Comparison of the performance of PCA-NN and PCA-MRM models for TEC over the Iberian Peninsula

Anna Morozova[1], Teresa Barata[1], and Tatiana Barlyaeva[1]

[1]University of Coimbra

November 23, 2022

Abstract

The total electron content (TEC) over the Iberian Peninsula was modeled using a PCA-based models based on the decomposition of the observed TEC series using the principal component analysis (PCA) and reconstruction of the daily modes' amplitudes either by a multiple linear regression model (MRM) or neural networks (NN) using several types of space weather parameters as regressors/predictors: proxies for the solar UV and XR fluxes, number of the solar flares of different types, parameters of the solar wind and of the interplanetary magnetic field, and geomagnetic indices. Lags of 1 and 2 days between the TEC and space weather parameters are used. The general performance of the PCA-MRM and PCA-NN models is tested for different months and in different space weather conditions.

NG45B-0551

AGU FALL MEETING
New Orleans, LA & Online Everywhere
13-17 December 2021

COMPARISON OF THE PERFORMANCE OF PCA-NN AND PCA-MRM MODELS FOR TEC OVER THE IBERIAN PENINSULA

ANNA MOROZOVA, TERESA BARATA, TATIANA BARLYAEVA

University of Coimbra, Instituto de Astrofísica e Ciências do Espaço, OGAUC, Coimbra, Portugal

1 2 90 FACULDADE DE CIÊNCIAS E TECNOLOGIA UNIVERSIDADE D COIMBRA

Observatório Geofísico e Astronómico da Universidade de Coimbra

ia instituto de astrofísica e ciências do espaço

CONTENT

Content

ABSTRACT

- The total electron content (TEC) over the Iberian Peninsula was modeled using a PCA-based models based on the decomposition of the observed TEC series using the principal component analysis (PCA) and reconstruction of the daily modes' amplitudes either by a multiple linear regression model (MRM) or neural networks (NN).
- Several types of space weather parameters are used as regressors/predictors: proxies for the solar UV and XR fluxes, number of the solar flares of different types, parameters of the solar wind and of the interplanetary magnetic field, and geomagnetic indices.
- Lags of 1 and 2 days between the TEC and space weather parameters are used.
- The general performance of the PCA-MRM and PCA-NN models is tested for different months and in different space weather conditions.

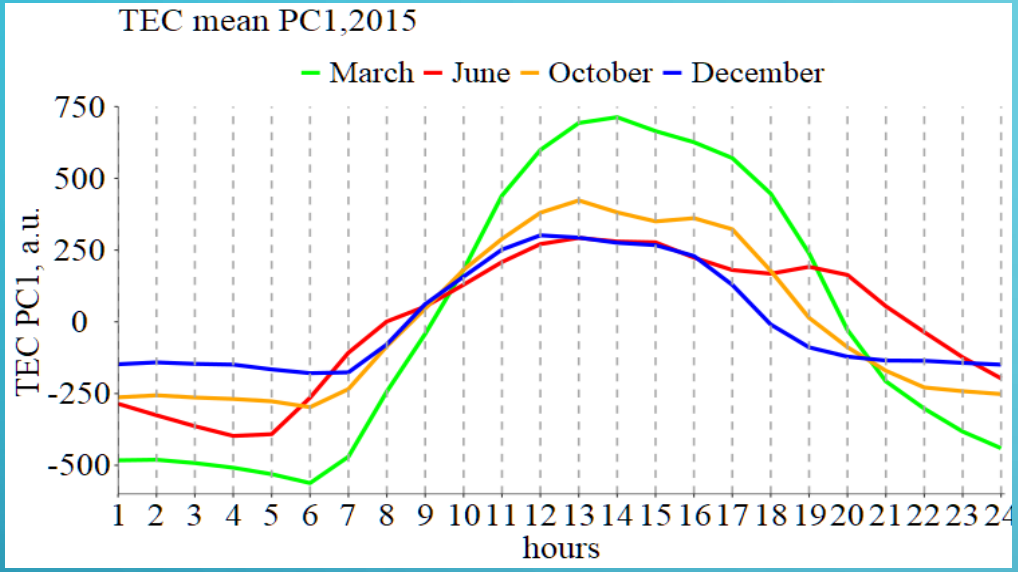
Content

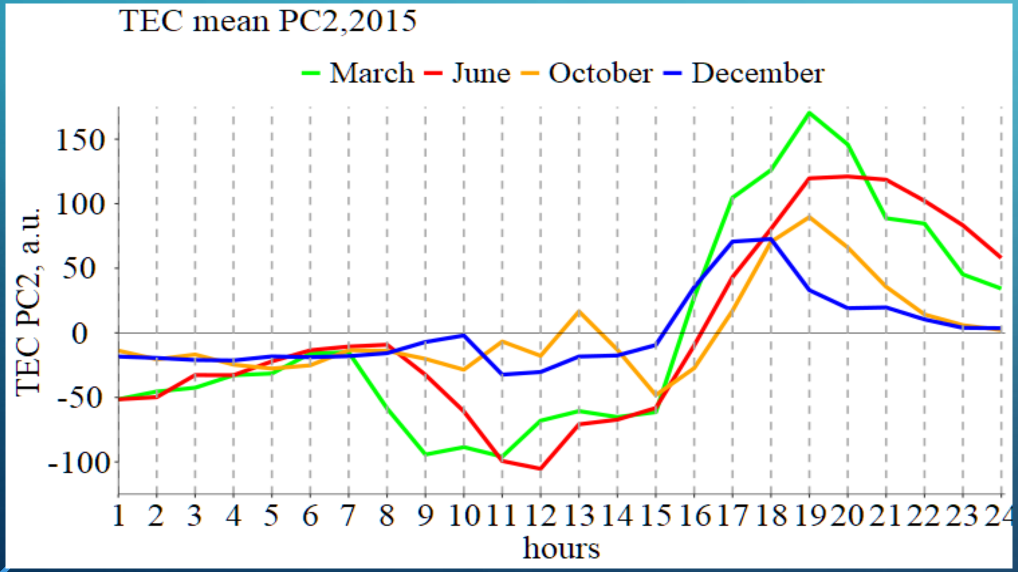
PCA-BASED MODELS

- The main feature of the PCA-based models is that the TEC series is decomposed into several PCA modes which represent TEC daily variations of different types
- The amplitude of each of the mode for each day is described by the EOF coefficients
- The EOF coefficients can be modelled using space weather parameters as predictors using, e.g., multiple regression models (MRM) or neural networks (NN)
- The MRM regression coefficients or a trained NN can be used to forecast the EOF coefficients, and, consequently, to forecast TEC
- The advantage of the PCA-based models is that there is no need for any assumption on the phase and amplitude or seasonal/regional features of TEC daily variations: the daily variations of correct shapes are extracted automatically by PCA from the input TEC data.

Content

PCA MODES





Mode 1:

Explains 77-95% of the TEC variations for different months

PC1 = regular daily variation due to the changes of the insolation

Mode 2:

Explains 1.5-8.4% of the TEC variations for different months

PC2 = shallow minimum of TEC around the noon and a maximum in the late afternoon

Back

Content

DATA: TEC

- Vertical TEC measured at Lisbon airport, Portugal (39° N, 9° W) by a GNSS receiver with SCINDA system
- Time interval: 01.01.2015- 31.12.2015
- Time resolution: 1h data
- Calibrated using data from Royal Observatory of Belgium (ROB) GNSS Research Group for the grid point that is very close to the location of the Lisbon airport.
  - To convert the relative TEC SCINDA data to the TEC data in TECu
  - The calibration was performed individually for each of the 12 month

Content

DATA: SPACE WEATHER PARAMETERS (*PREDICTORS*)

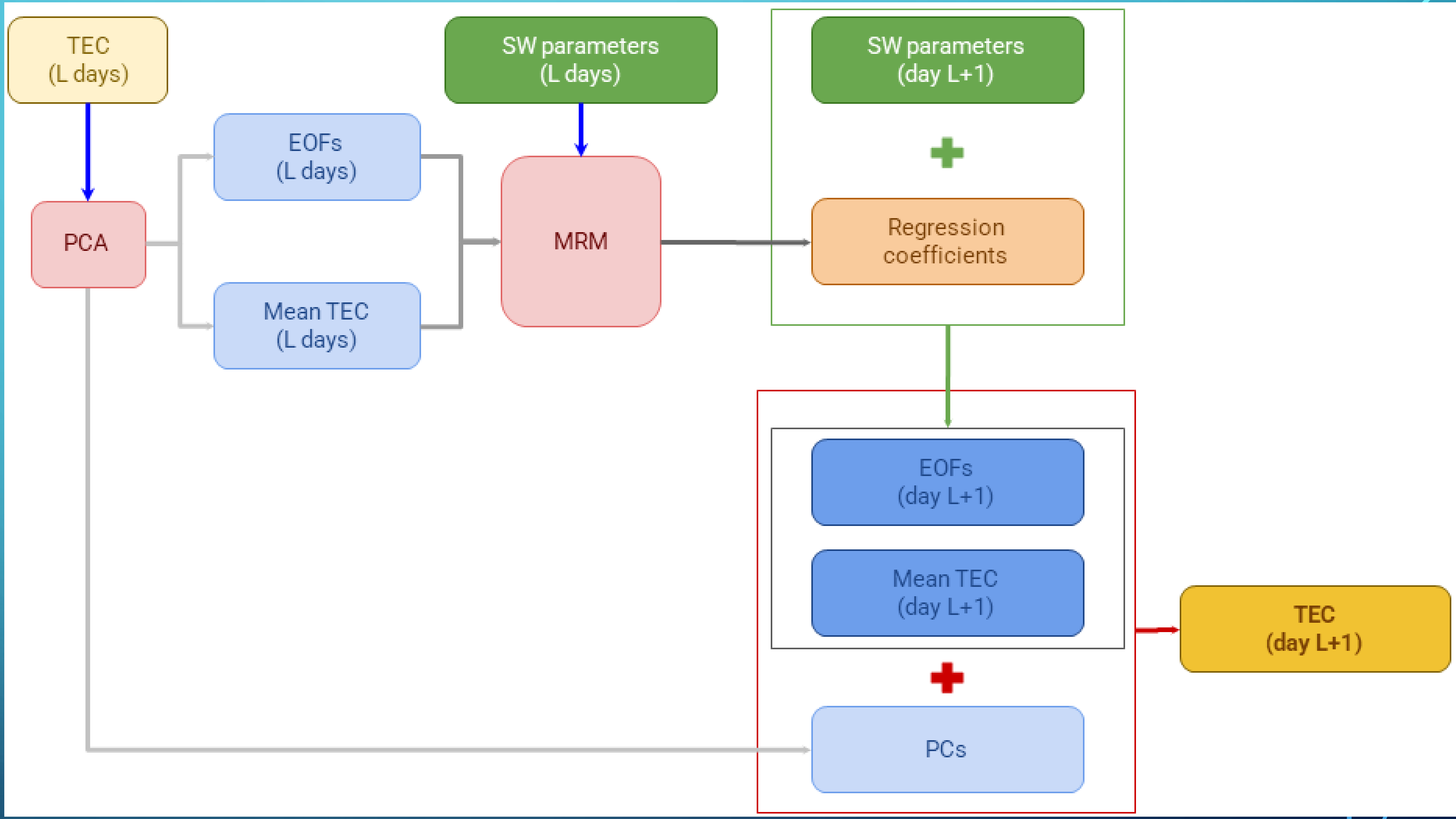
- Solar wind parameters:
  - Pressure (p), density (n), velocity (v)
- Interplanetary magnetic field:
  - Full interplanetary magnetic field (scalar B), GSM components (Bx, By, Bz)
- Geomagnetic indices:
  - Dst, ap, AE, local K_{COI}-index (Coimbra Geomagnetic Observatory, Portugal)
- Proxies for the solar UV & XR fluxes:
  - UV: Mg II composite series – a proxy for the spectral solar irradiance variability in the spectral range from UV to EUV (Snow et al., 2014);
  - F10.7 index (OMNI)
  - XR: Solar EUV Experiment (SEE) for the NASA TIMED mission at the wavelength 0.5 nm (LISIRD)
- Daily number of solar flares of classes C and M (NGDC)
- Time resolution: 1d data

Back

Content

PCA-MRM MODEL

- L = 31 or 32 days
- Lag between space weather parameters and TEC = 1 & 2 days (space weather parameters lead)
-

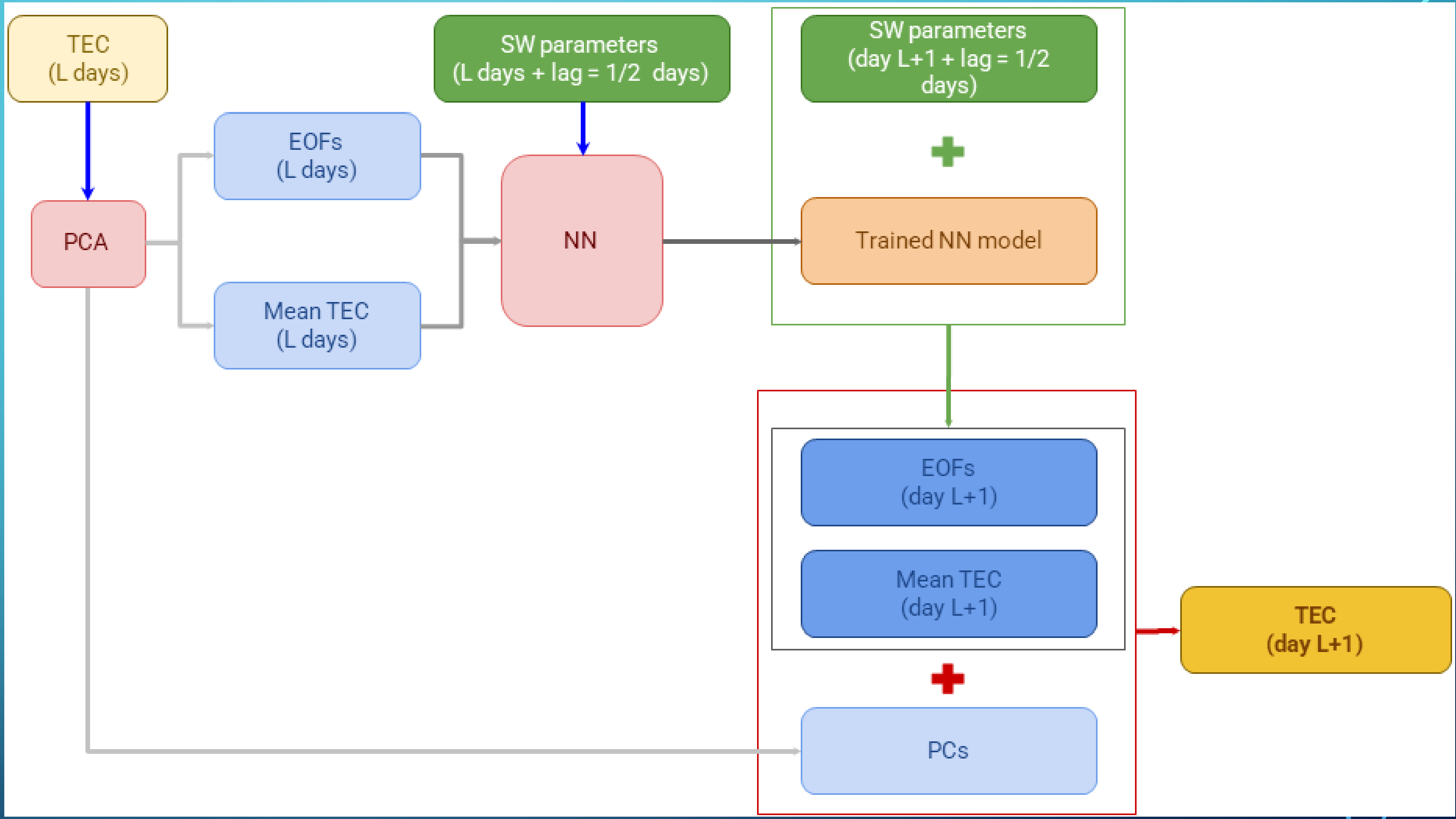


Back

Content

PCA-NN MODEL

- L = 31
- NN with "memory" (LSTM NN) or feedforward NN with weight backpropagation trained on the lagged series of predictors



Back

Content

NEURAL NETWORK

- R package: *neuralnet*
- NN algorithm: *the resilient backpropagation with weight backtracking*
- Input dataset length = 31 days → for NN models we cannot use all 16 predictors → we selected the predictors that were used by most of the PCA-MRM models
- Tested predictors: *Mg II, Dst, By lagged by 1 and 2 days (6 input series)*
  - These space weather parameters were used in 67-83% of the tested MRM models
- Tested TEC parameter: *daily mean TEC*

Content

COMPARISON OF THE MRM AND NN FORECASTS

- Different NN were trained to find the best combination of the
  - Input series
  - NN depth (number of hidden layers and number of nodes)
  - Usage of an "ensemble forecast": a number (e.g., 100) of NN models of the same architecture were trained on the same input dataset and were used to make a forecast for the day L+1; the final forecast is the arithmetic average of 100 forecasts
- The forecasts of the daily mean TEC made by NN models for all available days of 2015 were compared to the forecasts of the daily mean TEC made by the (PCA)-MRM model using the following metrics:
  1. Correlation coefficient (r)
  2. Mean absolute error (MAE)
  3. Root-mean-square error (RMSE)

Content

COMPARISON OF THE MRM AND NN FORECASTS

	(PCA) -MRM	NN 3 predictors			NN 3 predictors "ensemble"			NN 2 predictors			NN, 1 predictor		
predictors	16 predictors & "best subset"	MgII, Dst, By			MgII, Dst, By			MgII, Dst	Dst, By	MgII, By	MgII	Dst	By
Layers & nodes		(6,4,2)	(6,4)	(6,2)	(6,4,2)	(6,4)	(6,2)	(4,2)	(4,2)	(4,2)	(2)	(2)	(2)
r	**0.88**[1]	0.88	0.88	0.88	**0.9**[2,3]	0.89	**0.9**[2,3]	**0.9**[2,3]	0.87	0.89	0.89	0.87	0.85
MAE	**2.0**[1,2]	2.21	2.16	2.17	**2.08**[3]	2.1	**2.08**[3]	**2.05**[3]	2.28	2.13	2.11	2.31	2.91
RMSE	**2.8**[1]	2.79	2.82	2.79	**2.64**[2,3]	2.7	**2.63**[2,3]	**2.61**[2,3]	2.88	2.7	2.68	2.91	3.09

[1,2,3] Notes

Back

COMPARISON OF THE MRM AND NN FORECASTS

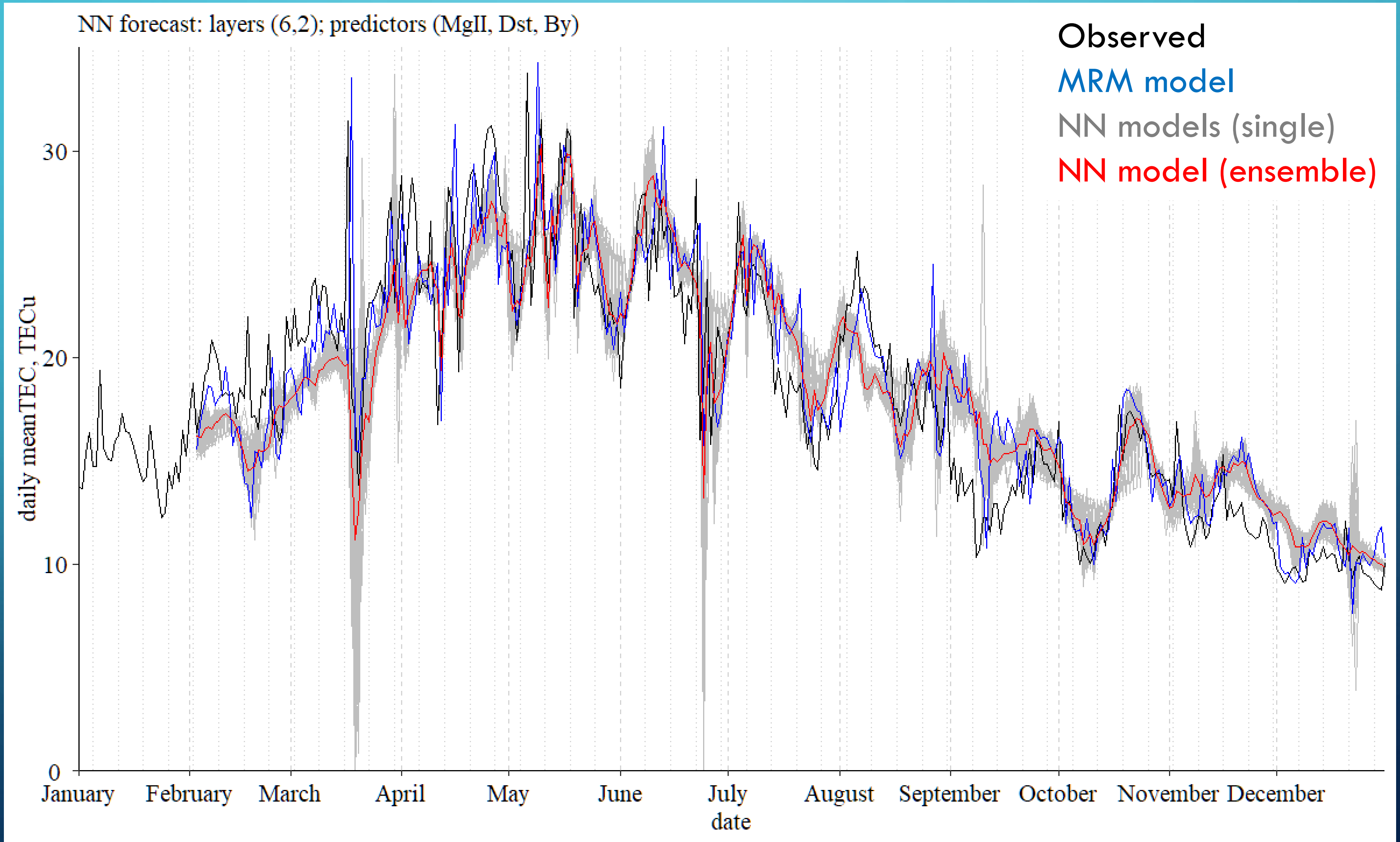
1 – metrics for the (PCA)-MRM model
2 – best metrics for all kind of models
3 – best metrics for NN models

Back

Content

COMPARISON OF THE MRM AND NN FORECASTS

- Observations vs MRM model vs NN models with single and ensemble forecasts

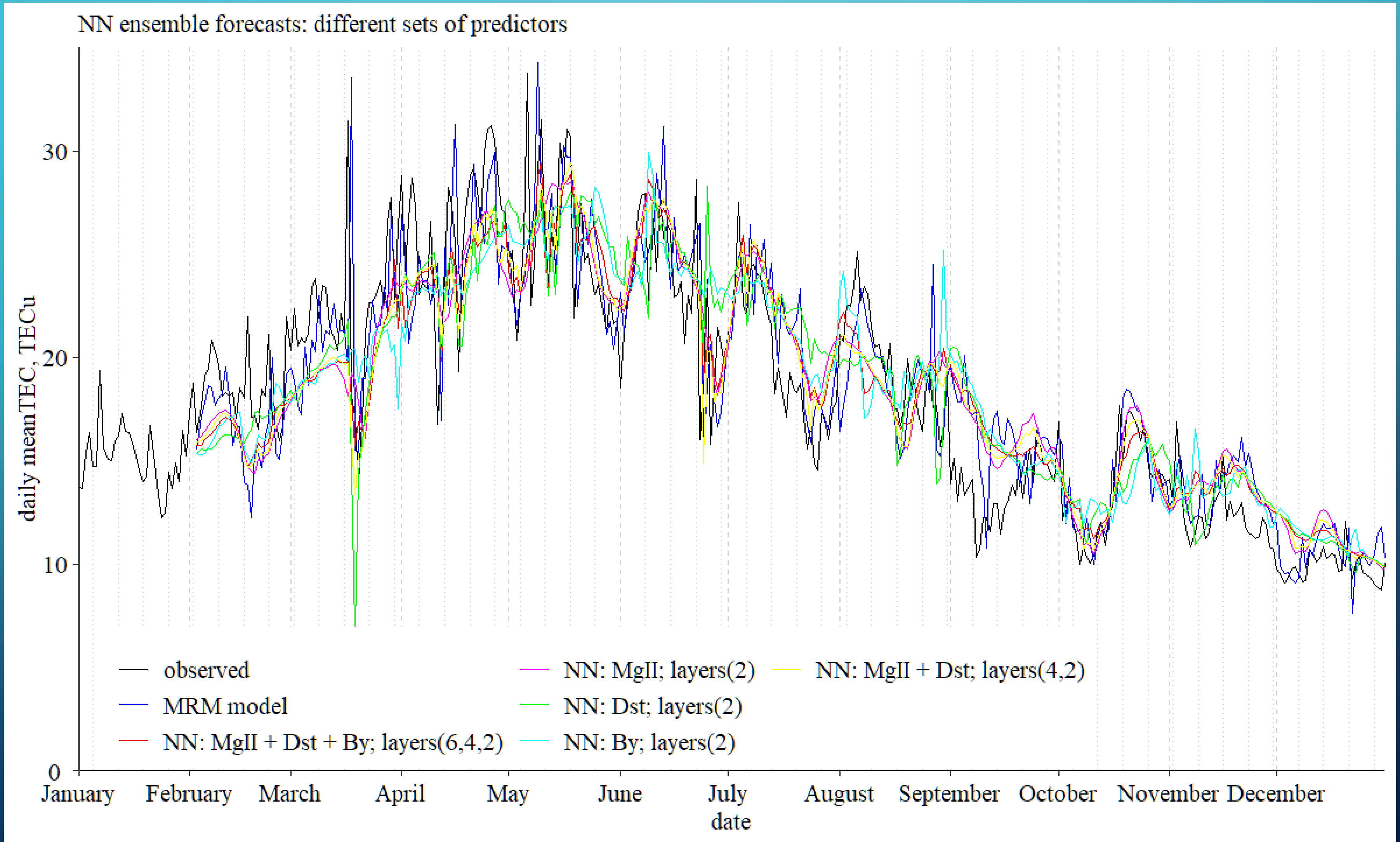


Back

Content

COMPARISON OF THE MRM AND NN FORECASTS

- Observations vs MRM model vs NN models with 1, 2 and 3 predictors

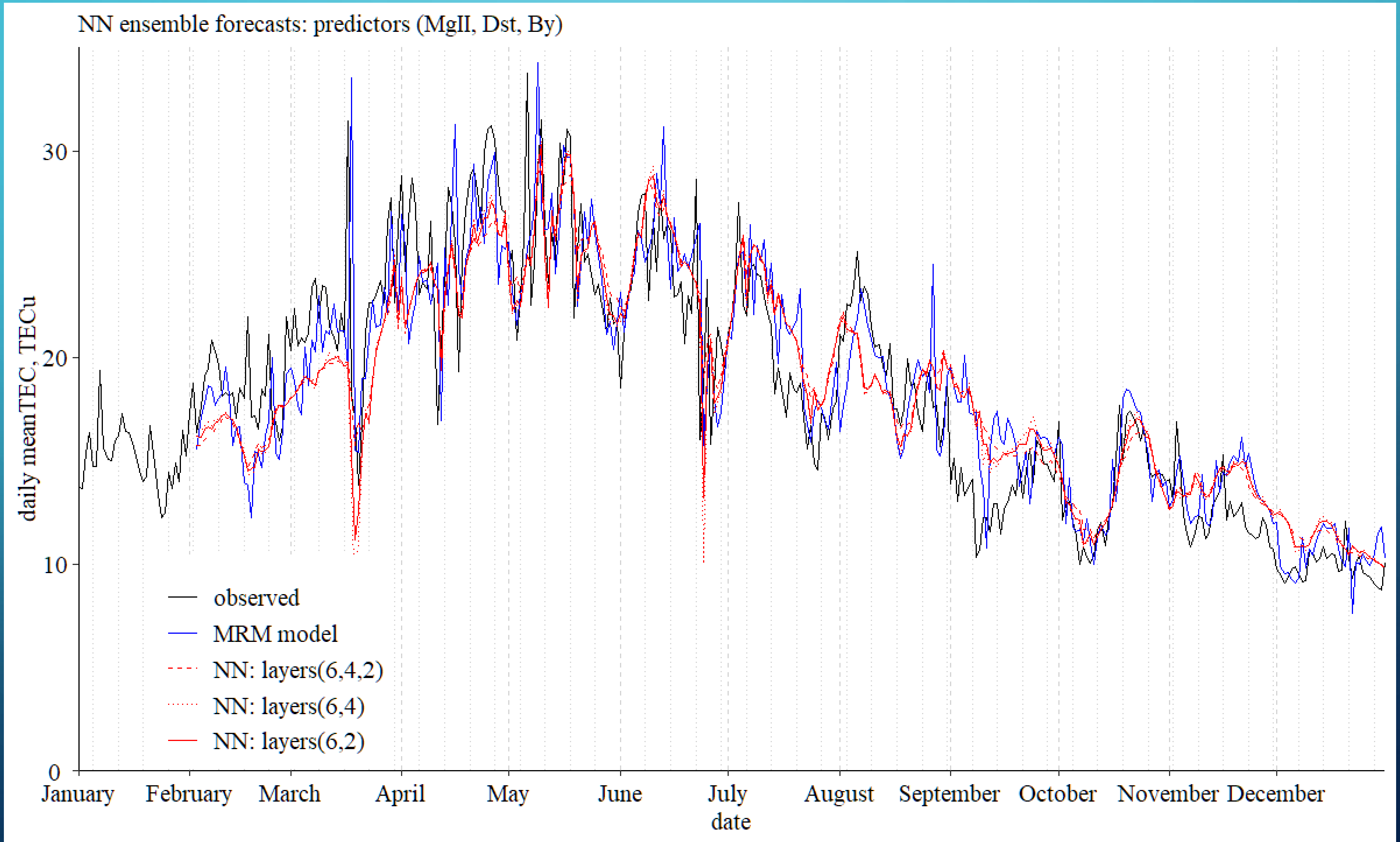


Back

Content

COMPARISON OF THE MRM AND NN FORECASTS

- Observations vs MRM model vs NN models with 3 predictors and different NN depth

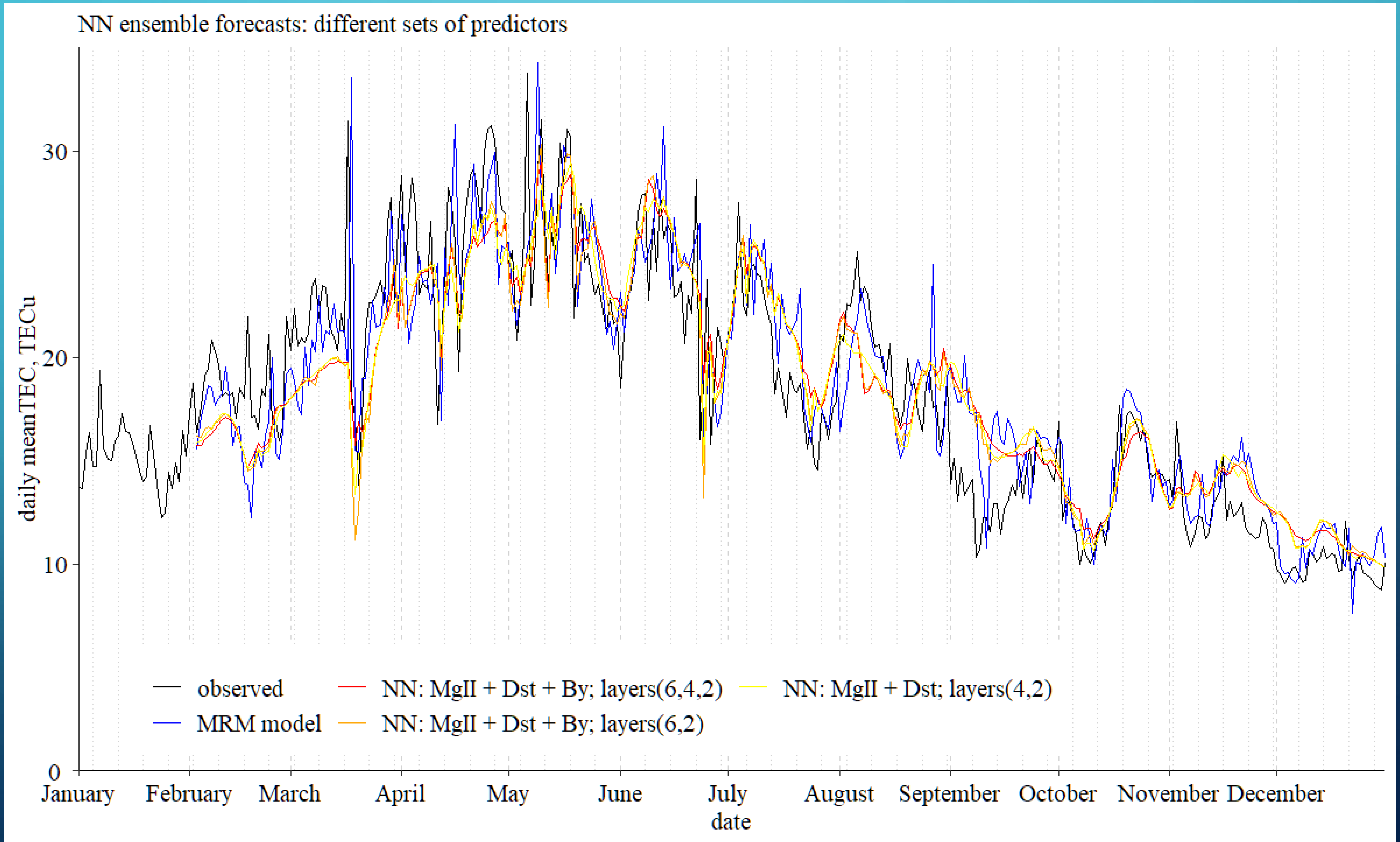


Back

Content

COMPARISON OF THE MRM AND NN FORECASTS

- Observations vs MRM model vs best NN models (marked in the Table)



Back

Content

CONCLUSIONS

- A simple NN (feedforward NN with weight backpropagation) with just 2 or 3 predictors with time lags trained on a 31-days long input dataset can forecast the daily mean TEC series with the same or even better quality than the multiple regression model with up to 16 regressors
  - For some time intervals both the MRM and NN models give similar predictions different from observations. Hypothesis: TEC variations for that time intervals have other (non-space weather) drivers
- Ensemble NN forecast perform better than the single forecast
- As predictors the solar UV proxy (MgII) is the most important predictor (models without MgII perform worse)
- The Dst index added to MgII improve the performance of NN
- Adding the By parameter slightly improve the forecast quality (the NN forecast with 3 space weather parameters is closer to the observations)

Content

NEXT STEPS

- To check other space weather parameter as predictors
- To find the optimal length and optimal list of predictors
- To test other NN architectures

Content

AUTHORS' INFO

ANNA MOROZOVA

annamorozovauc@gmail.com

ORCID iD: 0000-0002-8552-8052

Researcher at IA-U.Coimbra, Portugal

She is the author of 22 papers in peer review journals. The main areas of scientific activities are *Solar-Terrestrial Physics; Solar physics; Atmosphere physics; Ionosphere physics; Statistical data analysis; Geophysical data homogenization.*

TERESA BARATA

mtbarata@gmail.com

ORCID iD:0000-0001-6106-8285

Researcher at IA-U.Coimbra, Portugal

She is the author of 30 papers in peer review journals. The main areas of scientific activities are *Image Processing and Mathematical Morphology; Solar images; Space Weather.*

TATIANA BARLYAEVA

tvbarlyaeva@gmail.com

ORCID iD: 0000-0001-6562-594X

Invited Researcher at IA-U.Coimbra, Portugal

The author or co-author of 20 papers in peer review journals. The main areas of scientific activities are *Space Weather; Solar and Solar-Terrestrial Physics; GNSS; Remote Sensing.*

Content

ACKNOWLEDGEMENT

- This research was supported through the project "SWAIR - Space weather impact on GNSS service for Air Navigation", ESA Small ARTES Apps
- This work was supported by Fundação para a Ciência e a Tecnologia (FCT) through the research grants UIDB/04434/2020 and UIDP/04434/2020.

ARTES IAP Demonstration Projects

Cofinanciado por:

NG45B-0551

AGU FALL MEETING
New Orleans, LA & Online Everywhere
13-17 December 2021

COMPARISON OF THE PERFORMANCE OF PCA-NN AND PCA-MRM MODELS FOR TEC OVER THE IBERIAN PENINSULA

ANNA MOROZOVA, TERESA BARATA, TATIANA BARLYAEVA

University of Coimbra, Instituto de Astrofísica e Ciências do Espaço, OGAUC, Coimbra, Portugal

1 2 9 0 FACULDADE DE CIÊNCIAS E TECNOLOGIA UNIVERSIDADE D COIMBRA

Observatório Geofísico e Astronómico da Universidade de Coimbra

ia instituto de astrofísica e ciências do espaço

CONTENT

Content

ABSTRACT

- The total electron content (TEC) over the Iberian Peninsula was modeled using a PCA-based models based on the decomposition of the observed TEC series using the principal component analysis (PCA) and reconstruction of the daily modes' amplitudes either by a multiple linear regression model (MRM) or neural networks (NN).
- Several types of space weather parameters are used as regressors/predictors: proxies for the solar UV and XR fluxes, number of the solar flares of different types, parameters of the solar wind and of the interplanetary magnetic field, and geomagnetic indices.
- Lags of 1 and 2 days between the TEC and space weather parameters are used.
- The general performance of the PCA-MRM and PCA-NN models is tested for different months and in different space weather conditions.

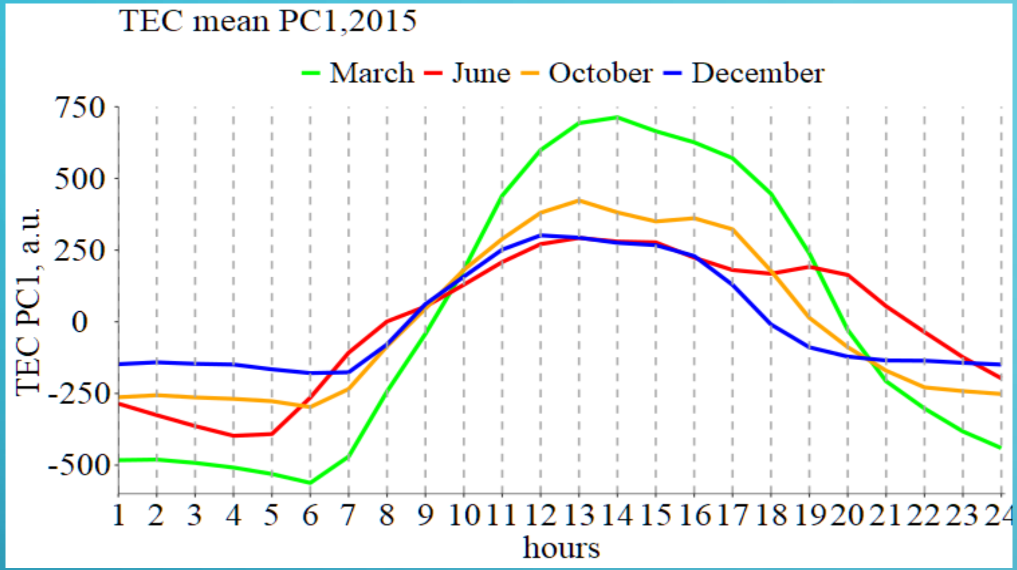
Content

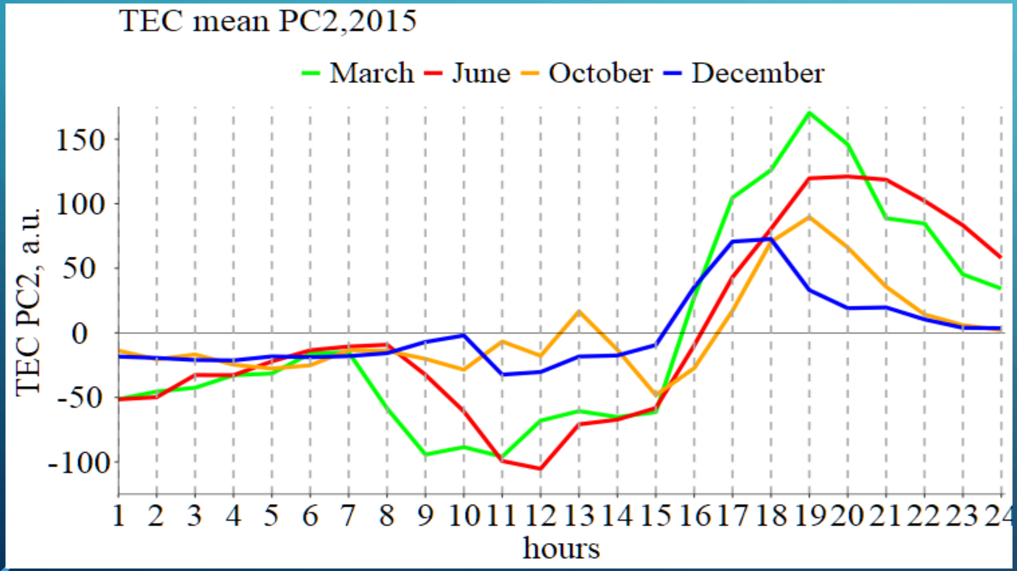
PCA-BASED MODELS

- The main feature of the PCA-based models is that the TEC series is decomposed into several PCA modes which represent TEC daily variations of different types
- The amplitude of each of the mode for each day is described by the EOF coefficients
- The EOF coefficients can be modelled using space weather parameters as predictors using, e.g., multiple regression models (MRM) or neural networks (NN)
- The MRM regression coefficients or a trained NN can be used to forecast the EOF coefficients, and, consequently, to forecast TEC
- The advantage of the PCA-based models is that there is no need for any assumption on the phase and amplitude or seasonal/regional features of TEC daily variations: the daily variations of correct shapes are extracted automatically by PCA from the input TEC data.

Content

PCA MODES





Mode 1:

Explains 77-95% of the TEC variations for different months

PC1 = regular daily variation due to the changes of the insolation

Mode 2:

Explains 1.5-8.4% of the TEC variations for different months

PC2 = shallow minimum of TEC around the noon and a maximum in the late afternoon

Back

Content

DATA: TEC

- Vertical TEC measured at Lisbon airport, Portugal (39° N, 9° W) by a GNSS receiver with SCINDA system
- Time interval: 01.01.2015- 31.12.2015
- Time resolution: 1h data
- Calibrated using data from Royal Observatory of Belgium (ROB) GNSS Research Group for the grid point that is very close to the location of the Lisbon airport.
  - To convert the relative TEC SCINDA data to the TEC data in TECu
  - The calibration was performed individually for each of the 12 month

Content

DATA: SPACE WEATHER PARAMETERS (*PREDICTORS*)

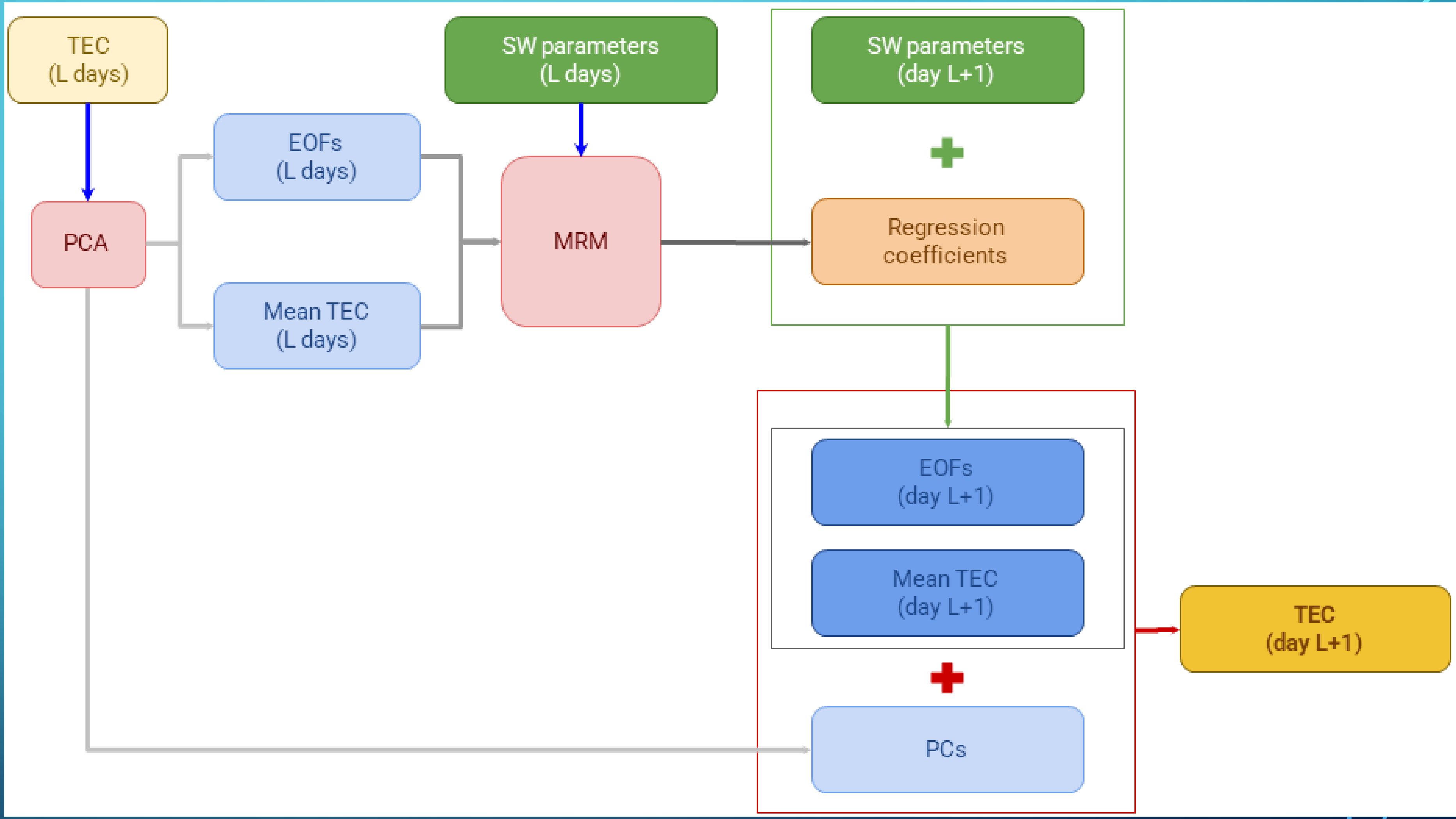
- Solar wind parameters:
  - Pressure (p), density (n), velocity (v)
- Interplanetary magnetic field:
  - Full interplanetary magnetic field (scalar B), GSM components (Bx, By, Bz)
- Geomagnetic indices:
  - Dst, ap, AE, local K_{COI}-index (Coimbra Geomagnetic Observatory, Portugal)
- Proxies for the solar UV & XR fluxes:
  - UV: Mg II composite series – a proxy for the spectral solar irradiance variability in the spectral range from UV to EUV (Snow et al., 2014);
  - F10.7 index (OMNI)
  - XR: Solar EUV Experiment (SEE) for the NASA TIMED mission at the wavelength 0.5 nm (LISIRD)
- Daily number of solar flares of classes C and M (NGDC)
- Time resolution: 1d data

Back

Content

PCA-MRM MODEL

- L = 31 or 32 days
- Lag between space weather parameters and TEC = 1 & 2 days (space weather parameters lead)
-

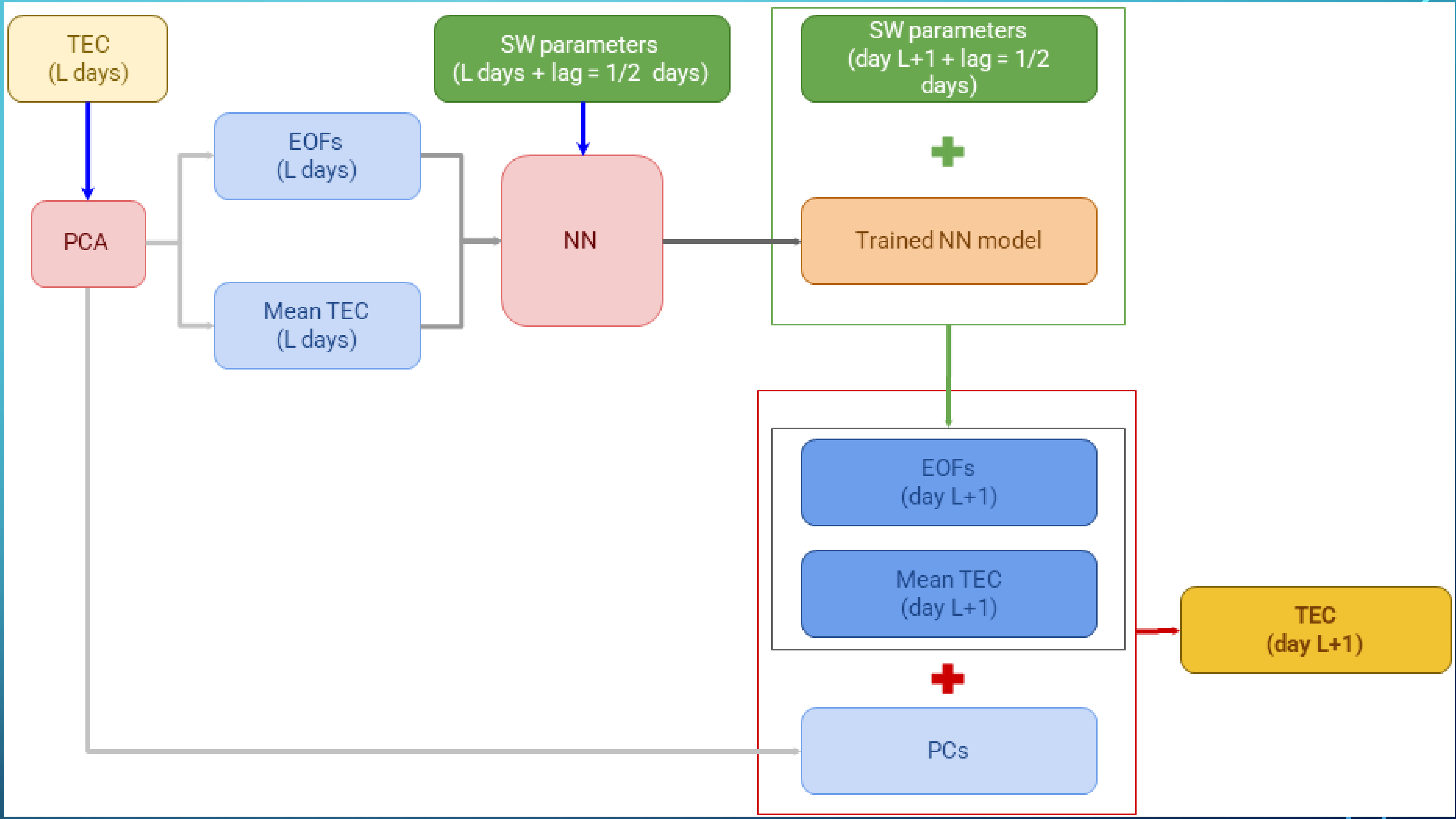


Back

Content

PCA-NN MODEL

- L = 31
- NN with "memory" (LSTM NN) or feedforward NN with weight backpropagation trained on the lagged series of predictors



Back

Content

NEURAL NETWORK

- R package: *neuralnet*
- NN algorithm: *the resilient backpropagation with weight backtracking*
- Input dataset length = 31 days → for NN models we cannot use all 16 predictors → we selected the predictors that were used by most of the PCA-MRM models
- Tested predictors: *Mg II, Dst, By lagged by 1 and 2 days (6 input series)*
  - These space weather parameters were used in 67-83% of the tested MRM models
- Tested TEC parameter: *daily mean TEC*

COMPARISON OF THE MRM AND NN FORECASTS

- Different NN were trained to find the best combination of the
  - Input series
  - NN depth (number of hidden layers and number of nodes)
  - Usage of an "ensemble forecast": a number (e.g., 100) of NN models of the same architecture were trained on the same input dataset and were used to make a forecast for the day L+1; the final forecast is the arithmetic average of 100 forecasts
- The forecasts of the daily mean TEC made by NN models for all available days of 2015 were compared to the forecasts of the daily mean TEC made by the (PCA)-MRM model using the following metrics:
  1. Correlation coefficient (r)
  2. Mean absolute error (MAE)
  3. Root-mean-square error (RMSE)

Content

COMPARISON OF THE MRM AND NN FORECASTS

	(PCA) -MRM	NN 3 predictors			NN 3 predictors "ensemble"			NN 2 predictors			NN, 1 predictor		
predictors	16 predictors & "best subset"	MgII, Dst, By			MgII, Dst, By			MgII, Dst	Dst, By	MgII, By	MgII	Dst	By
Layers & nodes		(6,4,2)	(6,4)	(6,2)	(6,4,2)	(6,4)	(6,2)	(4,2)	(4,2)	(4,2)	(2)	(2)	(2)
r	**0.88**[1]	0.88	0.88	0.88	**0.9**[2,3]	0.89	**0.9**[2,3]	**0.9**[2,3]	0.87	0.89	0.89	0.87	0.85
MAE	**2.0**[1,2]	2.21	2.16	2.17	**2.08**[3]	2.1	**2.08**[3]	**2.05**[3]	2.28	2.13	2.11	2.31	2.91
RMSE	**2.8**[1]	2.79	2.82	2.79	**2.64**[2,3]	2.7	**2.63**[2,3]	**2.61**[2,3]	2.88	2.7	2.68	2.91	3.09

[1,2,3] Notes

Back

COMPARISON OF THE MRM AND NN FORECASTS

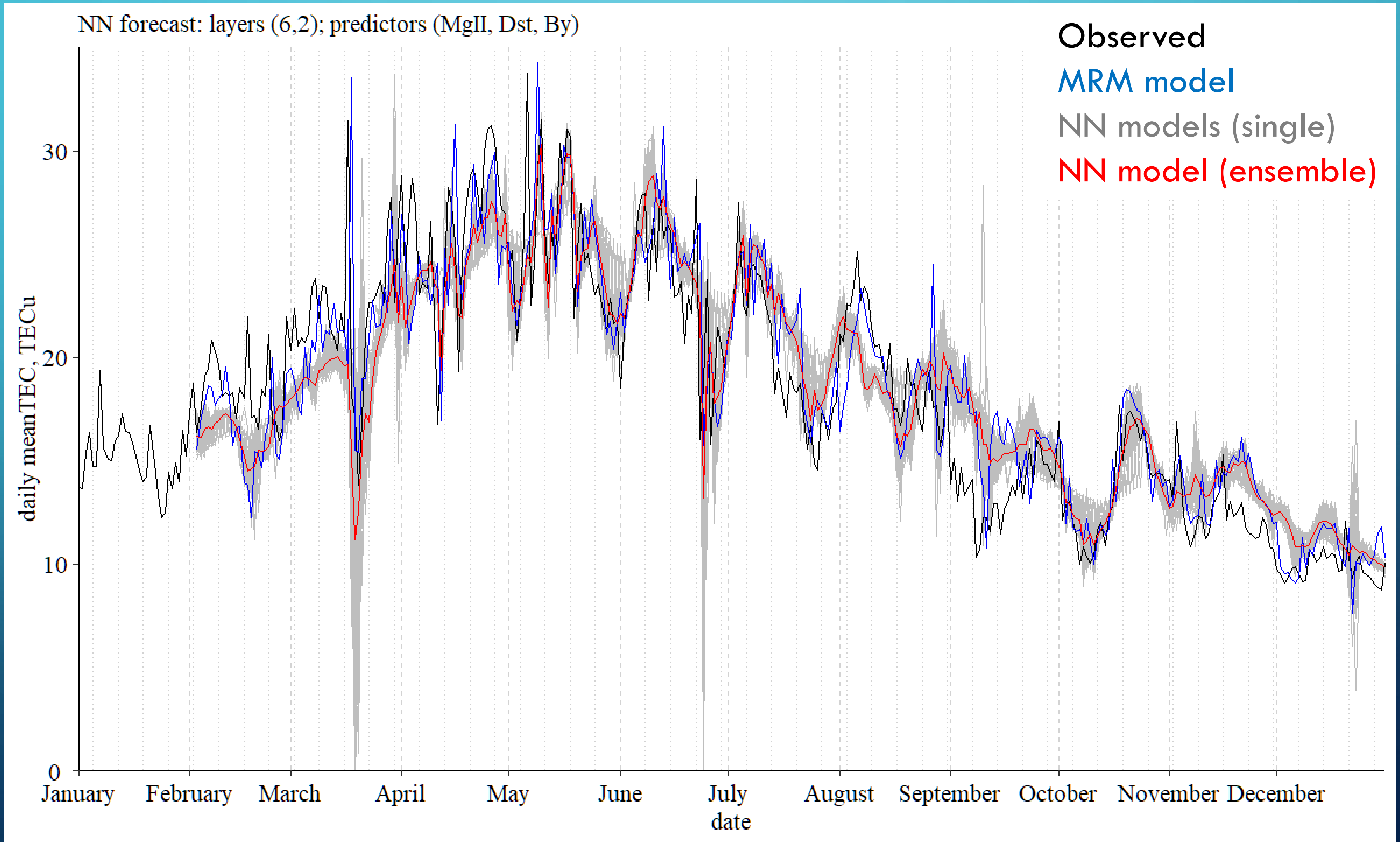
1 – metrics for the (PCA)-MRM model
2 – best metrics for all kind of models
3 – best metrics for NN models

Back

Content

COMPARISON OF THE MRM AND NN FORECASTS

- Observations vs MRM model vs NN models with single and ensemble forecasts

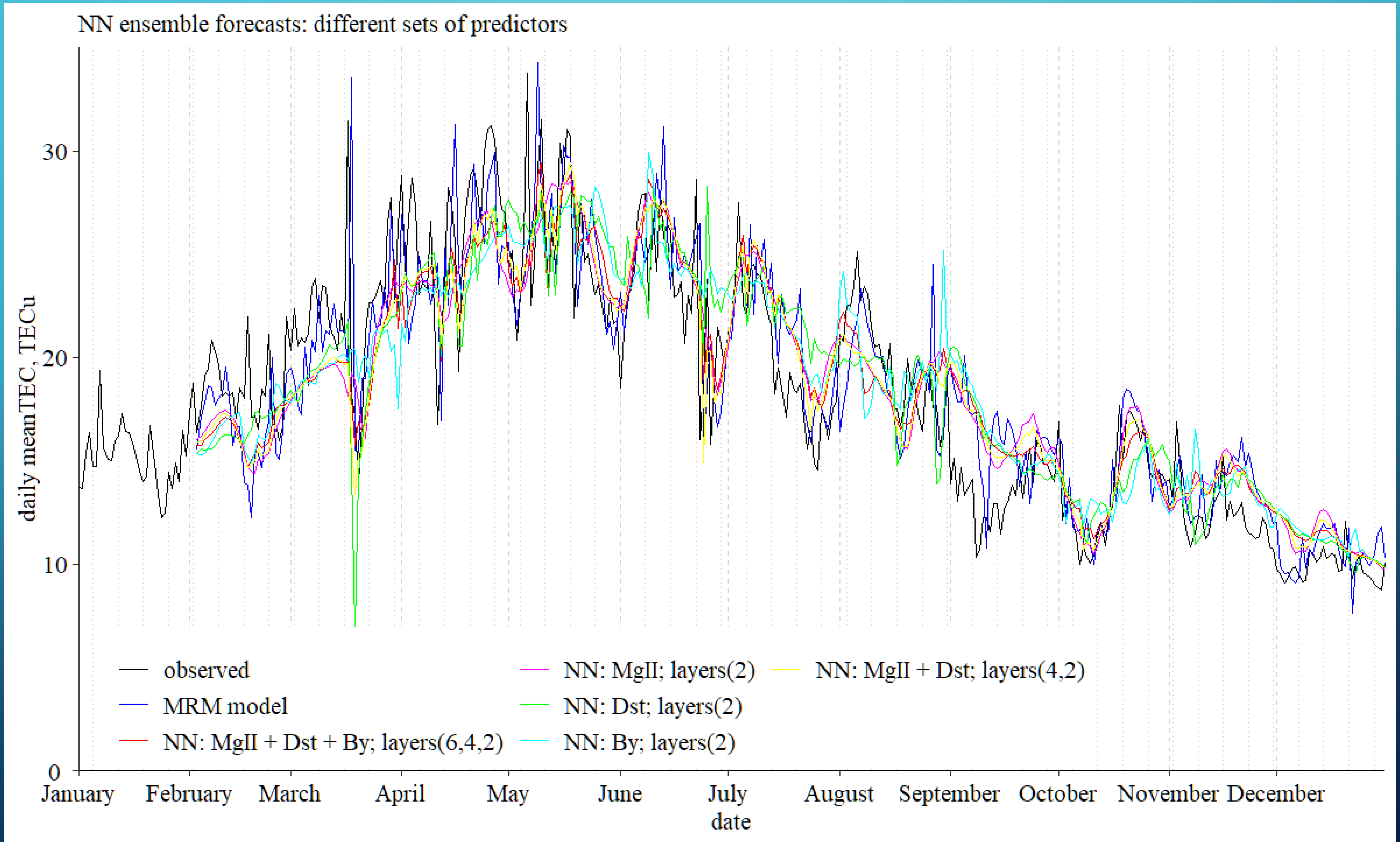


Back

Content

COMPARISON OF THE MRM AND NN FORECASTS

- Observations vs MRM model vs NN models with 1, 2 and 3 predictors

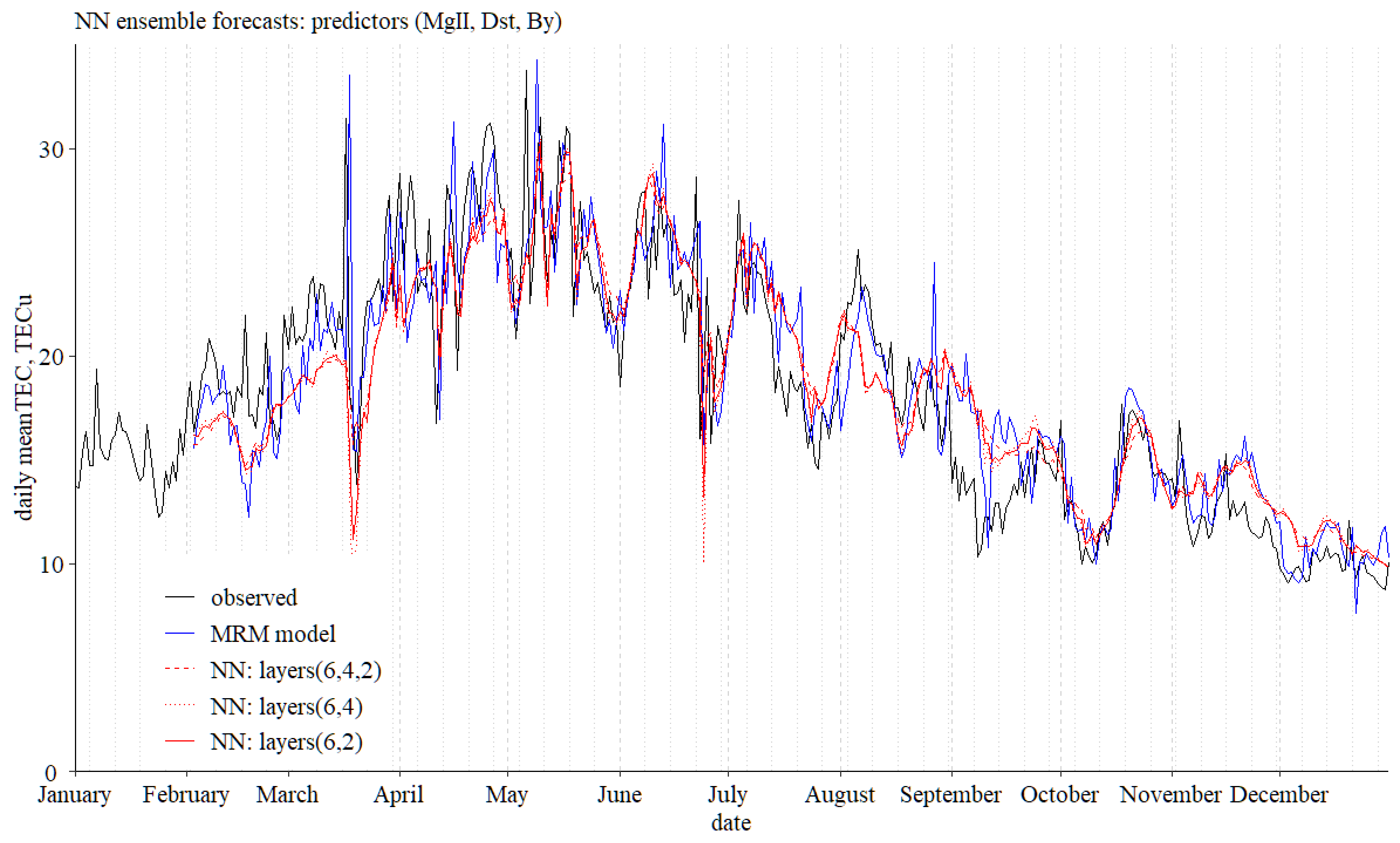


Back

Content

COMPARISON OF THE MRM AND NN FORECASTS

- Observations vs MRM model vs NN models with 3 predictors and different NN depth

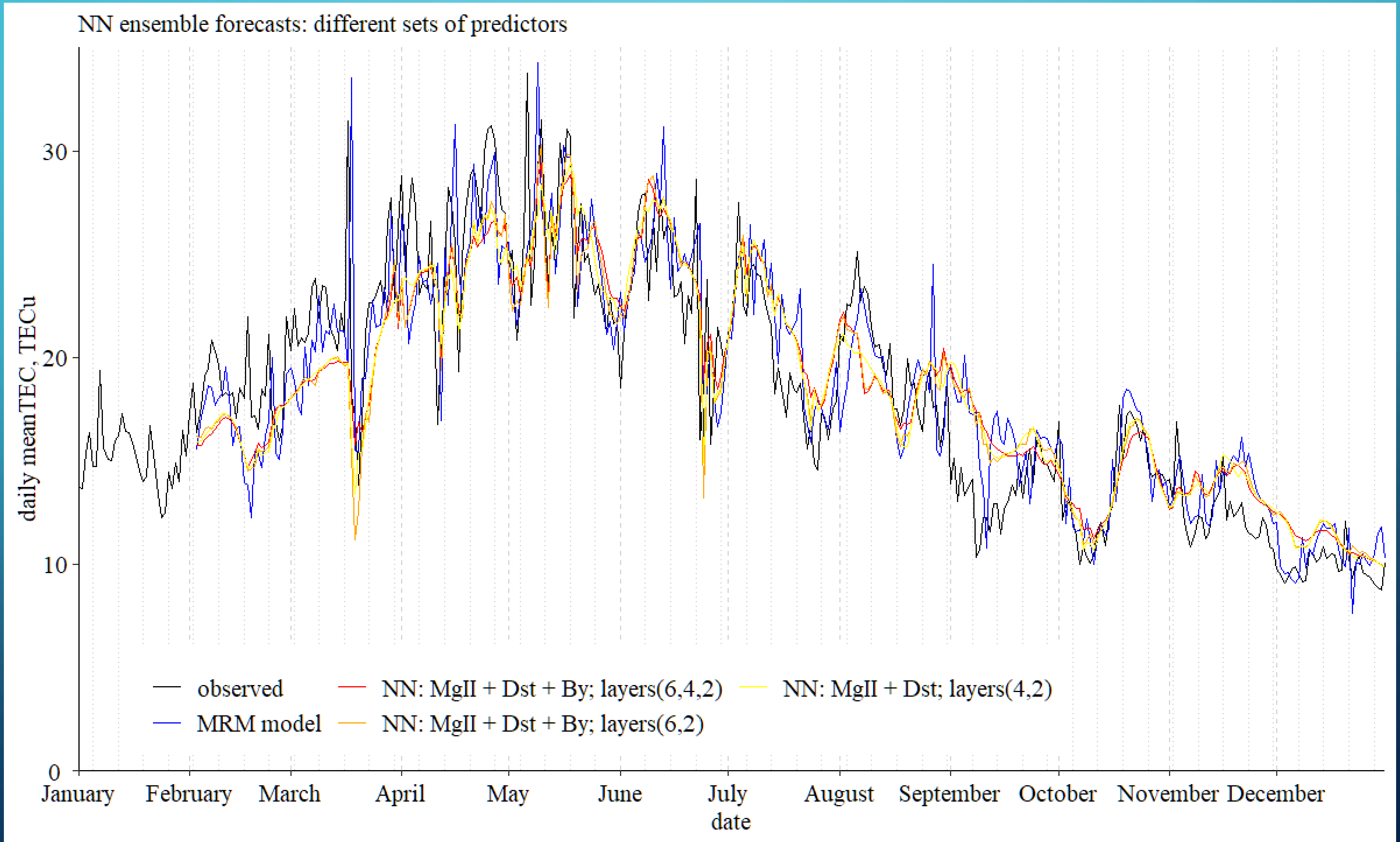


Back

Content

COMPARISON OF THE MRM AND NN FORECASTS

- Observations vs MRM model vs best NN models (marked in the Table)



Back

Content

CONCLUSIONS

- A simple NN (feedforward NN with weight backpropagation) with just 2 or 3 predictors with time lags trained on a 31-days long input dataset can forecast the daily mean TEC series with the same or even better quality than the multiple regression model with up to 16 regressors
  - For some time intervals both the MRM and NN models give similar predictions different from observations. Hypothesis: TEC variations for that time intervals have other (non-space weather) drivers
- Ensemble NN forecast perform better than the single forecast
- As predictors the solar UV proxy (MgII) is the most important predictor (models without MgII perform worse)
- The Dst index added to MgII improve the performance of NN
- Adding the By parameter slightly improve the forecast quality (the NN forecast with 3 space weather parameters is closer to the observations)

Content

NEXT STEPS

- To check other space weather parameter as predictors
- To find the optimal length and optimal list of predictors
- To test other NN architectures

Content

AUTHORS' INFO

ANNA MOROZOVA

annamorozovauc@gmail.com

ORCID iD: 0000-0002-8552-8052

Researcher at IA-U.Coimbra, Portugal

She is the author of 22 papers in peer review journals. The main areas of scientific activities are *Solar-Terrestrial Physics; Solar physics; Atmosphere physics; Ionosphere physics; Statistical data analysis; Geophysical data homogenization.*

TERESA BARATA

mtbarata@gmail.com

ORCID iD:0000-0001-6106-8285

Researcher at IA-U.Coimbra, Portugal

She is the author of 30 papers in peer review journals. The main areas of scientific activities are *Image Processing and Mathematical Morphology; Solar images; Space Weather.*

TATIANA BARLYAEVA

tvbarlyaeva@gmail.com

ORCID iD: 0000-0001-6562-594X

Invited Researcher at IA-U.Coimbra, Portugal

The author or co-author of 20 papers in peer review journals. The main areas of scientific activities are *Space Weather; Solar and Solar-Terrestrial Physics; GNSS; Remote Sensing.*

ACKNOWLEDGEMENT

- This research was supported through the project “SWAIR - Space weather impact on GNSS service for Air Navigation", ESA Small ARTES Apps
- This work was supported by Fundação para a Ciência e a Tecnologia (FCT) through the research grants UIDB/04434/2020 and UIDP/04434/2020.