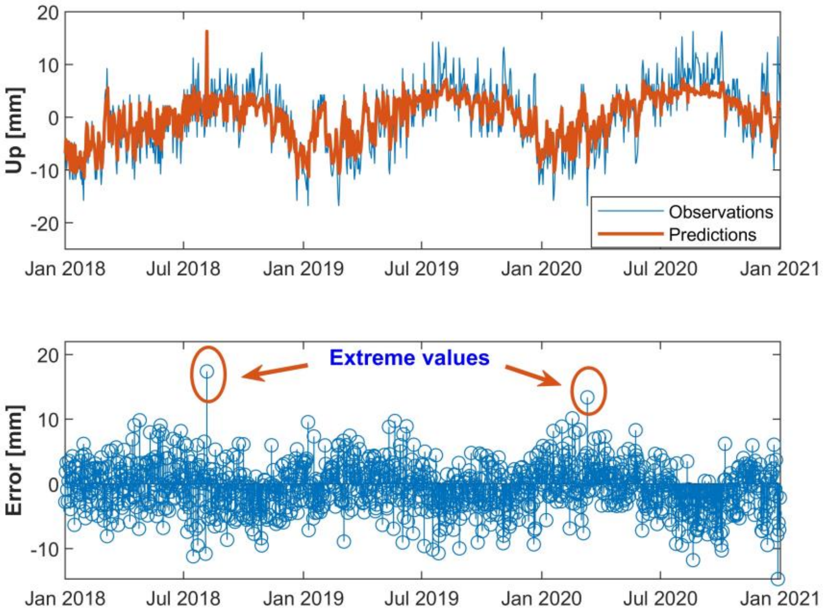

Fig 4. *One-month forecast of crustal motion (daily sampling rate) at the IGS station BADH (Germany) using Quadratic Gaussian algorithm.*

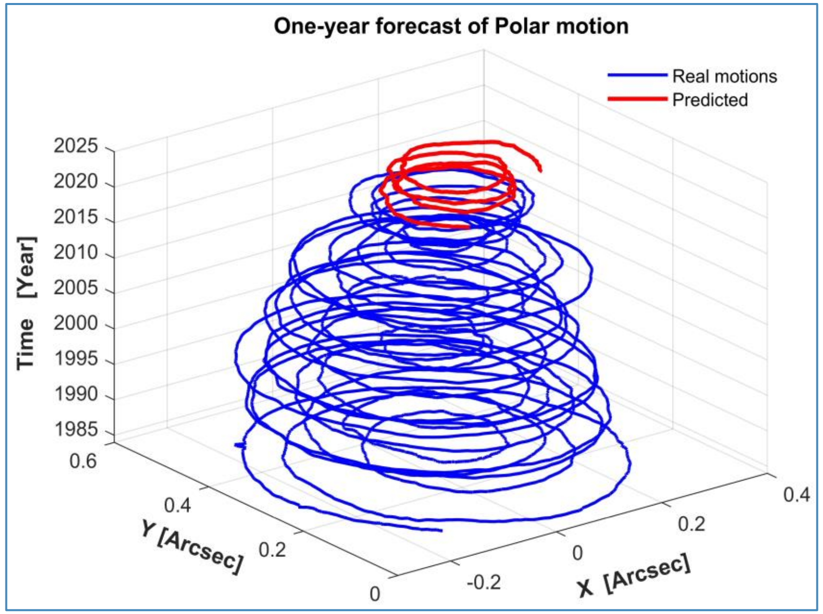

Fig 5. *One-year forecast of Polar motion (daily sampling rate) using Bagged-Tree Ensemble algorithm.*

Smoothen: Cut-offs
in forecast models

Remove: Extreme values/anomalies
in forecast models

Enhence: robustness
in long-term forecasts

Investigation schemes

Approach

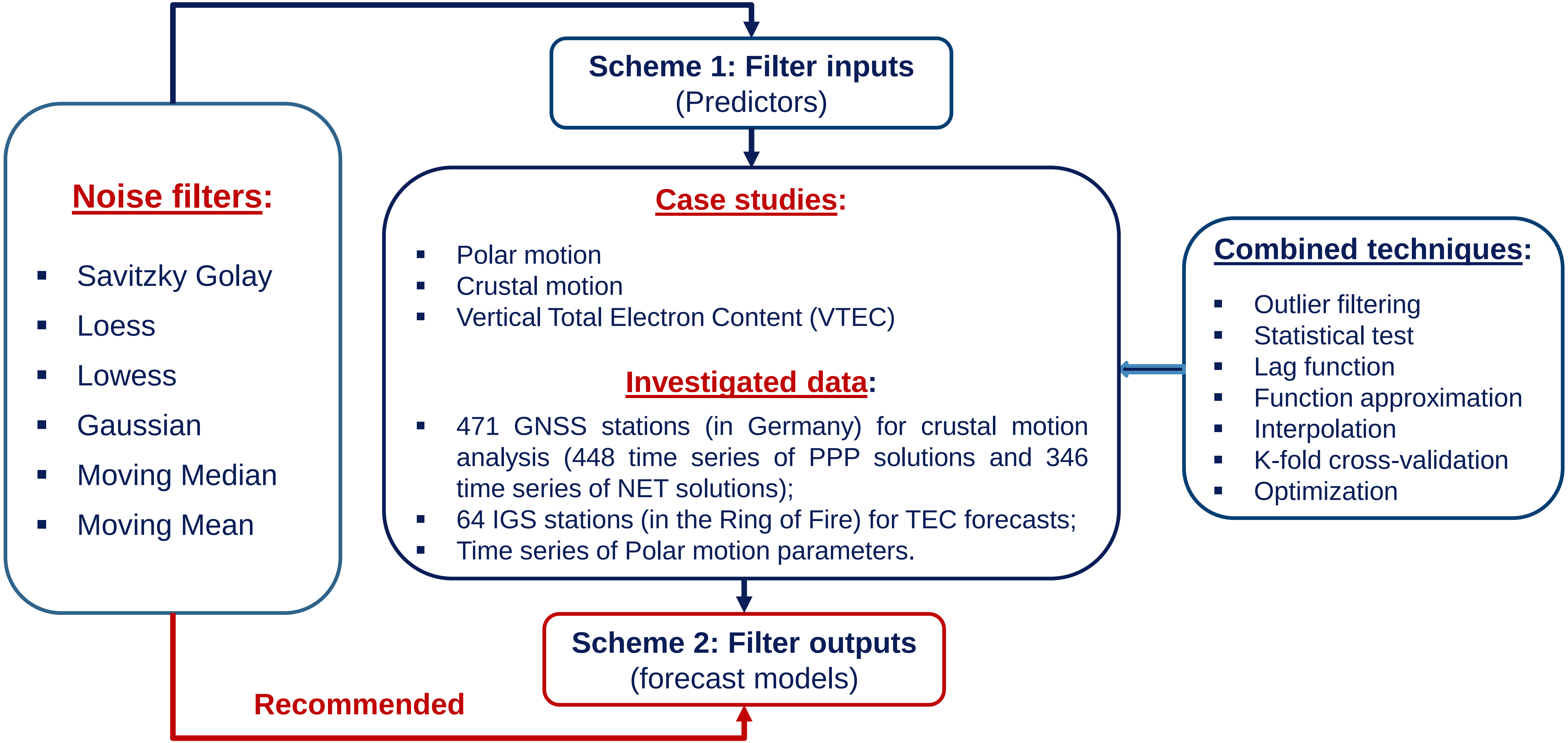

Fig 6. *Flowchart of investigation schemes performed in this study*

Sensitivity of noise filters on geodetic time series

Results

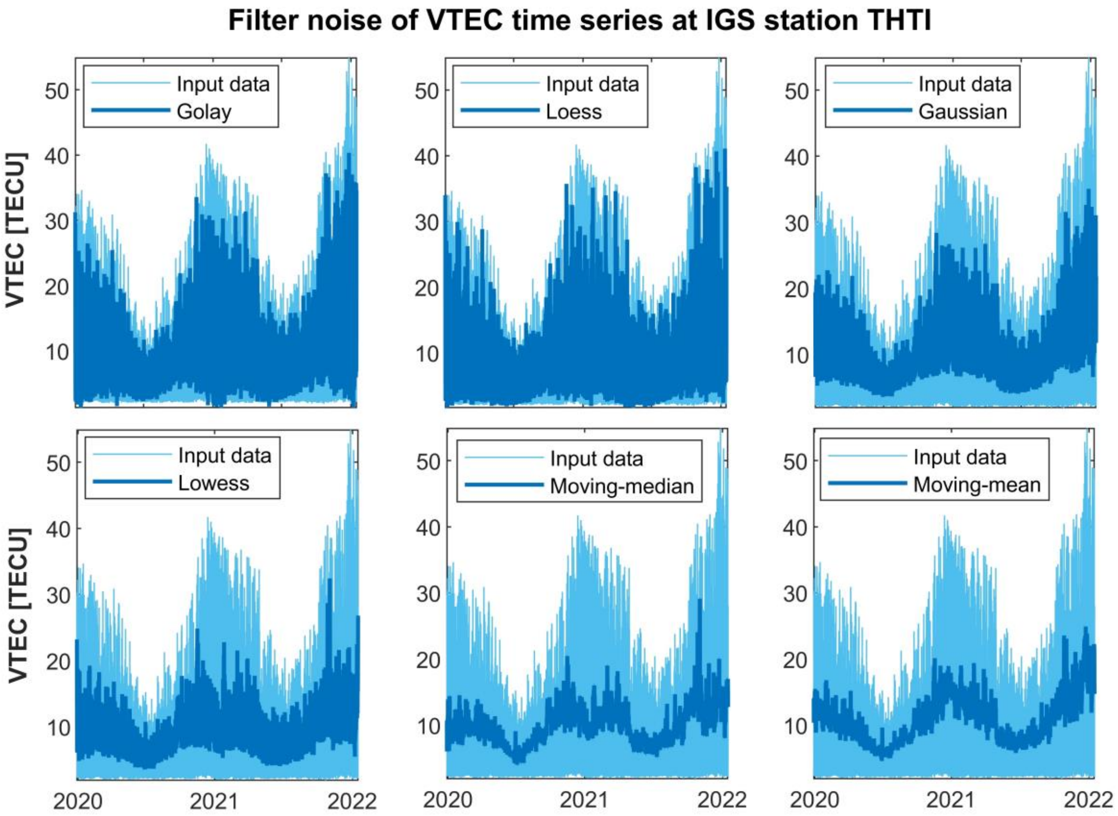

Fig 7. *Filter noise of VTEC time series (hourly sampling rate) at the IGS station THTI (French Polynesia) using 6 filtering algorithms, with the same moving window (24 hours) and statistical threshold (99%); "default" of polynomial function in Savitzky-Golay algorithm is degree 2.*

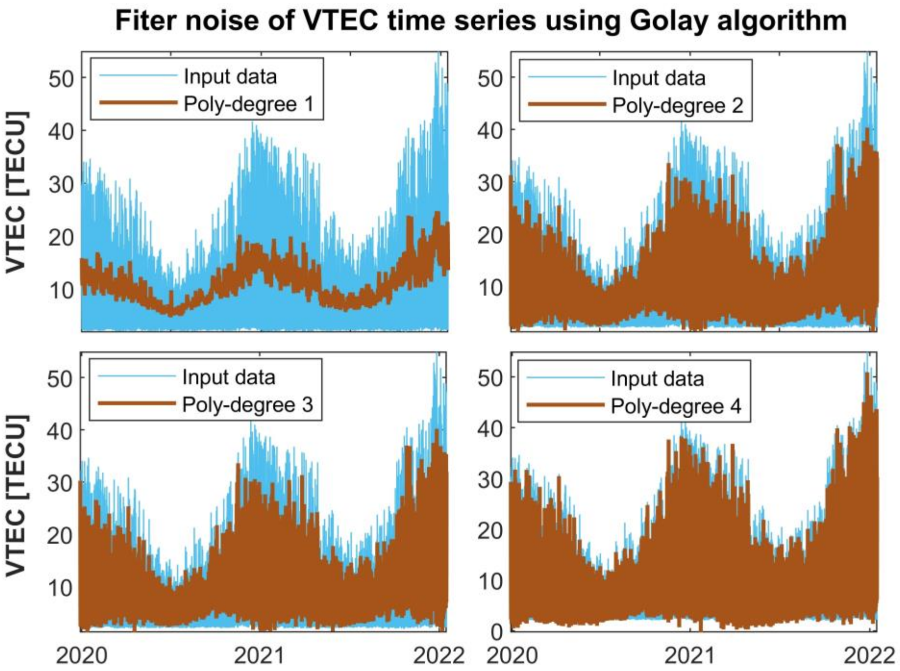

Fig 8. *Filter noise of VTEC time series (hourly sampling rate) at the IGS station THTI using* ***Savitzky-Golay,*** *with* ***polynomial degree 1,2,3, and 4****, moving window 24 hours, statistical threshold 99%.*

1. **Savitzky-Golay** filter is the **most sensitive** and **flexible;**
2. Moving-median and Moving-mean filters are the least effective;
3. Lowess and Moving-median models are biased by anomalies in observations.

Apply noise filters in Earth's crust motion forecast

Results

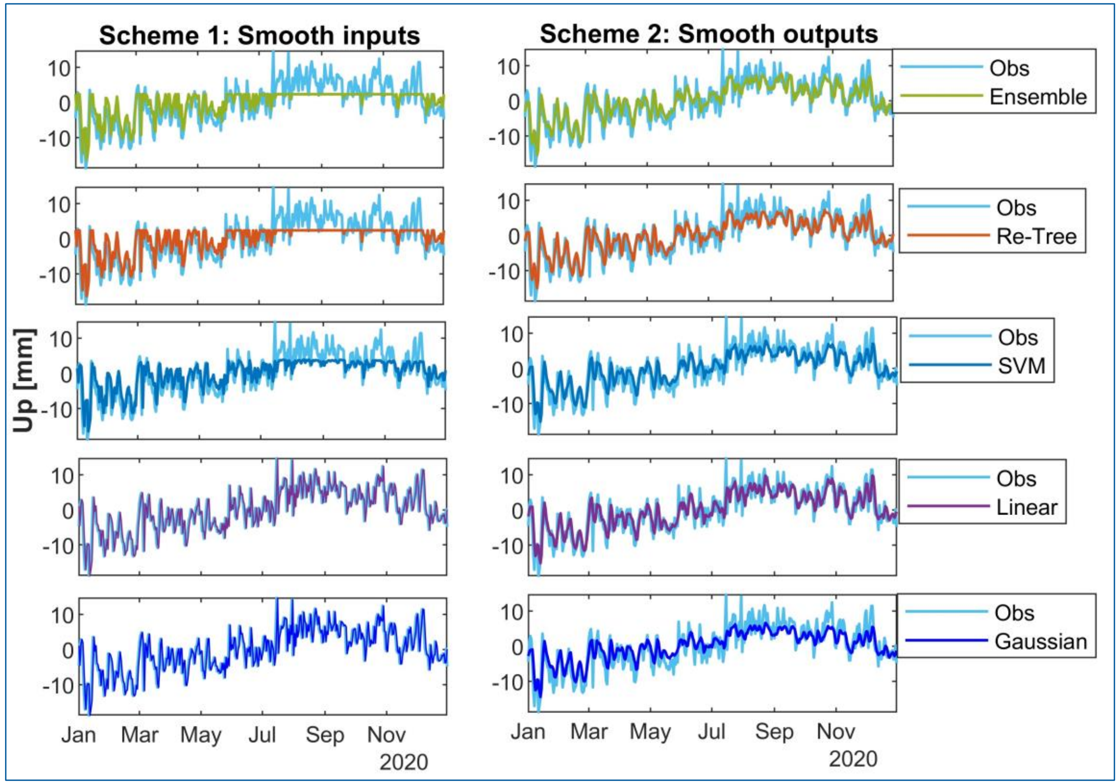

Fig 9. *Filter noise of GNSS time series for the **one-month forecasts** of Up component at IGS station POTS (Germany) in **2 schemes,** using **Savitzky-Golay** (statistical threshold 99%).*

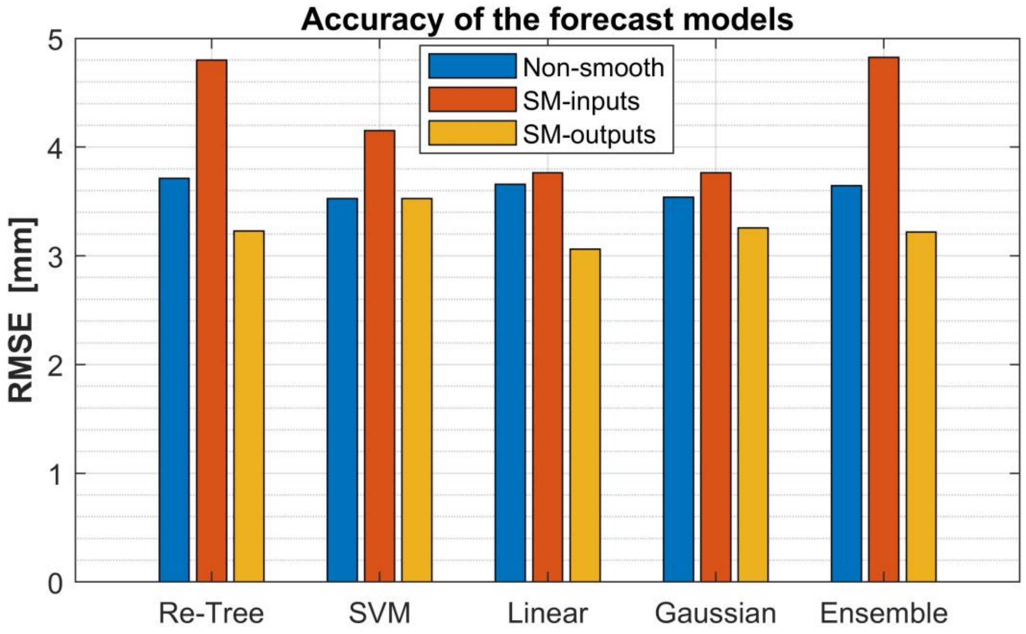

Fig 10. *Comparison of forecast accuracy in three cases: non-applied smooth, smooth inputs (scheme 1), and smooth outputs (scheme 2).*

1. Scheme 1 (Fig 9) can **cause underfitting** (Ensemble, Trees, and SVM) and **overfitting** (Linear and Gaussian).
2. In three investigated cases (Fig 10), scheme 2 gets the **highest accuracy** and can **overcome** both **underfitting** and **overfitting.**

Apply noise filters in VTEC forecast

Results

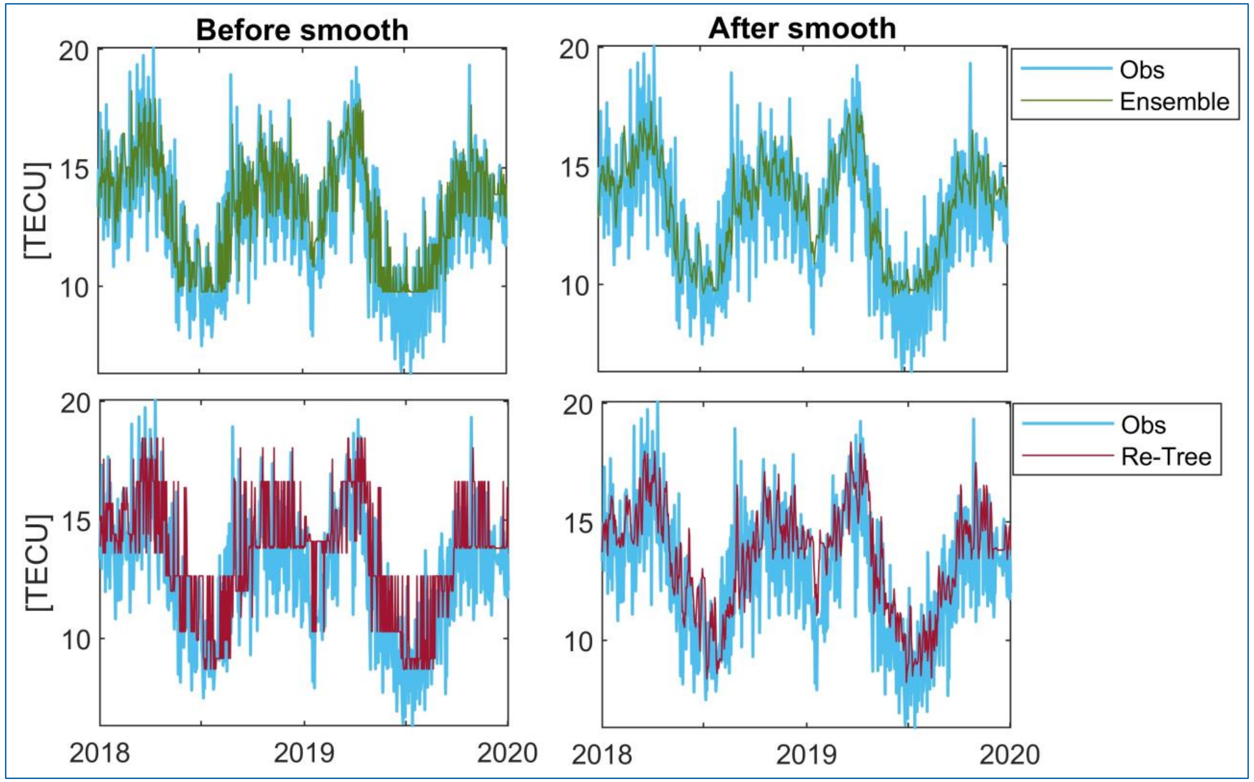

Fig 11. *Filter noise of the* ***one-month forecast*** *models of VTEC at the IGS station BAKO (Indonesia), using* ***Savitzky-Golay*** *at statistical threshold 99%.*

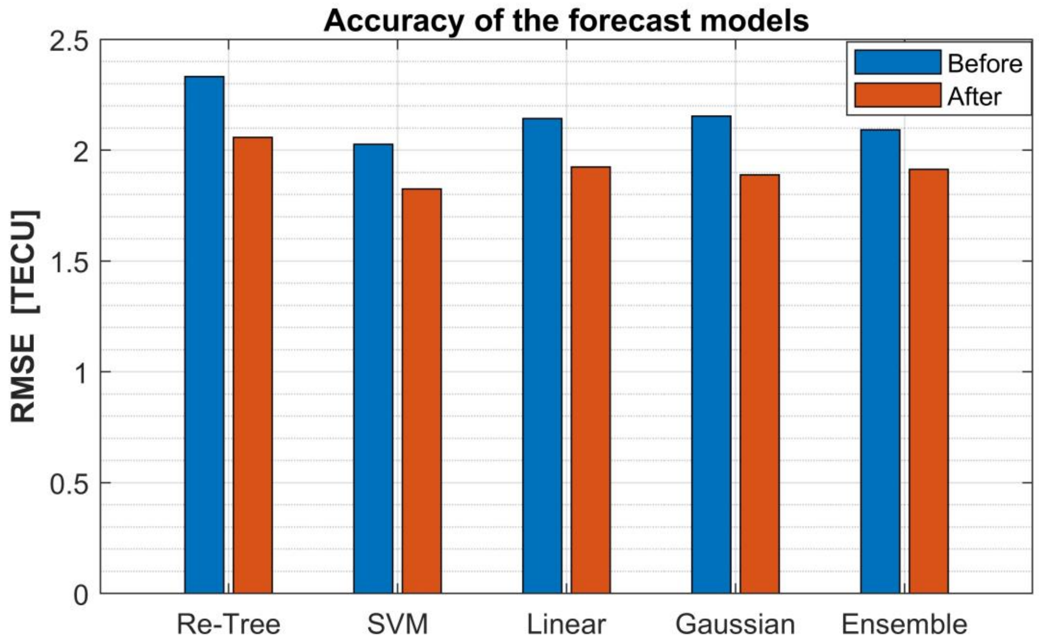

Fig 12. *Accuracy of forecast models* ***before*** *and* ***after*** *smoothed.*

Smoothen cut-offs in the models based on Regression-tree algorithms, and **improve forecast perfomance.**

Apply noise filters in Polar motion parameter forecast

Results

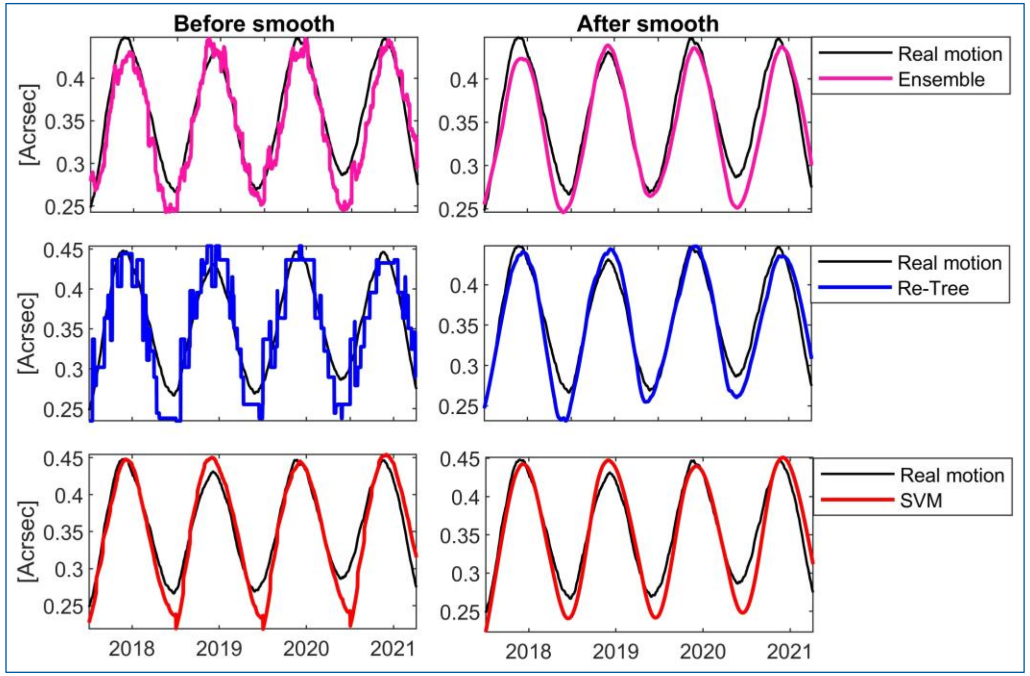

Fig 13. *Filter noise of the **one-year forecast** models (for y component) using **Savitzky-Golay** with polynomial degree 3, moving windows of 90, 180, and 210 days for Bagged-tree ensemble, Regression Trees, SVM, respectively.*

➡ Noise filters can **remove anomalies, extreme values** and **reduce variations** in the forecast models.

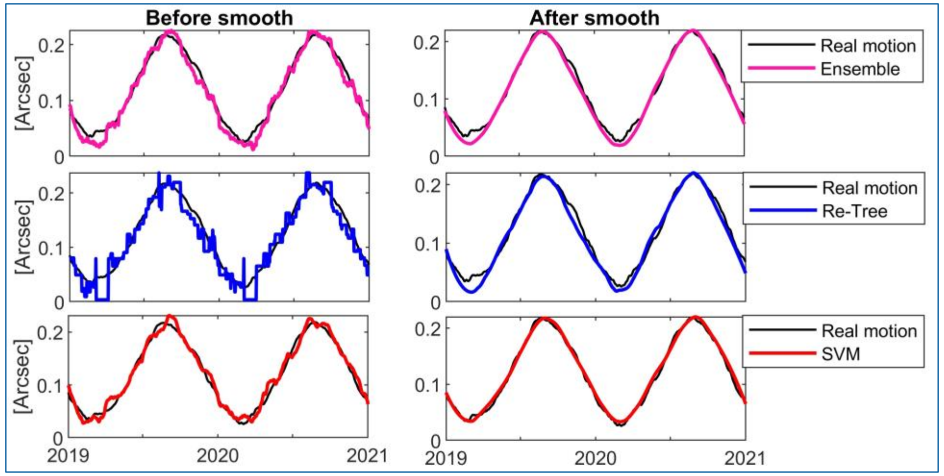

Fig 14. *Filter noise of the **one-month forecast** models (for x component) using **Savitzky-Golay** at statistical threshold 99%.*

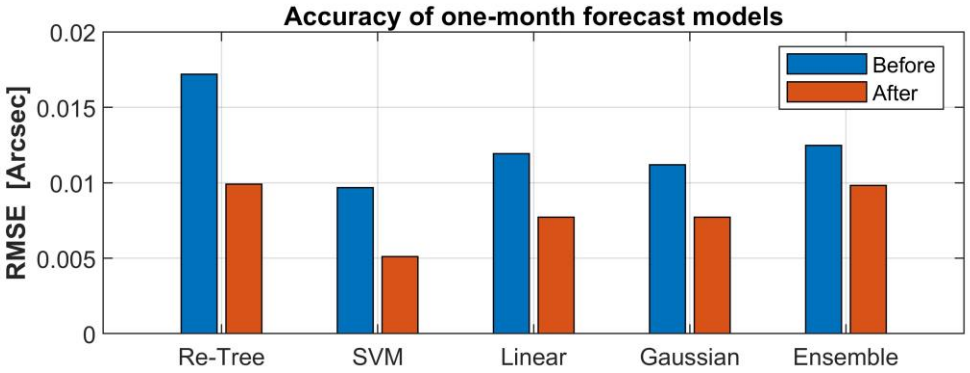

Fig 15. *Accuracy of the **one-month forecast** models (for x component) **before** and **after** smoothed.*

Apply noise filters in geodetic time series analysis

Discussion

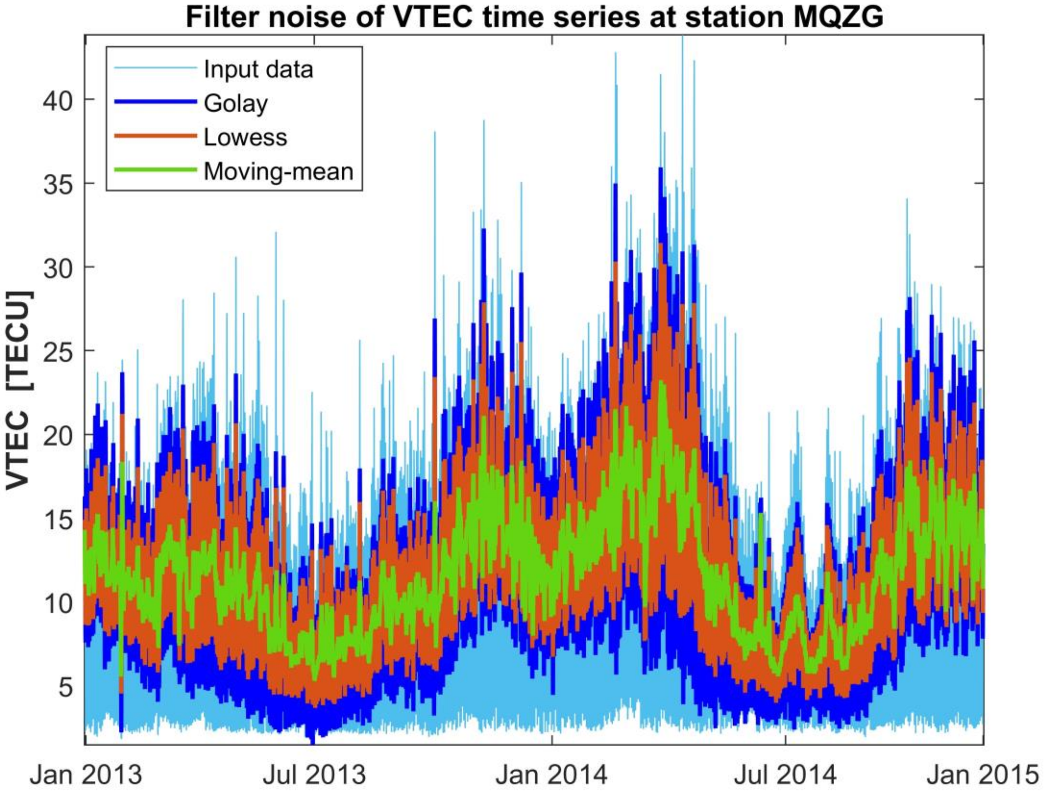

Fig 16. *Filter noise of VTEC time series (hourly sampling rate) at the IGS station MQZG (New Zealand) using* ***Golay of degree 2, Lowess,*** *and* ***Moving-mean*** *algorithms, with statistical threshold 99%, moving window 24 hours.*

Problems:

1. Applying different noise filters gives different patterns (Fig. 16). Thus, **data characteristics** might be **affected** after noise filtering.

2. Noise filtering can **mitigate anomalies** and **variation amplitude** in time series, but might also **lose valuable information** (e.g. in deformation analysis).

Solutions:

1. Identify firmly **data characteristics** and **verify anomalies** in time series before filtering noise;

2. Tune **moving window sizes** via trials to get the best-suited one;

3. Select **statistical thresholds** (95%, 97%, 99%...) based on the required accuracy of forecast models.

EGU General Assembly 2022

Summary & Outlook

GFZ Helmholtz Centre Potsdam

Conclusion

Summary:

1. **Savitzky-Golay** filter is the **most sensitive** and **flexible** (via variety options of polynomial degrees and sliding windows)**,** while Moving-median and Moving-mean filters are the least effective;
2. In Machine Learning, noise filtering should be applied in **scheme 2;**
3. Efficiency of noise filters on the **Polar motion forecast** models is **highest;**
4. Noise filters are **more sensitive** in the forecast models based on **Regression-Tree** and **SVM** algorithms but **less sufficient** on **Ensemble** and **Gaussian.**

Outlook:

1. **Extend investigations** in forecast models based on **deep learning**;
2. Research further **potentials of combination** of noise filtering and other hyperparameter tuning techniques in Machine learning;
3. **Apply** noise filtering techniques for **global forecast models** based on Machine learning.

Acknowledgements

The authors gratefully acknowledge:

- International GNSS Service (IGS) center;
- NOAA space weather prediction centre, USA;
- The world data centre for geomagnetism, Kyoto, Japan;
- Satellitenpositionierungsdienst der Deutschen Landesvermessung (SAPOS®), Germany;
- Space weather live, Belgium;
- The IERS (International Earth Rotation and Reference System services);
- MathWorks (MATLAB® Release 2022a);
- Dr. Gopi's GPS-TEC software 3.03, Release 2021 (Indian Institute of Geomagnetism), India.

for providing data, software, and open-code sources.

The authors would like to thank colleagues at GFZ German Research Centre for Geosciences; the Vietnamese AI4E group for constructive discussion and valuable suggestion.

Thank you!

Nhung Le*[1,2,3], Benjamin Männel[1], Randa Natras[4], Pierre Sakic[5], Zhiguo Deng[1], Harald Schuh[1,3]

[1]*GFZ German Research Centre for Geosciences, Germany (Corresponding: nhung@gfz-potsdam.de)*

[2]*Hanoi University of Natural Resources and Environment (HUNRE), Vietnam*

[3]*Technische Universität Berlin, Germany*

[4]*Deutsches Geodätisches Forschungsinstitut der Technischen Universität München (DGFI-TUM), Germany*

[5]*Institut de physique du globe de Paris, Université de Paris, Paris, France*