Koninklijk Nederlands
Meteorologisch Instituut
Ministerie van Infrastructuur en Waterstaat

Improving MOS Random Forests for Post-processing Extreme Wind Gust Forecasts

Bastien François[1], Harun Kivril[1], Maurice Schmeits[1], Kirien Whan[2]

[1]Royal Netherlands Meteorological Institute (KNMI), R&D Weather and Climate Modelling, De Bilt, Netherlands
[2]Royal Netherlands Meteorological Institute (KNMI), R&D Observations and Data technology (RDWD), De Bilt, Netherlands

15th International Meeting on Statistical Climatology (IMSC 2024)
June 26th 2024

Introduction & objectives

- **Ensemble forecasts** for extreme events: can be **biased** and **under/over-dispersed** compared to obs.

- **Statistical post-processing**: correcting for systematic errors and offering better information about forecast uncertainty.

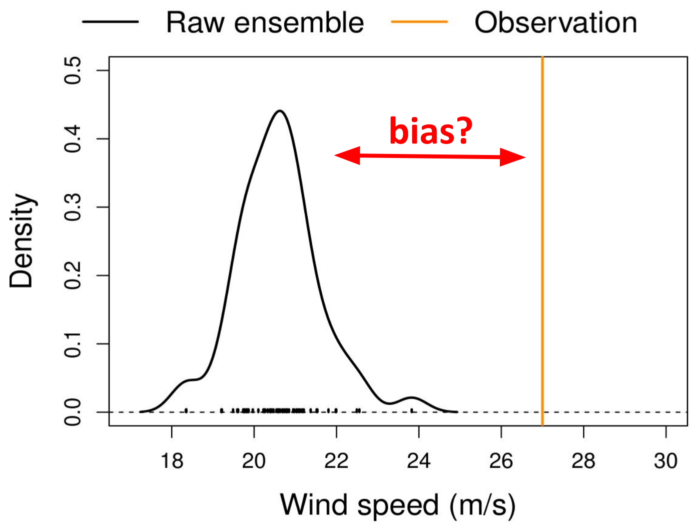


Figure: Biased forecast of 6-hourly maximum wind gusts (ECMWF-IFS ENS) and the observation at IJmuiden.

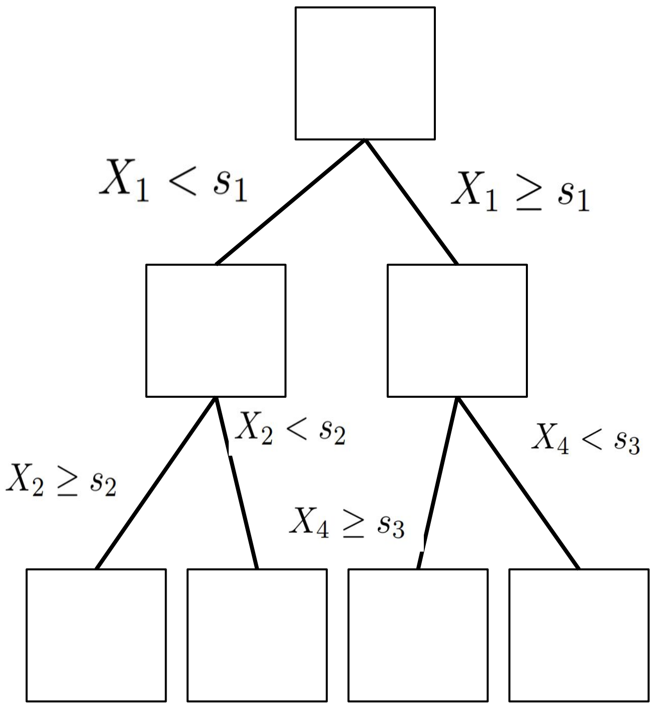
Introduction & objectives

- **Ensemble forecasts** for extreme events: can be **biased** and **under/over-dispersed** compared to obs.

- **Statistical post-processing**: correcting for systematic errors and offering better information about forecast uncertainty.

- **Machine Learning: Forest-based methods**
  - use an ensemble of decision trees to adapt the post-processing depending on covariates. (Schlosser et al., 2019, Taillardat et al., 2019)
  - post-process bulk of the data well… **and extremes?** (Schulz et al., 2022)

Inputs: Raw ensemble (biased)



Outputs: Post-processed forecast

Forest-based methods

Three main methods… with different **extrapolation capability!**

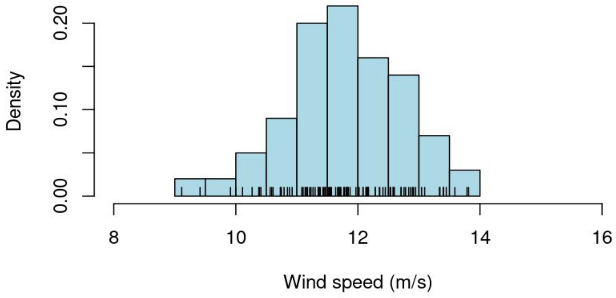
Forest-based methods

Three main methods… with different **extrapolation capability!**

Generalized Regression Forests (GRF)

Predict collection of **observed quantiles**.

No extrapolation.



Taillardat et al., 2019

Forest-based methods

Three main methods… with different **extrapolation capability!**

Generalized Regression Forests (GRF)

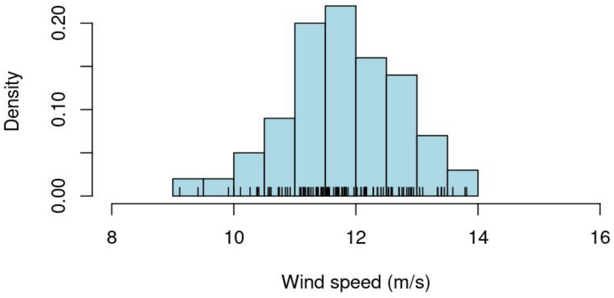
Predict collection of **observed quantiles**.

No extrapolation.

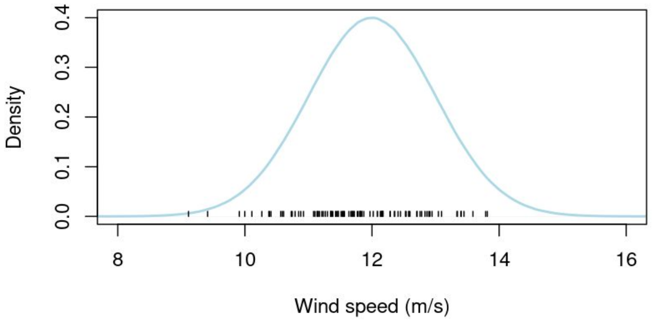
Distributional Regression Forests (DRF)

Predict the parameters of a **probability distribution**.

Limited extrapolation.



Taillardat et al., 2019

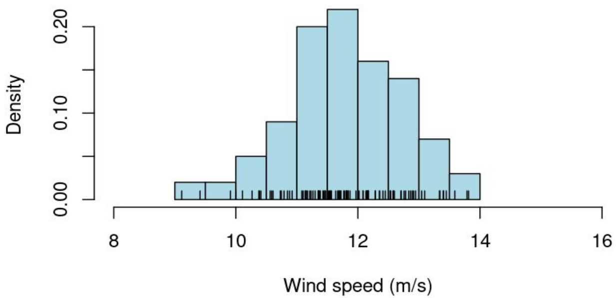


Schlosser et al., 2019

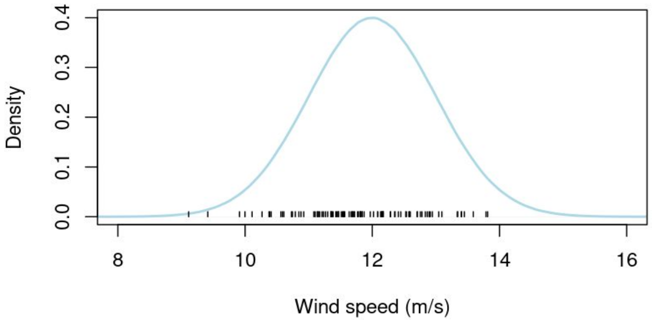
Forest-based methods

Three main methods… with different **extrapolation capability!**

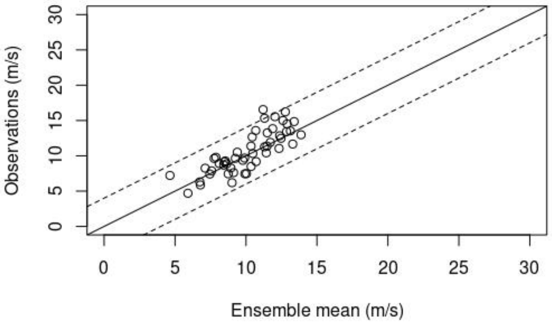
Generalized Regression Forests (GRF)	Distributional Regression Forests (DRF)	MOS Random Forests (MOS-RF)
Predict collection of **observed quantiles**. No extrapolation.	Predict the parameters of a **probability distribution**. Limited extrapolation.	Predict the **parameters** of a **MOS model**. Extrapolation possible.



Taillardat et al., 2019



Schlosser et al., 2019



Muschinski et al., 2023

Forest-based methods

Three main methods… with different **extraptolation capability!**

Generalized Regression Forests (GRF)

Predict collection of **observed quantiles**.

No extrapolation.

Distributional Regression Forests (DRF)

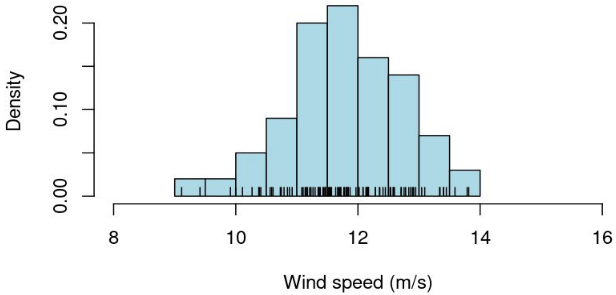
Predict the parameters of a **probability distribution**.

Limited extrapolation.

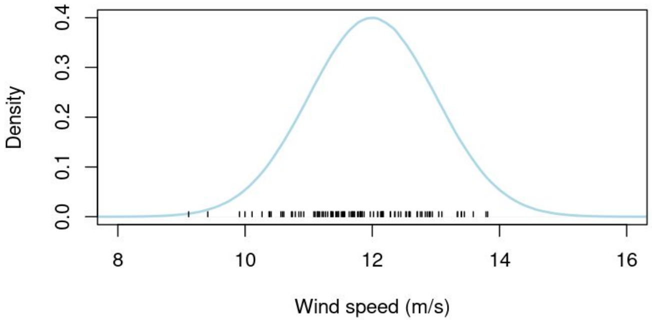
MOS Random Forests (MOS-RF)

Predict the **parameters** of a **MOS model**.

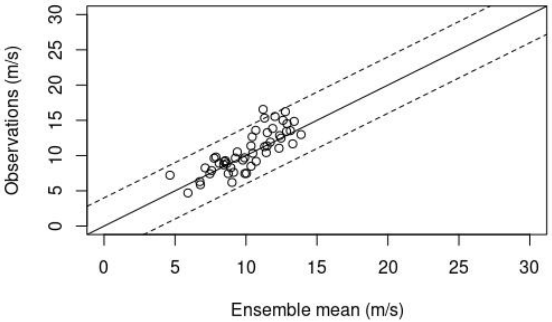
Extrapolation possible.



Taillardat et al., 2019



Schlosser et al., 2019



Muschinski et al., 2023

… but assume one MOS model: **suitable for wind gust extremes**?

Adding more flexibility to better model extremes

<u>MOS-RF method:</u> Adapt MOS model to weather situations.

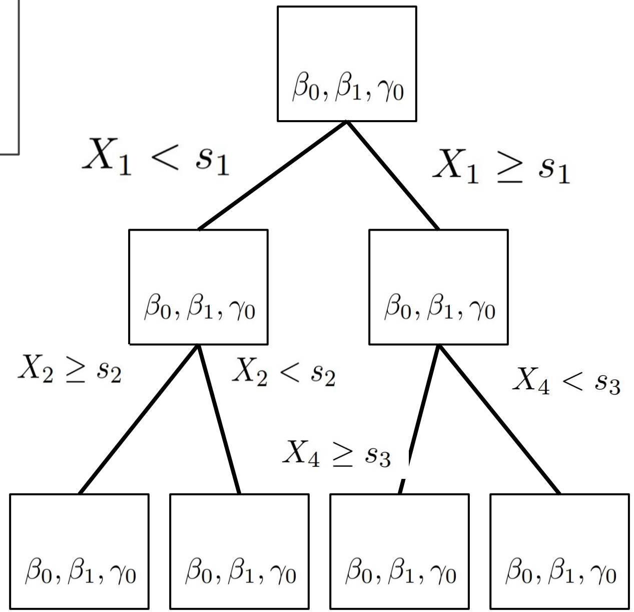
MOS model

$$y \sim \mathcal{N}_0(\mu, \sigma^2)$$

$$\text{with } \mu = \beta_0 + \beta_1 \times x_1$$

$$\text{and } log(\sigma) = \gamma_0 \times x_1$$

<u>Inputs:</u> Raw ensemble (biased)



<u>Outputs:</u> Adjusted distribution

Adding more flexibility to better model extremes

<u>MOS-RF method:</u> Adapt MOS model to weather situations.

<u>"Regime-switching" approach:</u>

- Fit a base MOS-RF model.

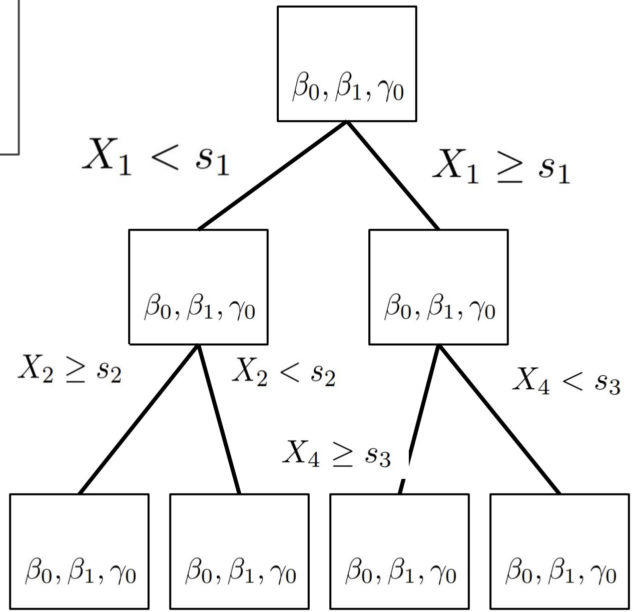
MOS model

$$y \sim \mathcal{N}_0(\mu, \sigma^2)$$

$$\text{with } \mu = \beta_0 + \beta_1 \times x_1$$

$$\text{and } log(\sigma) = \gamma_0 \times x_1$$

<u>Inputs:</u> Raw ensemble (biased)



<u>Outputs:</u> Adjusted distribution

Adding more flexibility to better model extremes

MOS-RF method: Adapt MOS model to weather situations.

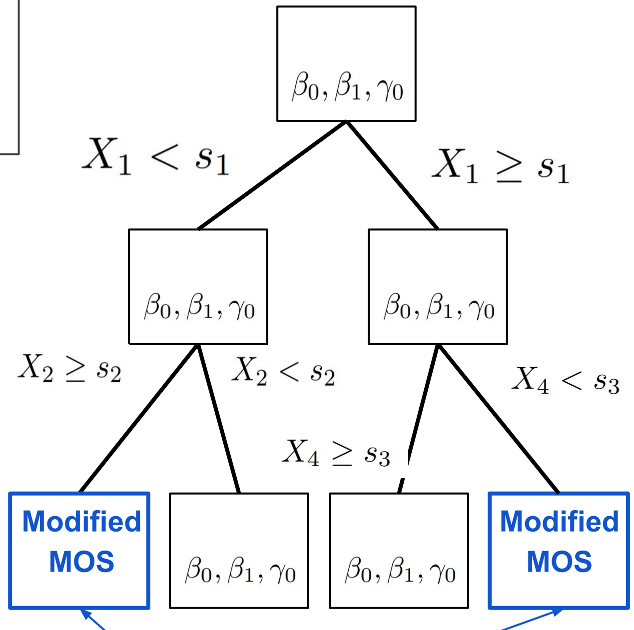
MOS model

$$y \sim \mathcal{N}_0(\mu, \sigma^2)$$
$$\text{with } \mu = \beta_0 + \beta_1 \times x_1$$
$$\text{and } log(\sigma) = \gamma_0 \times x_1$$

"Regime-switching" approach:

- Fit a base MOS-RF model.
- **Modify** the MOS model for **terminal nodes** associated with high wind gust forecasts:

Inputs: Raw ensemble (biased)



Regime-switching when wind gust forecast is high (median ensemble mean above a ***threshold***)

Adding more flexibility to better model extremes

MOS-RF method: Adapt MOS model to weather situations.

MOS model

$$y \sim \mathcal{N}_0(\mu, \sigma^2)$$
$$\text{with } \mu = \beta_0 + \beta_1 \times x_1$$
$$\text{and } log(\sigma) = \gamma_0 \times x_1$$

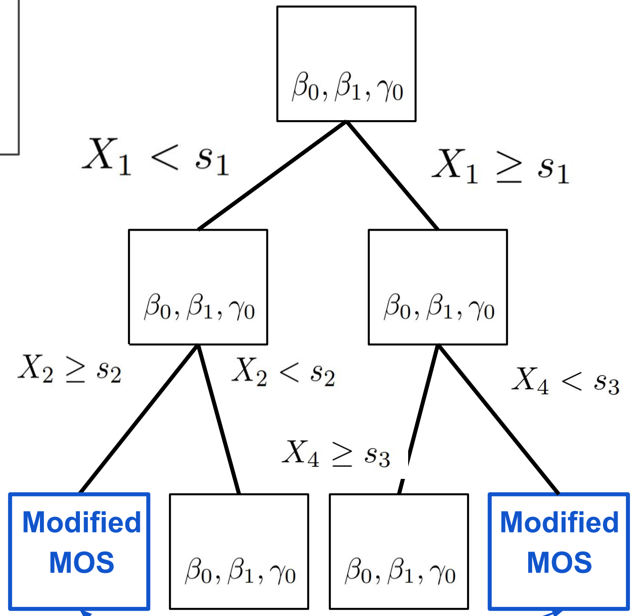
"Regime-switching" approach:

- Fit a base MOS-RF model.
- **Modify** the MOS model for **terminal nodes** associated with high wind gust forecasts:
    - **adding covariates** related to extremes

Add cov.

$$y \sim \mathcal{N}_0(\mu, \sigma^2)$$
$$\text{with } \mu = \beta_0 + \beta_1 \times x_1 + \boldsymbol{\beta_2} \times \boldsymbol{x_2}$$
$$log(\sigma) = \gamma_0 \times x_1$$

Inputs: Raw ensemble (biased)



Regime-switching when wind gust forecast is high (median ensemble mean above a ***threshold***)

Adding more flexibility to better model extremes

MOS-RF method: Adapt MOS model to weather situations.

MOS model

$$y \sim \mathcal{N}_0(\mu, \sigma^2)$$
$$\text{with } \mu = \beta_0 + \beta_1 \times x_1$$
$$\text{and } log(\sigma) = \gamma_0 \times x_1$$

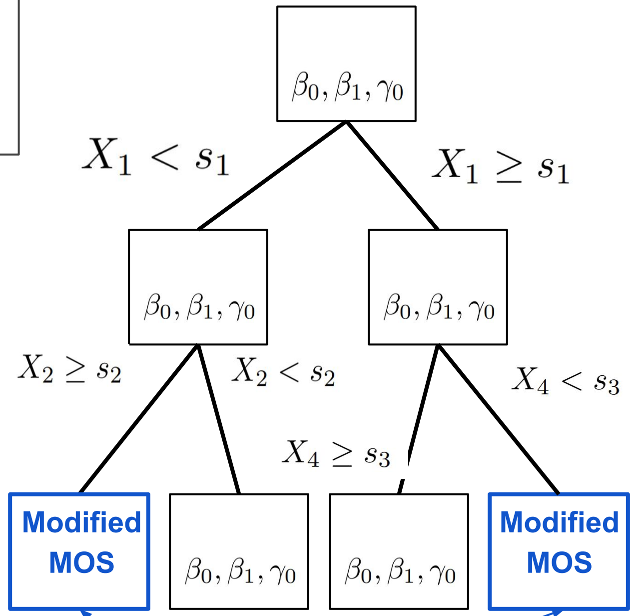
"Regime-switching" approach:

- Fit a base MOS-RF model.
- **Modify** the MOS model for **terminal nodes** associated with high wind gust forecasts:
  - **adding covariates** related to extremes
  - using **other statistical distributions**

Truncated GEV

$$y \sim \mathcal{GEV}_0(\mu, \sigma, \xi)$$
$$\text{with } \mu = \beta_0 + \beta_1 \times x_1$$
$$log(\sigma) = \gamma_0 \times x_1$$
$$\xi = \delta_0$$

Inputs: Raw ensemble (biased)



Regime-switching when wind gust forecast is high (median ensemble mean above a ***threshold***)

Adding more flexibility to better model extremes

MOS-RF method: Adapt MOS model to weather situations.

MOS model

$$y \sim \mathcal{N}_0(\mu, \sigma^2)$$
$$\text{with } \mu = \beta_0 + \beta_1 \times x_1$$
$$\text{and } log(\sigma) = \gamma_0 \times x_1$$

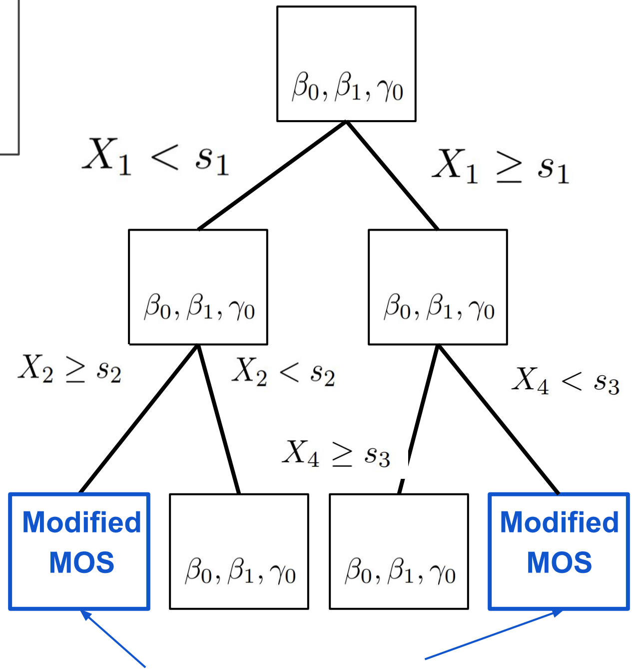
"Regime-switching" approach:

- Fit a base MOS-RF model.
- **Modify** the MOS model for **terminal nodes** associated with high wind gust forecasts:
  - **adding covariates** related to extremes
  - using **other statistical distributions**
  - Both

Add cov. + Truncated GEV

$$y \sim \mathcal{GEV}_0(\mu, \sigma, \xi)$$
$$\text{with } \mu = \beta_0 + \beta_1 \times x_1 + \boldsymbol{\beta_2} \times \boldsymbol{x_2}$$
$$log(\sigma) = \gamma_0 \times x_1$$
$$\xi = \delta_0$$

Inputs: Raw ensemble (biased)



Regime-switching when wind gust forecast is high (median ensemble mean above a ***threshold***)

Adding more flexibility to better model extremes

MOS-RF method: Adapt MOS model to weather situations.

MOS model

$$y \sim \mathcal{N}_0(\mu, \sigma^2)$$

$$\text{with } \mu = \beta_0 + \beta_1 \times x_1$$

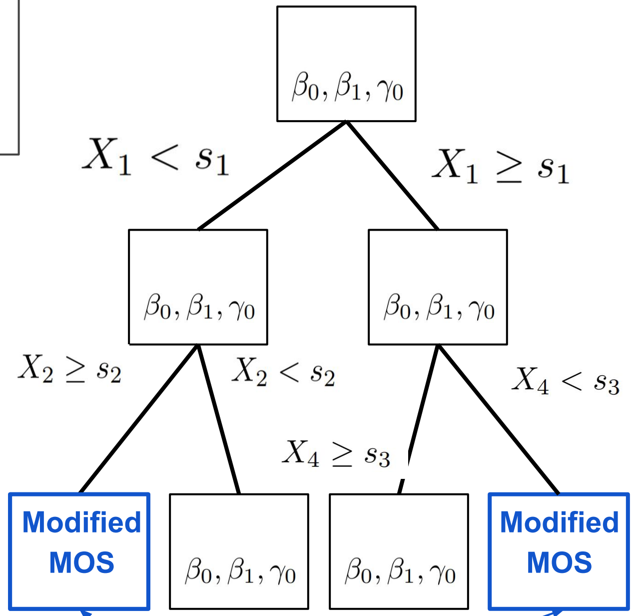
$$\text{and } log(\sigma) = \gamma_0 \times x_1$$

"Regime-switching" approach:

- Fit a base MOS-RF model.
- **Modify** the MOS model for **terminal nodes** associated with high wind gust forecasts:
  - **adding covariates** related to extremes
  - using **other statistical distributions**
  - Both

Advantage: Improve extremes while keeping initial performances for non-extremes.

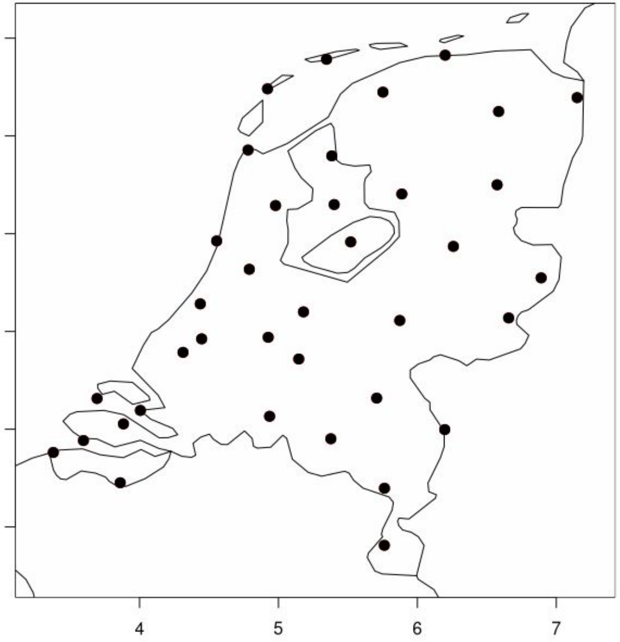
Inputs: Raw ensemble (biased)



Regime-switching when wind gust forecast is high (median ensemble mean above a ***threshold***)

Data and applications

- Data: 6-hourly maximum **wind gusts**
- Model: **ECMWF-IFS ensemble** data (51 members)
- Reference: observations at KNMI stations (37).
- over **Netherlands** during winter 2018-2022.
- Number of covariates: 71.
- Data splitting: train-validation (2018-2021), test (2022).
- **Lead time: 30h**, initialisation time: 00 UTC.
- **Pooling** of stations

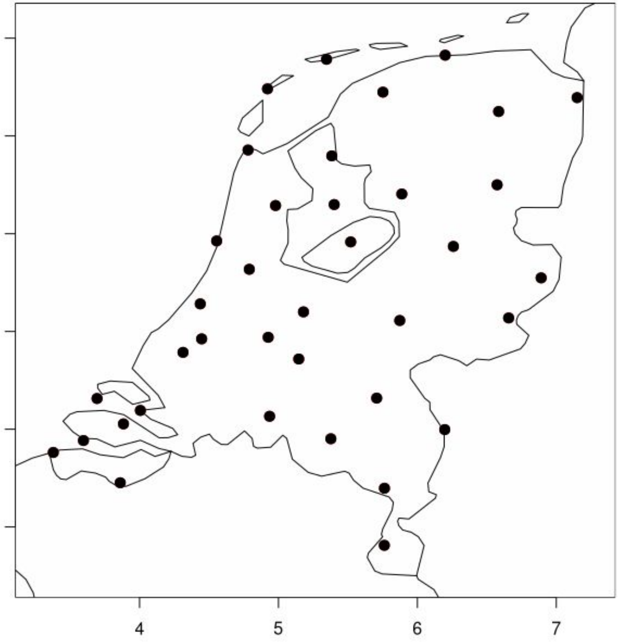


<u>MOS-RF versions:</u>

- **MOS-RF:** base model with truncated Gaussian (cov: ensemble mean)

Data and applications

- Data: 6-hourly maximum **wind gusts**
- Model: **ECMWF-IFS ensemble** data (51 members)
- Reference: observations at KNMI stations (37).
- over **Netherlands** during winter 2018-2022.
- Number of covariates: 71.
- Data splitting: train-validation (2018-2021), test (2022).
- **Lead time: 30h**, initialisation time: 00 UTC.
- **Pooling** of stations



MOS-RF versions:

- **MOS-RF:** base model with truncated Gaussian (cov: ensemble mean)
- Regime-switching:
    - MOS-RF-AddCov to model location (90th percentile of wind gust ensemble).
    - MOS-RF-GEV0 with truncated GEV.
    - **MOS-RF-AddCovGEV0** truncated GEV + additional cov. for location.

Regime-switching: finding the optimal threshold

Select best threshold on validation set:

- Mean CRPSS wrt MOS-RF on modified predictions.

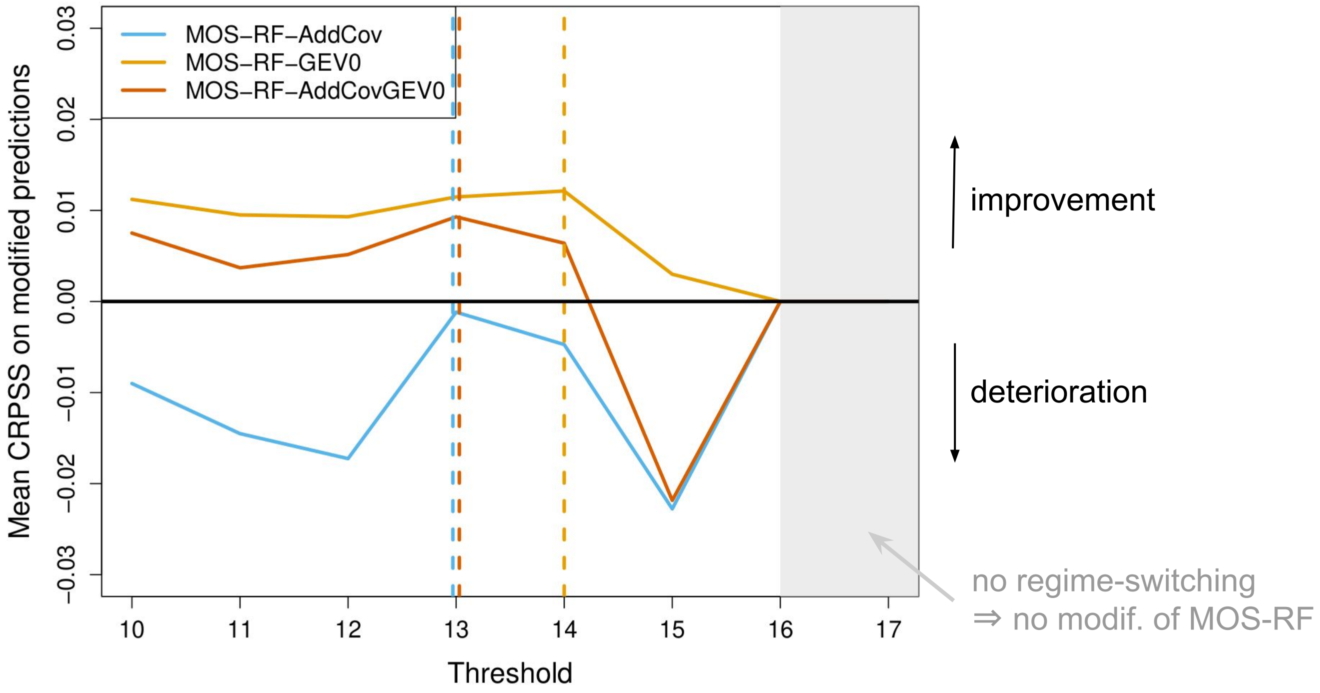


Figure: Mean CRPSS with respect to MOS-RF depending on the threshold.

Results for wind gust extremes

Comparison of **MOS-RF versions** with:

- **GRF** (Taillardat et al., 2019) - **DRF** (Schlosser et al., 2019) - **Raw forecasts**

Metrics: **twCRPS** at stations for events above 21m/s (~97.5th obs. percentiles)

Results for wind gust extremes

Comparison of **MOS-RF versions** with:

- **GRF** (Taillardat et al., 2019) - **DRF** (Schlosser et al., 2019) - **Raw forecasts**

Metrics: **twCRPS** at stations for events above 21m/s (~97.5th obs. percentiles)

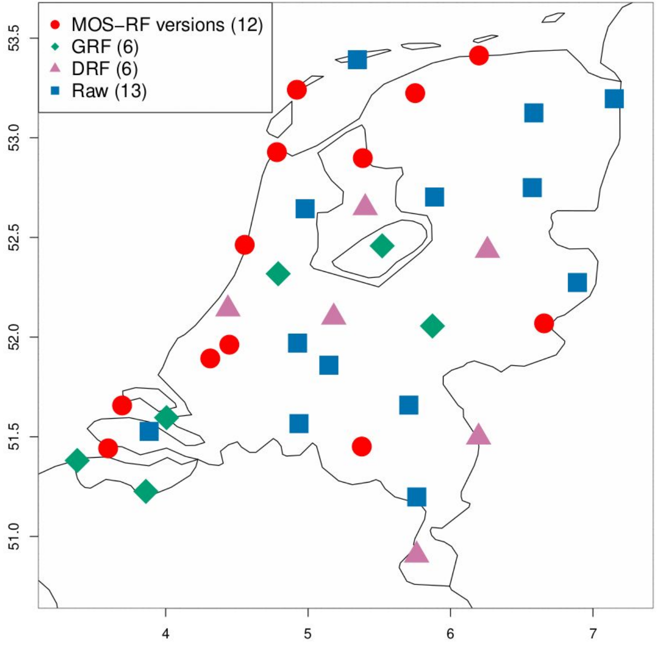


Figure: Post-processing method that performs best at each station.

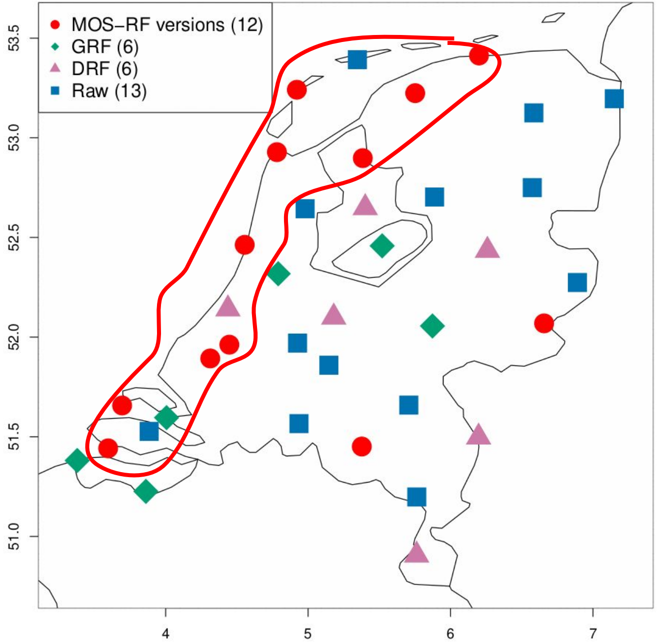
- **MOS-RF versions** better along the coast.
- **Raw forecasts still better** for 1/3 of stations.

Results for wind gust extremes

Comparison of **MOS-RF versions** with:

- **GRF** (Taillardat et al., 2019) - **DRF** (Schlosser et al., 2019) - **Raw forecasts**

Metrics: **twCRPS** at stations for events above 21m/s (~97.5th obs. percentiles)



- **MOS-RF versions** better along the coast.
- **Raw forecasts still better** for 1/3 of stations.

Figure: Post-processing method that performs best at each station.

Results for wind gust extremes

Comparison of **MOS-RF versions** with:

- **GRF** (Taillardat et al., 2019)
- **DRF** (Schlosser et al., 2019)
- **Raw forecasts**

Metrics: **twCRPS** at stations for events above 21m/s (~97.5th obs. percentiles)

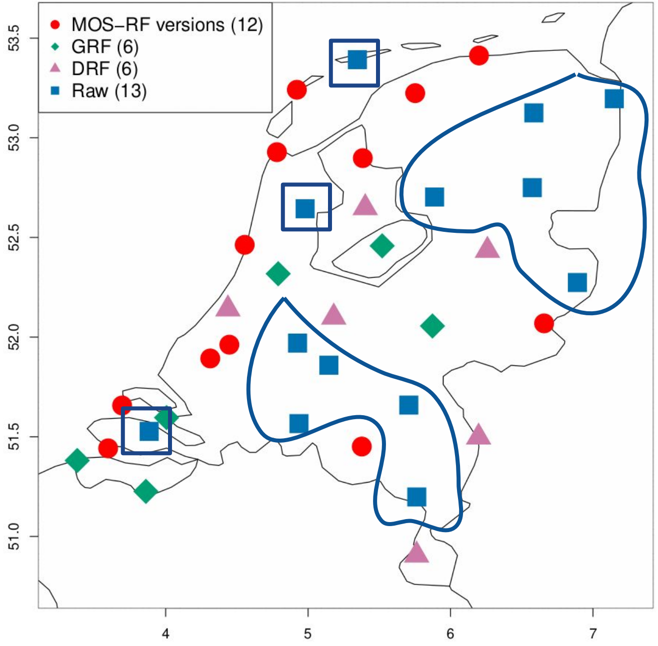


Figure: Post-processing method that performs best at each station.

- **MOS-RF versions** better along the coast.
- **Raw forecasts still better** for 1/3 of stations.

Results for wind gust extremes

Comparison of **MOS-RF versions** with:

- **GRF** (Taillardat et al., 2019)
- **DRF** (Schlosser et al., 2019)
- **Raw forecasts**

Metrics: **twCRPS** at stations for events above 21m/s (~97.5th obs. percentiles)

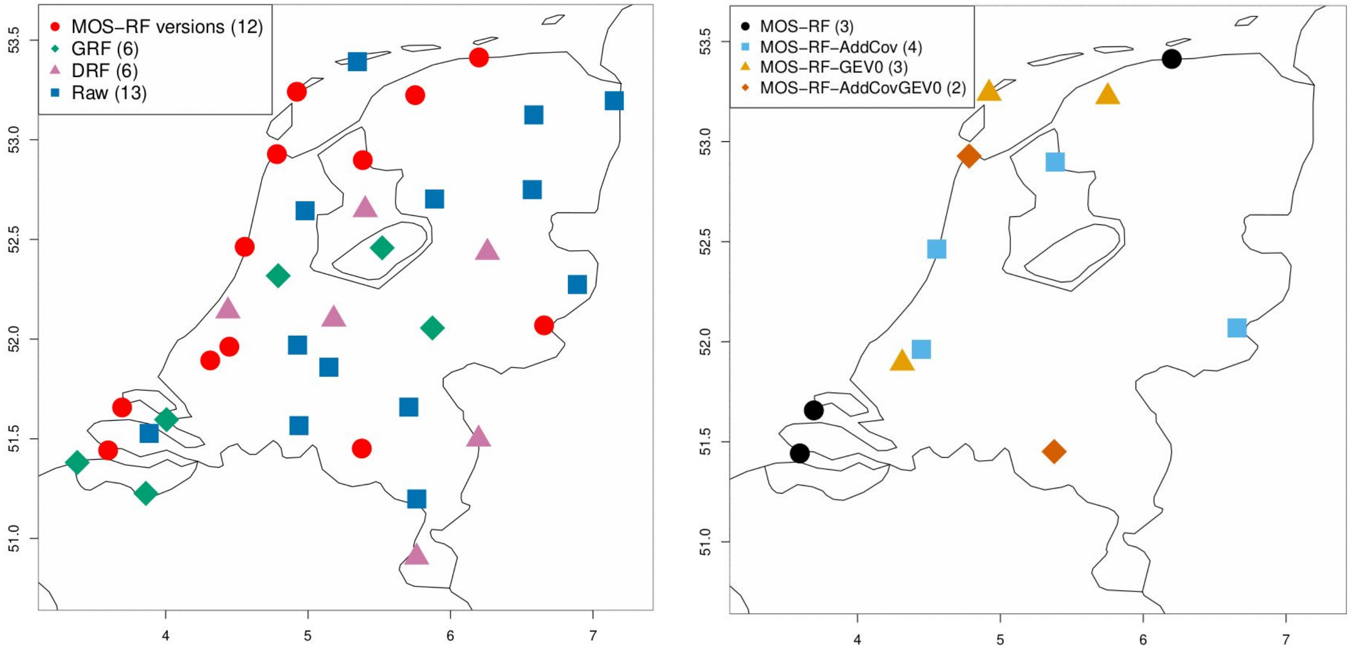


Figure: Post-processing method that performs best at each station.

- **MOS-RF versions** better along the coast.
- **Raw forecasts still better** for 1/3 of stations.
- **Regime-switching** improves extremes (wrt MOS-RF) for **9 out of 12 stations.**

Results for global forecast performances

Comparison of **MOS-RF versions** with:

- **GRF** (Taillardat et al., 2019) - **DRF** (Schlosser et al., 2019) - **Raw forecasts**

Metrics: **CRPS** at stations

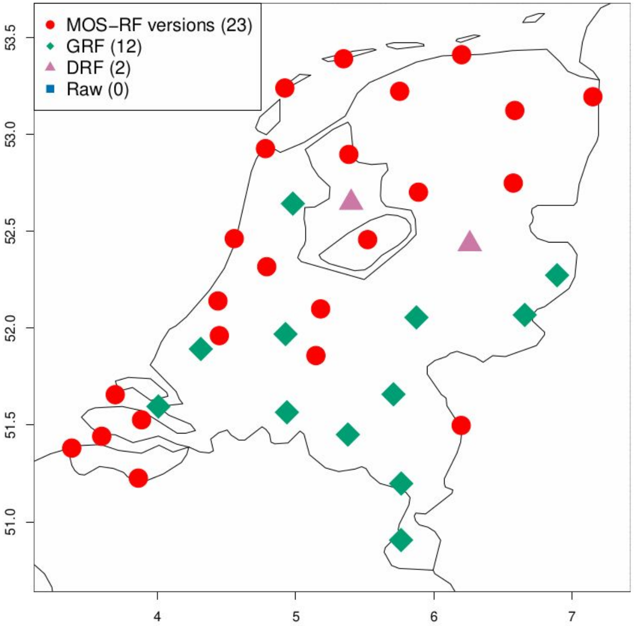


Figure: Post-processing method that performs best at each station.

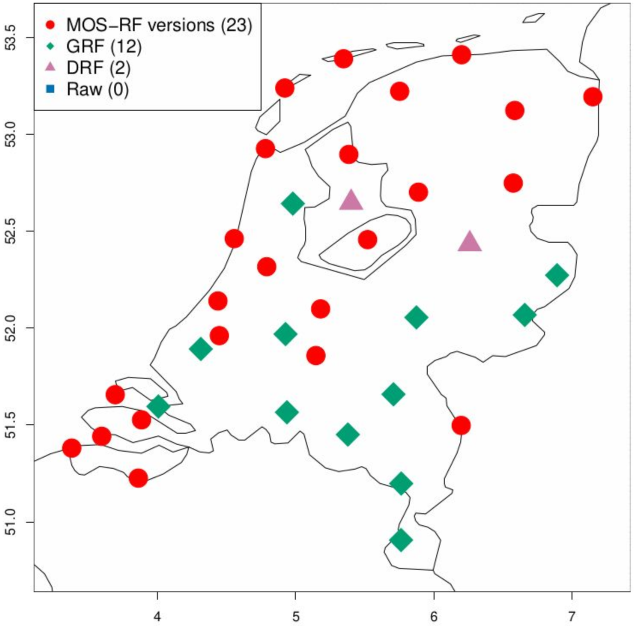
- **Forest-based methods** improve **Raw forecasts**
- **MOS-RF versions** generally better.

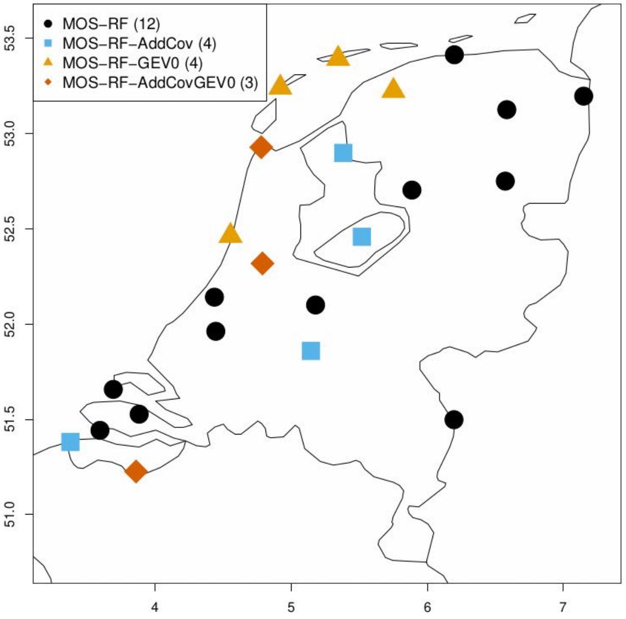
Results for global forecast performances

Comparison of **MOS-RF versions** with:

- **GRF** (Taillardat et al., 2019) - **DRF** (Schlosser et al., 2019) - **Raw forecasts**

Metrics: **CRPS** at stations





- **Forest-based methods** improve **Raw forecasts**
- **MOS-RF versions** generally better.
- **Regime-switching** not always improves CRPS wrt MOS-RF (11 out of 23 stations)

Figure: Post-processing method that performs best at each station.

Conclusions and Perspectives

- **MOS-RF + “Regime-switching”** approach to post-process wind gust extremes.
- **Mixed results for extremes** depending on:
  - the stations
  - metrics/intensity of extremes (not shown).
- Sometimes **difficult to beat Raw forecasts** for extremes.
- “Improving extremes comes at a cost” (Jakob Wessel's poster)

  ⇒ but methods based on MOS-RF present the best results

Conclusions and Perspectives

- **MOS-RF + “Regime-switching”** approach to post-process wind gust extremes.
- **Mixed results for extremes** depending on:
  - the stations
  - metrics/intensity of extremes (not shown).
- Sometimes **difficult to beat Raw forecasts** for extremes.
- “Improving extremes comes at a cost” (Jakob Wessel's poster)

  ⇒ but methods based on MOS-RF present the best results

- Further research is needed:
  - How can we beat **Raw** for (very high) extremes?
    - Fitting residuals instead of original target variable ⇒ better results?
    - Considering other Machine Learning tools (e.g., neural networks)
  - How can we better assess improvements to extremes?
  - Differences of extremes between the training/test set: Stratified sampling?

Additional slides

Preliminary results for wind gust extremes

Comparison of **MOS-RF versions** with:

- **GRF** (Taillardat et al., 2019)
- **DRF** (Schlosser et al., 2019)
- **Raw forecasts**

Metrics: **CRPS** at stations

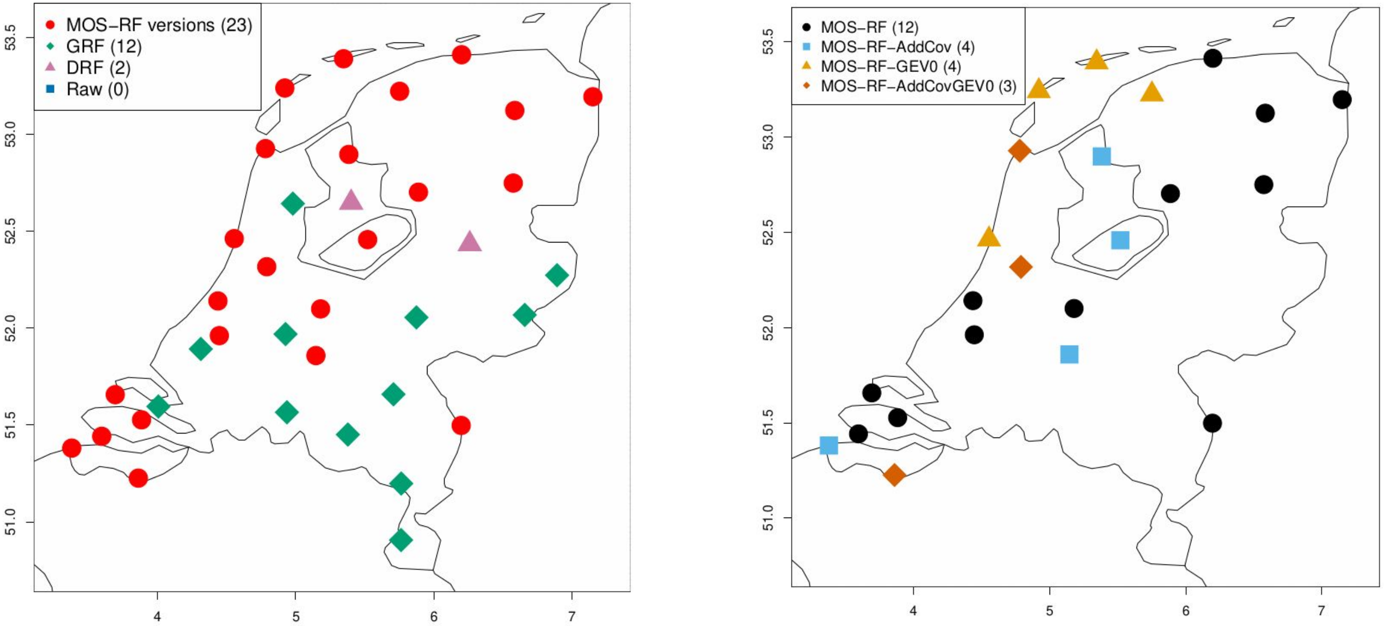


Figure: Post-processing method that performs best at each station.

Additional slides

<u>Comparison with MOS-RF:</u> **twCRPS** at stations for events above 21m/s (~97.5th obs. percentiles)

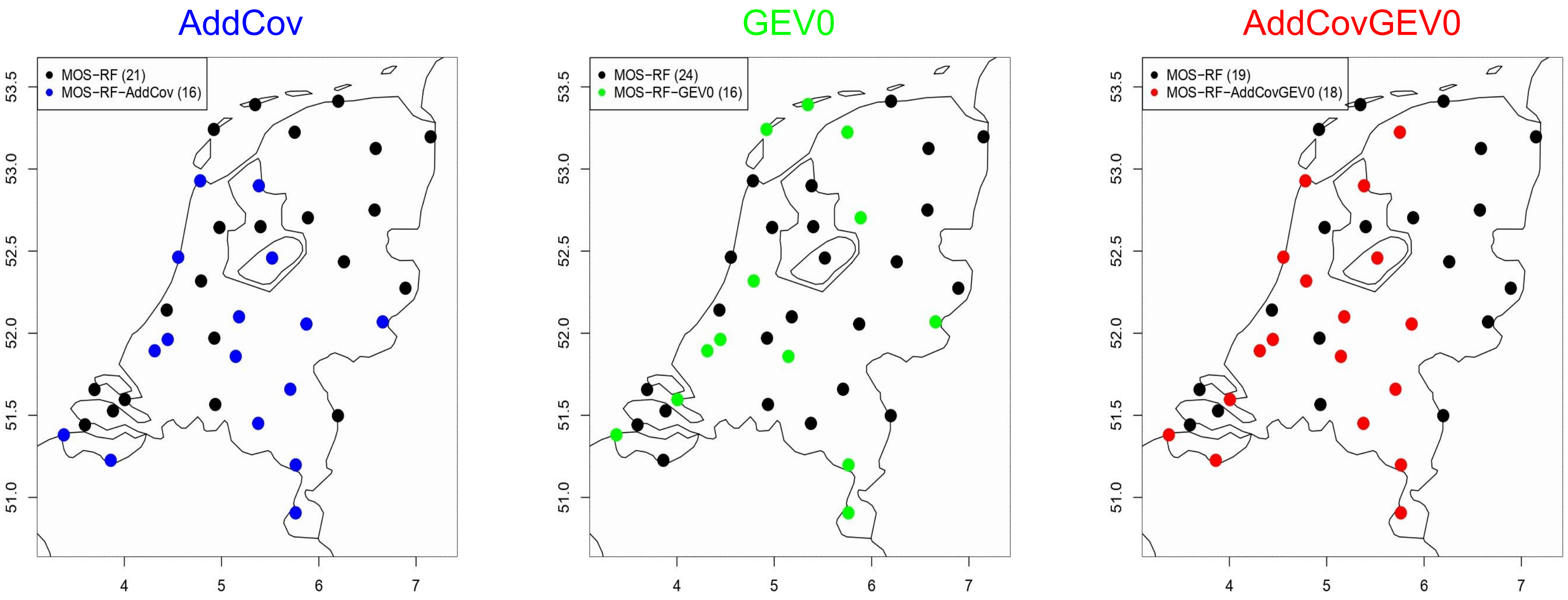


Figure: Post-processing method that performs best at each station.

Regime-switching: finding the optimal threshold

Observations for modified cases vs. threshold

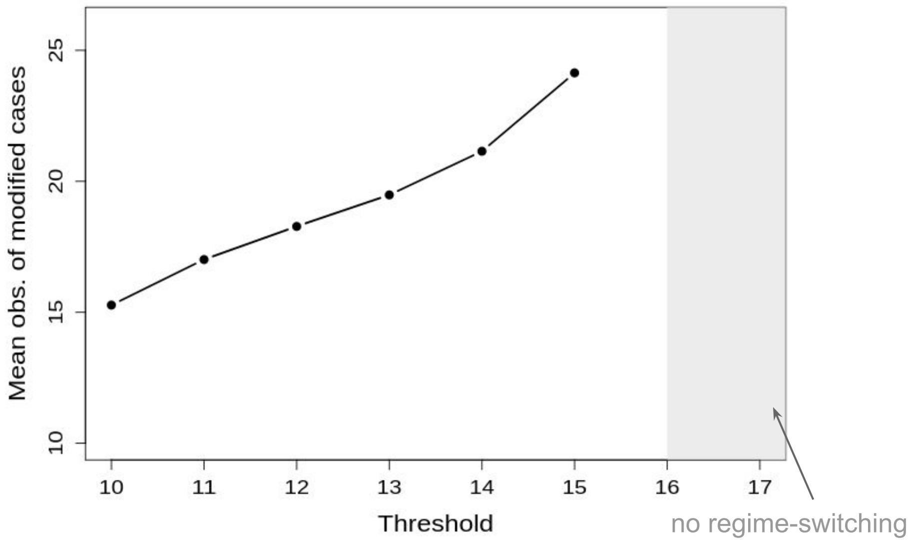


Figure: Mean of obs. for modified cases wrt threshold.

- Higher threshold ⇒ forecast associated with higher obs.

Select best threshold on validation set:

- Mean CRPSS wrt MOS-RF on modified cases.

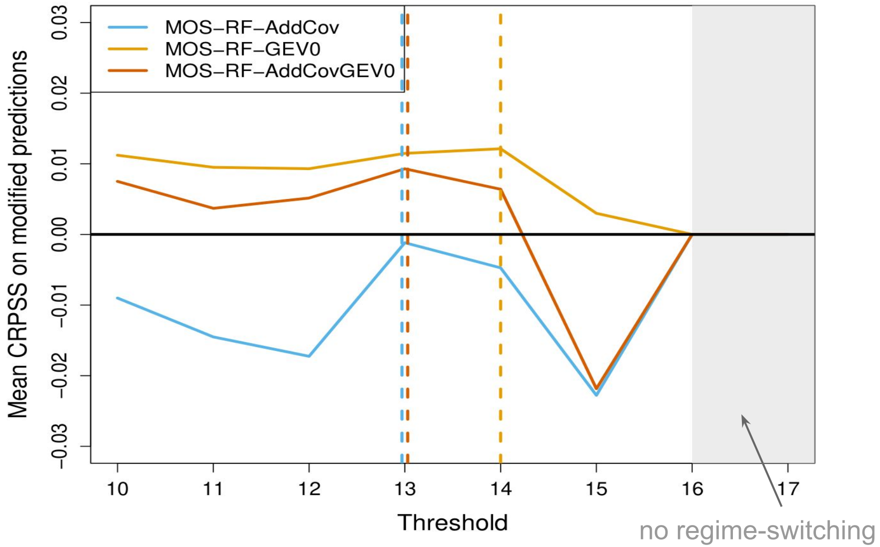


Figure: Mean CRPSS with respect to threshold.

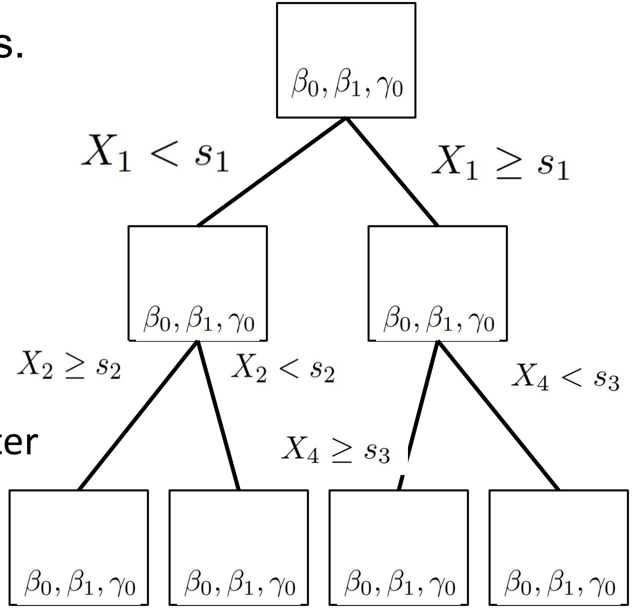
MOS-RF: how does it work?

<u>The MOS-RF method:</u> Adapt MOS model to weather situations.

- Assume a MOS model

$$y \sim \mathcal{N}_0(\mu,\, \sigma^2)$$
$$\text{with } \mu = \beta_0 + \beta_1 \times x_1$$
$$\text{and } log(\sigma) = \gamma_0 \times x_1$$

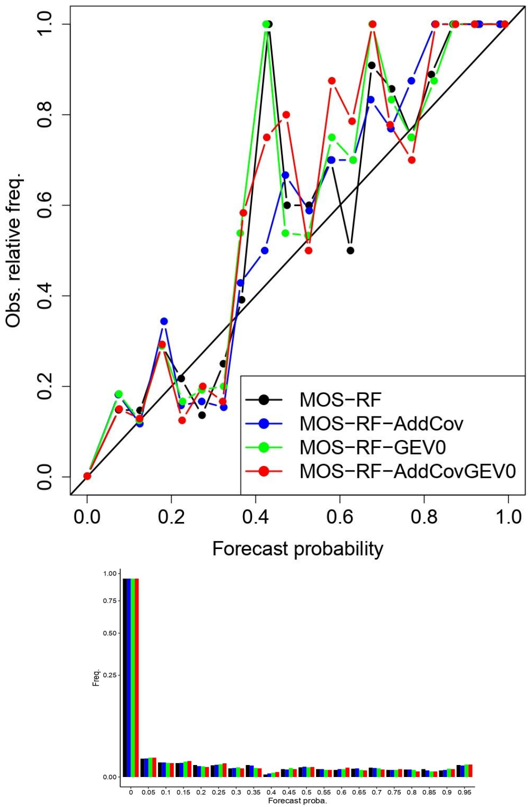
- Training steps:
  - Estimate MOS coefficients at a parent node.
  - Select optimal covariate and split point to fit better MOS models
  - Repeat.

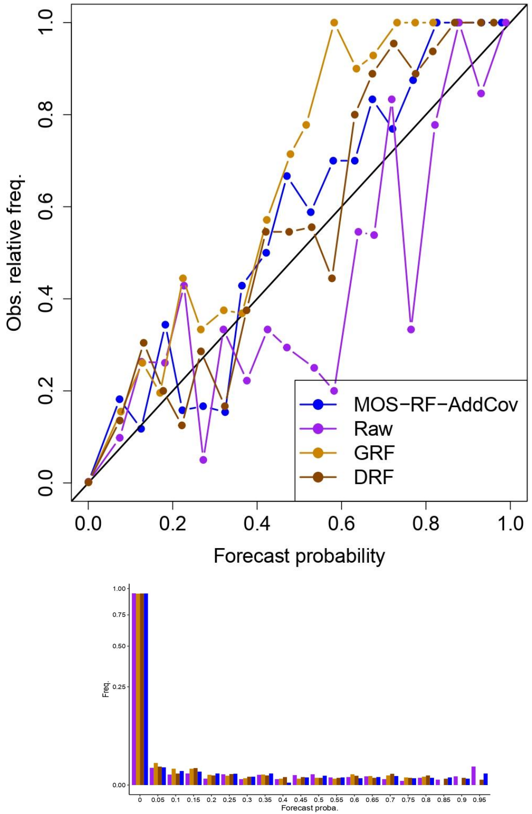


… but assume one MOS model: **suitable for wind gust extremes**?

Results for the post-processing of wind gust extremes

<u>Reliability and sharpness diagrams for events above 21m/s</u>





- **MOS-RF-AddCov** seems more reliable
- Sharpness better for **Raw**.

Results for the post-processing of wind gust extremes

Comparison of MOS-RF versions with:

- **GRF** (Taillardat et al., 2019) - **DRF** (Schlosser et al., 2019) - **Raw forecasts**

Metrics: **twCRPS** at stations for events above **21m/s** (~97.5th obs. percentiles)

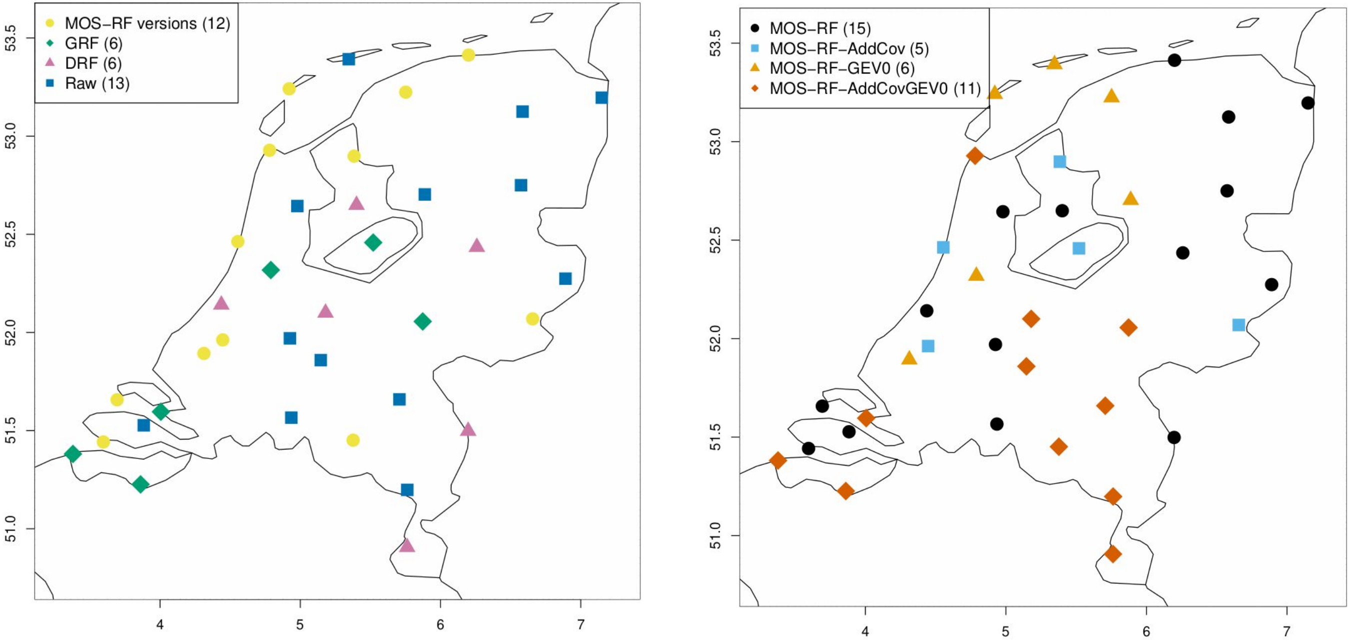


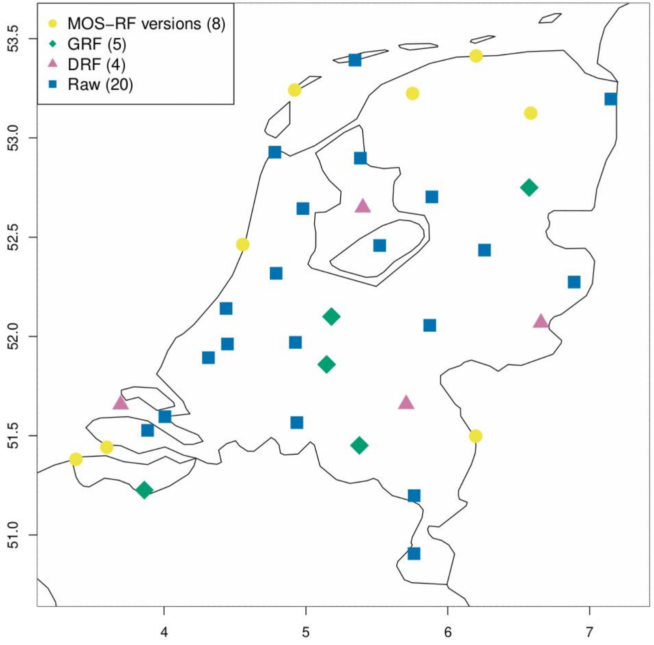
Figure: Post-processing method that performs best at each station.

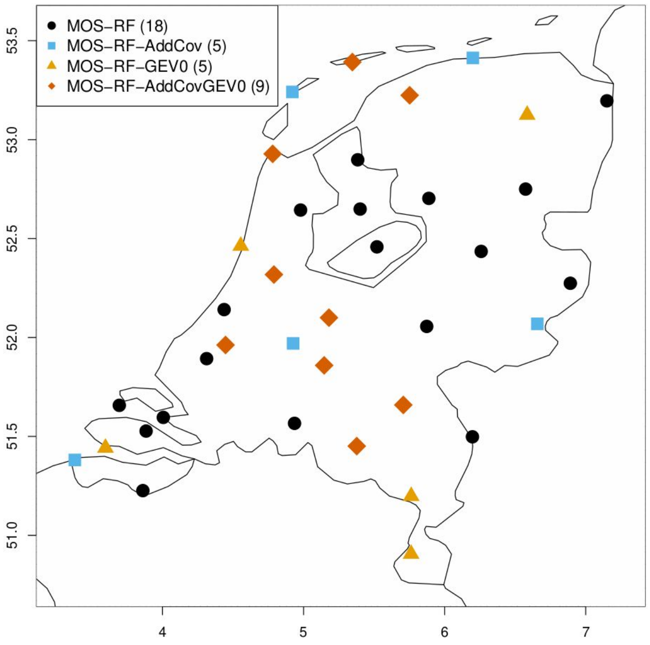
Results for the post-processing of wind gust extremes

Comparison of MOS-RF versions with:

- **GRF** (Taillardat et al., 2019) - **DRF** (Schlosser et al., 2019) - **Raw forecasts**

Metrics: **Brier Score** at stations for events above **21m/s** (~97.5th obs. percentiles)





- **Raw forecasts better** for Brier Scores

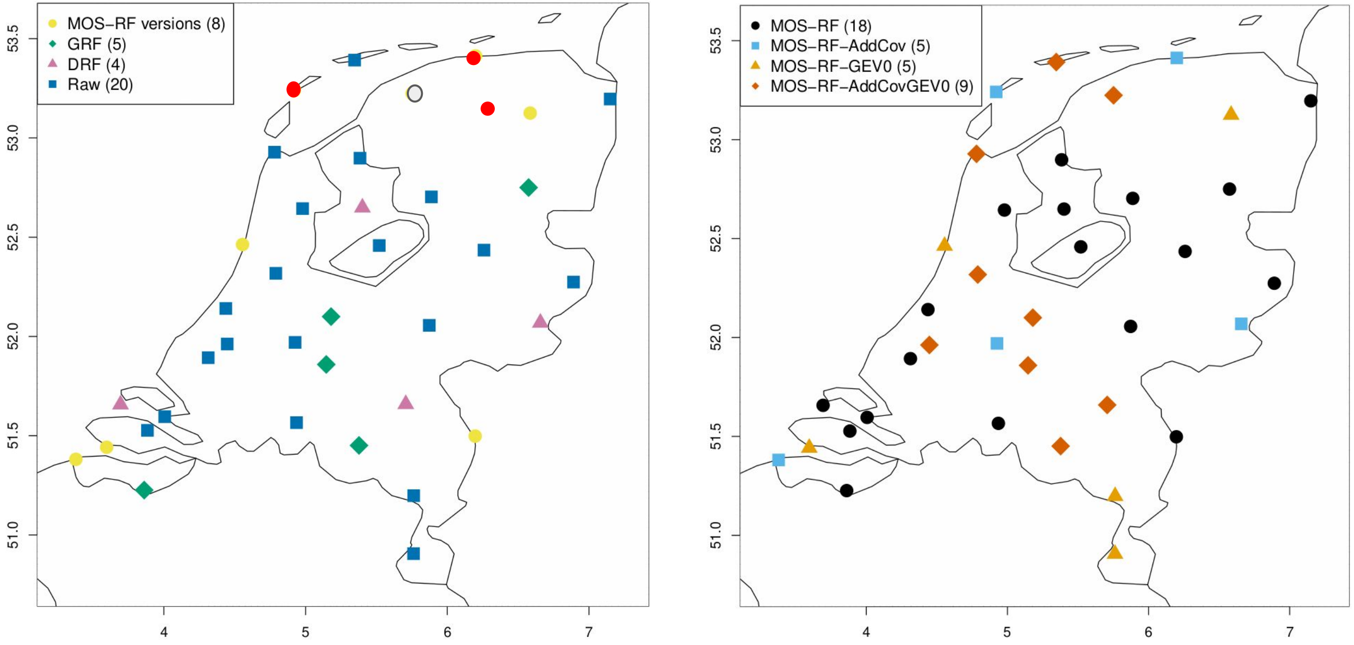
Figure: Post-processing method that performs best at each station.

Preliminary results for wind gust extremes

Comparison of **MOS-RF versions** with:

- **GRF** (Taillardat et al., 2019) - **DRF** (Schlosser et al., 2019) - **Raw forecasts**

Metrics: **Brier Score** at stations for events above 21m/s (~97.5th obs. percentiles)



- **Raw forecasts** **better** for Brier Scores

Figure: Post-processing method that performs best at each station.

Results for the post-processing of wind gust extremes

Comparison of MOS-RF versions with:

- **GRF** (Taillardat et al., 2019)
- **DRF** (Schlosser et al., 2019)
- **Raw forecasts**

Metrics: **twCRPS** at stations for events above **23m/s** (~99th obs. percentiles)

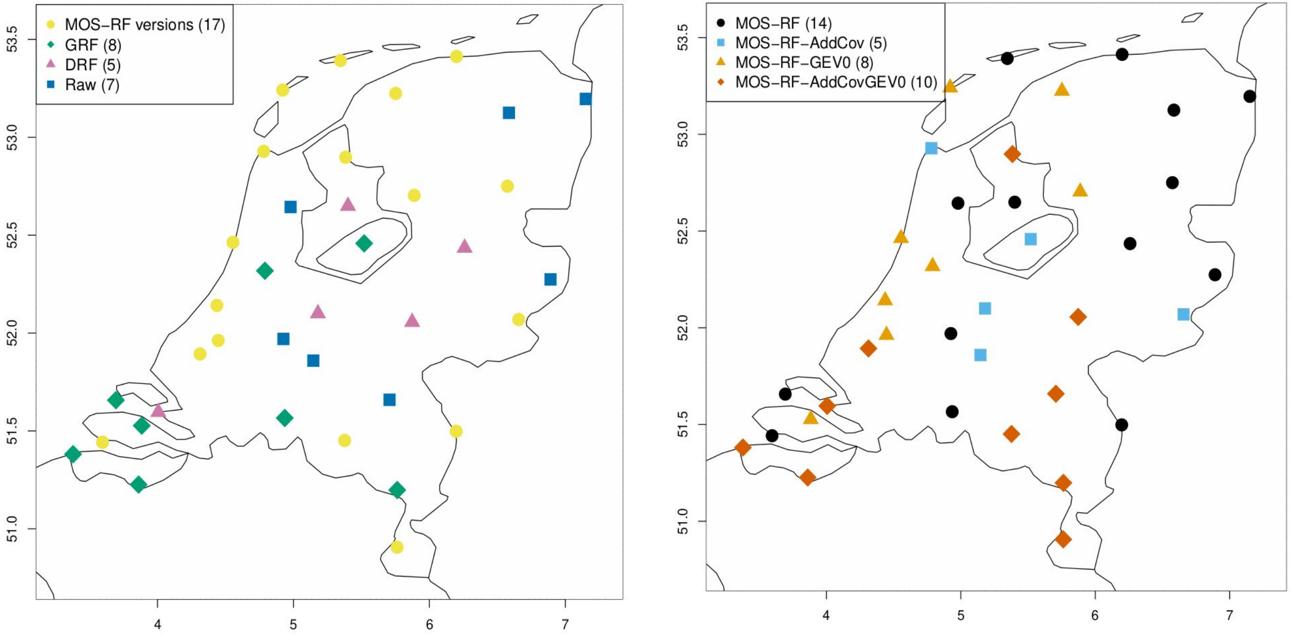


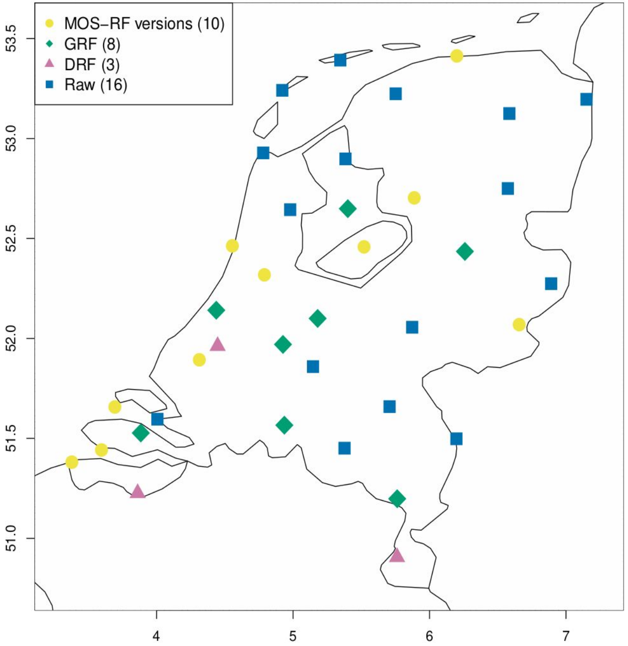
Figure: Post-processing method that performs best at each station.

Results for the post-processing of wind gust extremes

Comparison of MOS-RF versions with:

- **GRF** (Taillardat et al., 2019) - **DRF** (Schlosser et al., 2019) - **Raw forecasts**

Metrics: **Brier Score** at stations for events above **23m/s** (~99th obs. percentiles)



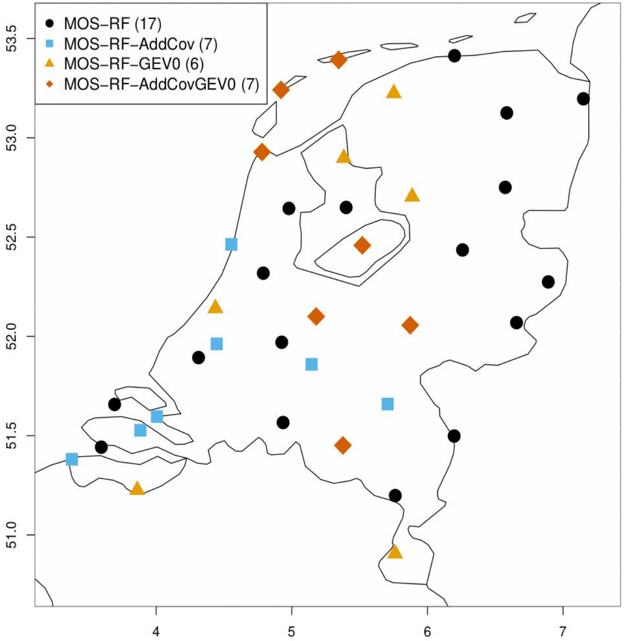


Figure: Post-processing method that performs best at each station.

- **Raw forecasts better** for Brier Scores.
- Mixed results with higher thresholds