EGU22-5220

EGU General Assembly 2022

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

THE ROLE OF SPACE WEATHER PARAMETERS AS PREDICTORS FOR TEC

ANNA MOROZOVA,

TERESA BARATA, TATIANA BARLYAEVA

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

ABSTRACT

The total electron content (TEC) over the Iberian Peninsula was modeled using a PCA-NN models based on the decomposition of the observed TEC series using the principal component analysis (PCA) and reconstruction of the daily mean TEC and daily PCA modes' amplitudes by different types of neural networks (NN) using several types of space weather parameters as predictors. Lags of 1 and 2 days between the TEC and space weather parameters are used.

Two main goals are set:

1. To find a NN configuration(s) that produces forecasts of reasonable quality with minimal amount of input data
2. To find a best set of space weather parameters that work as predictors for PCA-NN models

Here we present preliminary results related to the **second goal**: PCA-NN models with different sets of predictors are compared. Among predictors we consider 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, geomagnetic indices, solar wind coupling functions and TEC series themselves.

TOC

PCA-NN MODELS FOR TEC

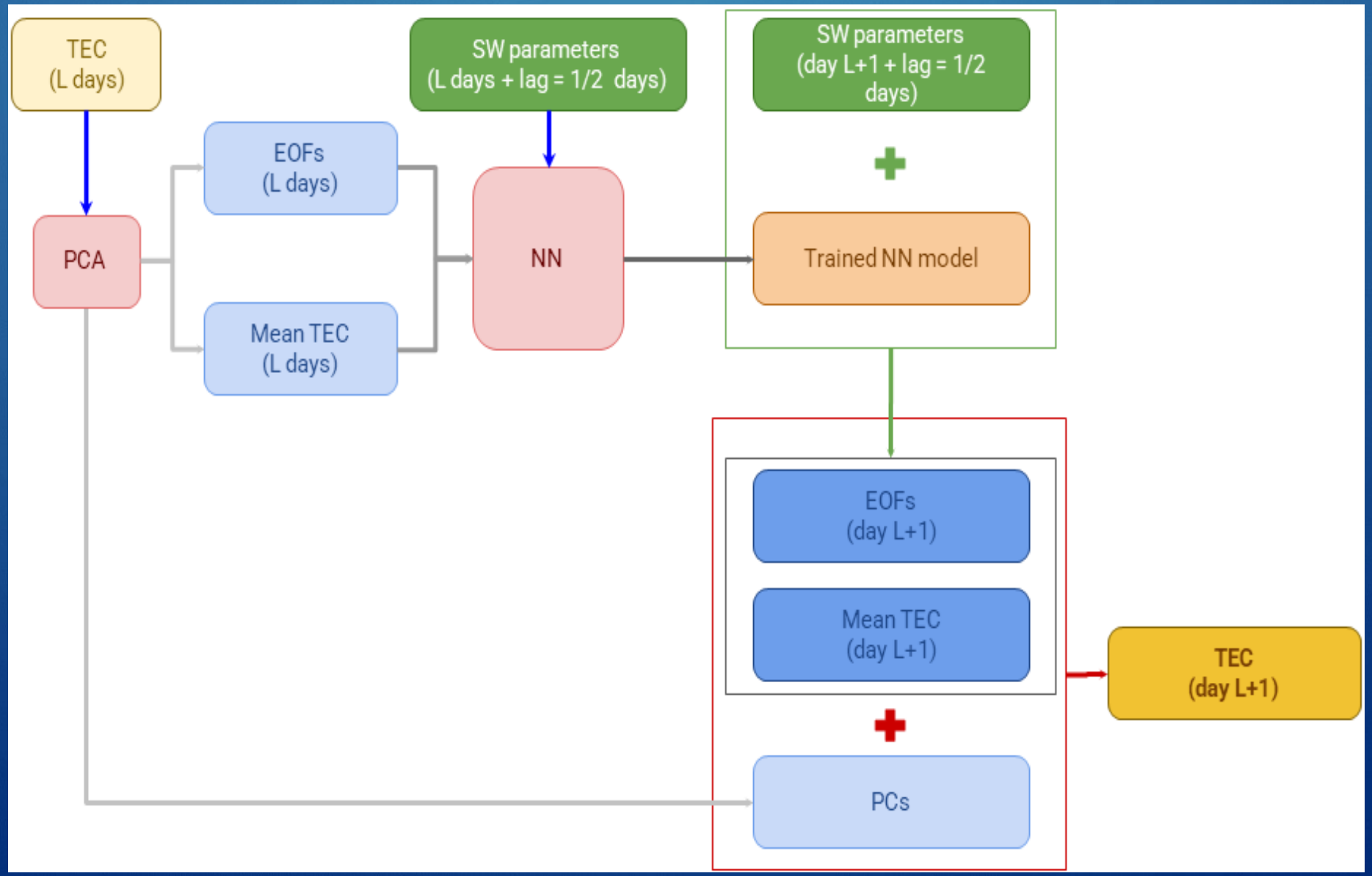
PCA-BASED MODELS

- The total electron content (TEC) over the Iberian Peninsula was modeled using a PCA-NN model based on
  1. decomposition of the observed TEC series using the principal component analysis (PCA)
  2. reconstruction of the daily mean TEC and daily PCA modes' amplitudes by, for example, regression models (PCA-MRM models*) or neural networks (PCA-NN models) using several types of space weather parameters (SW) as predictors.
- Lags of 1 and 2 days between the TEC and SW predictors are used

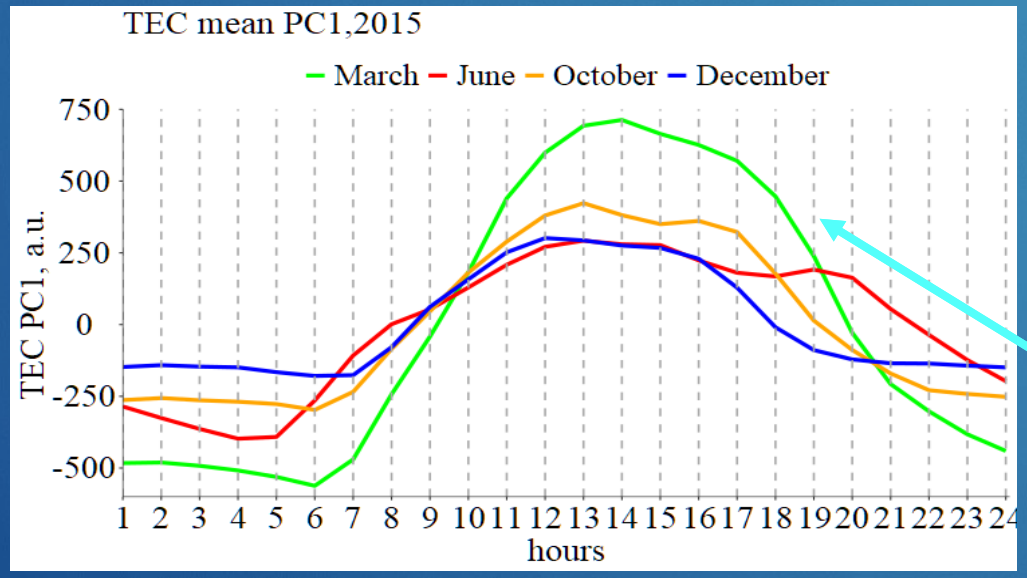
* *A. L. Morozova, T. Barata, T. Barlyaeva (2022) PCA-MRM model to forecast TEC at middle latitudes, Atmosphere, 13(2), 323; https://doi.org/10.3390/atmos13020323*

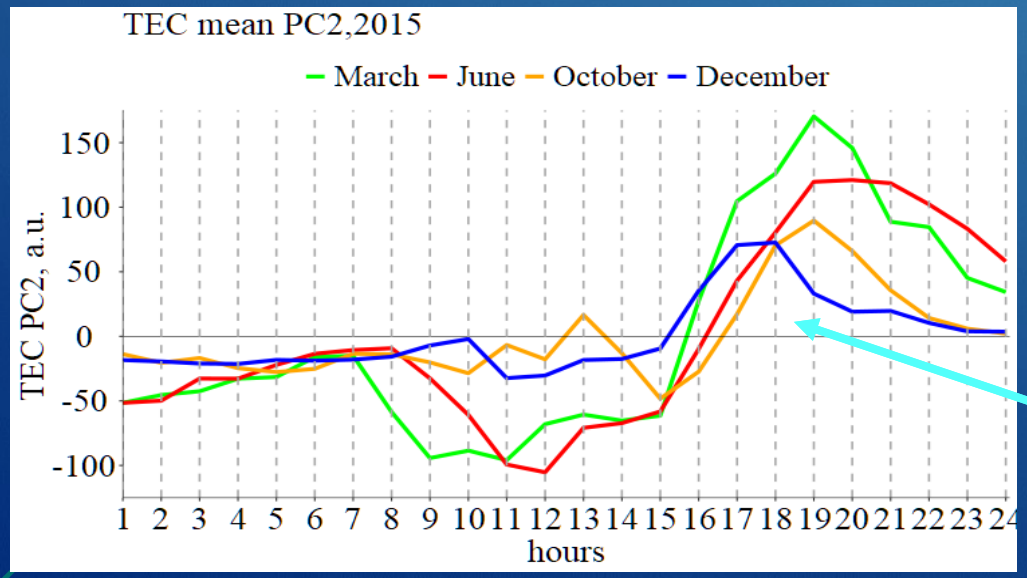
PCA-NN MODEL

L = 31 days



TEC PCA MODES 1 AND 2





Mode 1 (PC1 & EOF1):

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

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

Mode 2 (PC2 & EOF2):

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

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 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

DATA

DATA: TEC

- Vertical TEC measured at Lisbon airport, Portugal (39° N, 9° W) by a GNSS receiver with SCINDA firmware
- Time interval: 01.01.2015- 31.12.2015
- Time resolution: 1h data

- TEC series: daily mean TEC, EOF1, EOF2

DATA: SPACE WEATHER PARAMETERS (SW PREDICTORS)

- Solar wind parameters:
  - Pressure (p), density (n), velocity (v)
- Interplanetary magnetic field:
  - Full interplanetary magnetic field (scalar B), GSM components (Bx, By, Bz)
- Coupling functions*: $d\Phi_{MP}/dt$, E_{WAV} , E_{WV}, ε_3, E_{KLV}, E_{KL}, $v \cdot B_S$, E_{SR}, E_{TL}
- 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 or F10.7 index
  - XR: Solar EUV Experiment (TIMED mission)
  - Daily number of solar flares of classes ≤C and M, and the number of all flares N
- Time resolution: 1d data

Newell, P. T., T. Sotirelis, K. Liou, C.-I. Meng, and F. J. Rich (2007), A nearly universal solar wind-magnetosphere coupling function inferred from 10 magnetospheric state variables, J. Geophys. Res., 112, A01206, doi:10.1029/2006JA012015

SPACE WEATHER PARAMETERS: CORRELATED PREDICTORS

- Some of the SW predictors correlate with each other
- Highest correlations:
  - n & p ($|r| = 0.75$)
  - Dst & ap & AE & Kcoi ($|r| = 0.75 \div 0.85$)
  - Mg II & F10.7 & XR ($|r| \geq 0.82$)
  - Number of C and all flares ($|r| = 0.98$)
  - ($d\Phi_{MP}/dt$, E_{WAV} , E_{WV}, ε_3, E_{KLV}, E_{KL}, $v \cdot B_S$, E_{SR}, E_{TL}) vs (K_{coi}, ap, AE) ($|r| \geq 0.75$)

To test: Can we use correlated predictors?

TEC SERIES: AUTO-REGRESSION

- Lag1 autocorrelation coefficients
  - Daily mean TEC: $\alpha_1 = 0.91$
  - EOF1 : $\alpha_1 = 0.14$
  - EOF2 : $\alpha_1 = 0.38$

- Hypothesis to test: Adding lagged TEC series as the NN input parameters could improve NN forecasting skills for daily mean TEC and EOF2, but not for EOF1

NN MODELS: CONFIGURATION

PCA-NN MODEL – PREVIOUS RESULTS

- NN algorithm: feedforward NN with the resilient backpropagation with weight backtracking (ready-to-use package `neuralnet` (R))
- The input dataset length L = 31 days
- SW predictors are submitted as lagged series (lag = 1 and 2 days) together (for X predictors there are 2*X input series)
- "Monte-Carlo approach": 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

PCA-NN MODEL – CURRENT QUESTIONS

- ? Best set of SW predictors for the daily mean TEC and EOFs series
  - ? Can correlated SW predictors be used
  - ? Can TEC series be used as predictors

- ? Best NN configuration (hyperparameters) that produces forecasts of reasonable quality with minimal number of SW predictors for the daily mean TEC and EOFs series

SELECTION OF THE BEST SETS OF SW PREDICTORS
&
OPTIMIZATION OF HYPERPARAMETERS

BEST NN ARCHITECTURE AND BEST SETS OF SW PREDICTORS (no solar wind coupling functions, no auto-regression)

	Daily mean TEC	EOF1	EOF2
NN (layers & nodes)	(12,6,4)	(6,4)	(8,4,2)
Predictors	6 predictors	3 predictors	4 predictors
Best predictors	MgII Dst N. C flares ap AE XR	AE MgII Dst	AE Dst F10.7 v
r	0.91	**0.31**	0.33
MAE/RMSE (TECu)	1.82/2.40	**0.03/0.04**	0.13/0.18

BEST NN ARCHITECTURE AND BEST SETS OF SW PREDICTORS
(+ solar wind coupling functions, no auto-regression)

	Daily mean TEC	EOF1	EOF2
NN (layers & nodes)	(14,7,3) (14,6,4) (14,4,2)	(8,4)	(10,8,4)
Predictors	7 predictors	4 predictors	5 predictors
Best predictors	MgII Dst N. C flares ap AE XR E_{SR} or E_{KLV} or E_{WV}	AE MgII Dst E_{SR}	AE Dst F10.7 v E_{SR} or E_{TL}
r	0.91	0.29	0.32
MAE/RMSE (TECu)	1.84/2.42	0.03/0.04	0.14/0.18

BEST NN ARCHITECTURE AND BEST SETS OF SW PREDICTORS (no solar wind coupling functions, + auto-regression)

	Daily mean TEC	EOF1	EOF2
NN (layers & nodes)	(14,7,3) (14,6,4) (14,4,2)	(8,4)	(10,8,4)
Predictors	7 predictors	4 predictors	5 predictors
Best predictors	MgII Dst N. C flares ap AE XR Daily mean TEC	AE MgII Dst EOF1	AE Dst F10.7 v EOF2
r	**0.93**	0.28	**0.39**
MAE/RMSE (TECu)	**1.73/2.40**	0.03/0.04	**0.13/0.17**

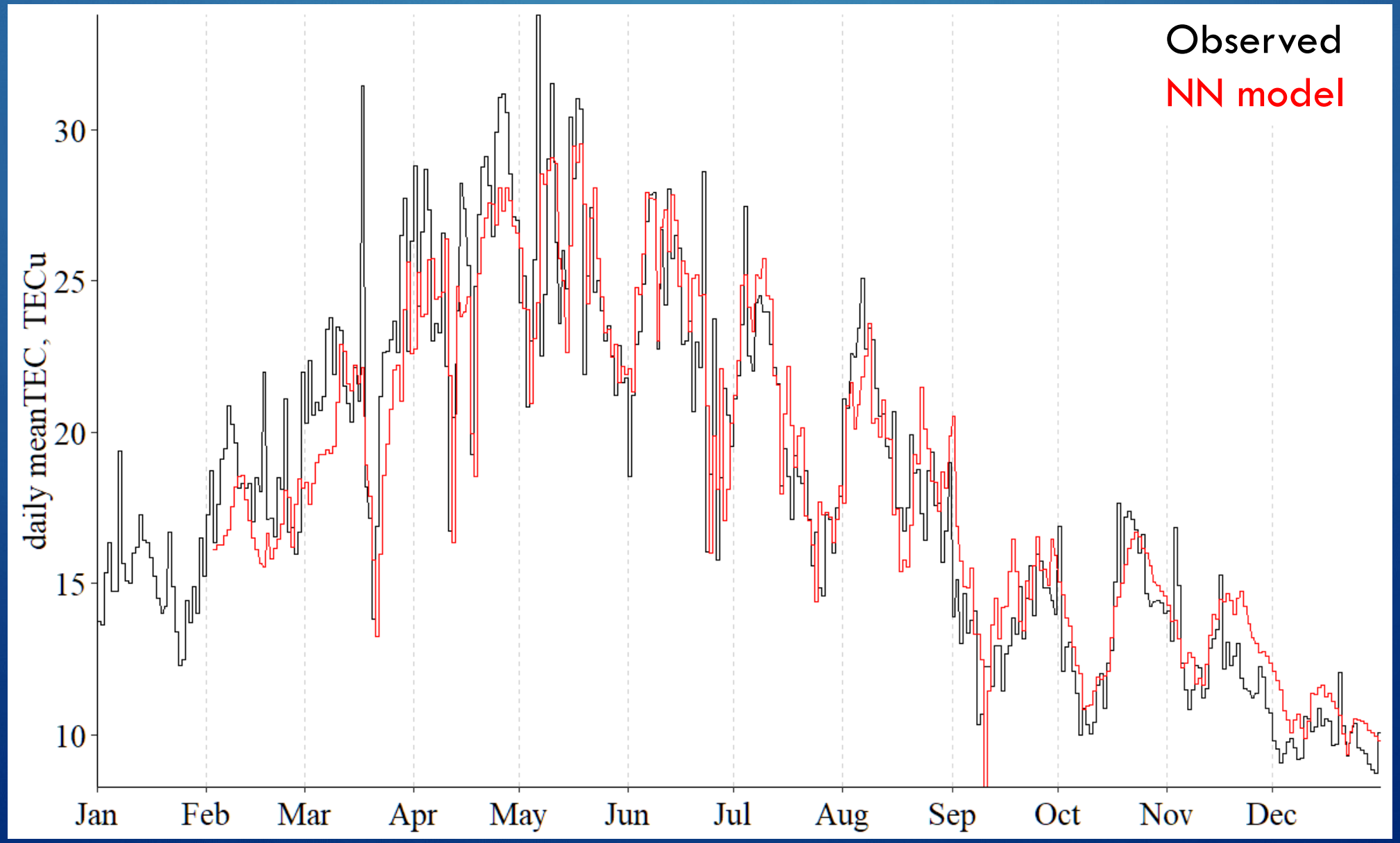
Legend:

About the same skills 

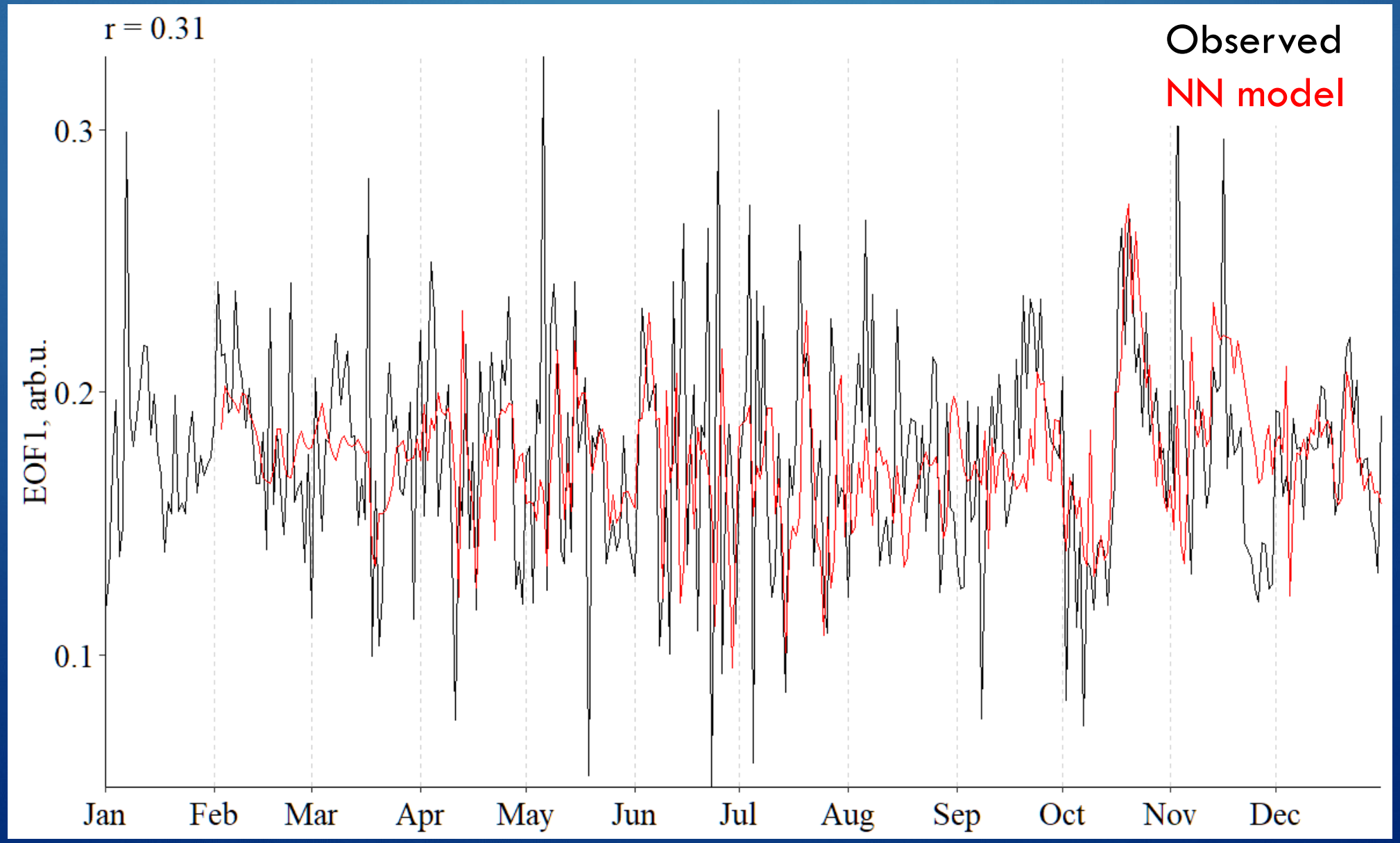
Worse skills 

Better skills 

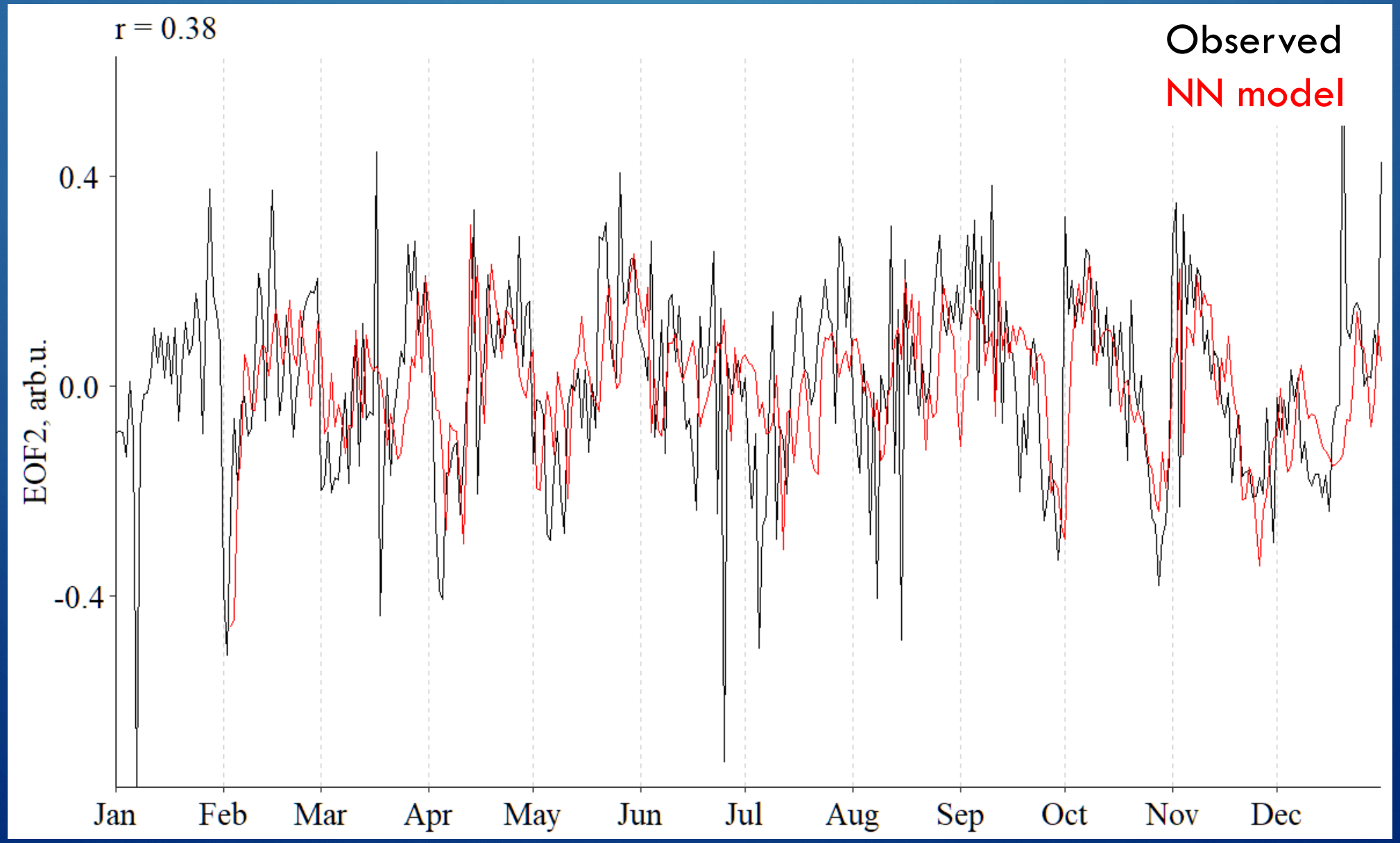
COMPARISON OF PC-NN FORECASTS AND OBSERVATIONS: daily mean TEC: forecasted vs observed



COMPARISON OF PC-NN FORECASTS AND OBSERVATIONS: EOF1: forecasted vs observed



COMPARISON OF PC-NN FORECASTS AND OBSERVATIONS: EOF2: forecasted vs observed



PCA-NN MODEL: FINAL CONFIGURATION

- SW predictors:
  - Geomagnetic indices
  - Proxies of the solar UV & XR fluxes/number of the flares
  - Only for the daily mean TEC and EOF2 – auto-regressive component
  - Only for EOF2 – solar wind velocity
- NN with 2 or 3 layers
- The forecasting quality of NN for EOF1 and EOF2 is lower than for the daily mean TEC due to their higher day-to-day variability

BEST NN ARCHITECTURE AND BEST SETS OF SW PREDICTORS

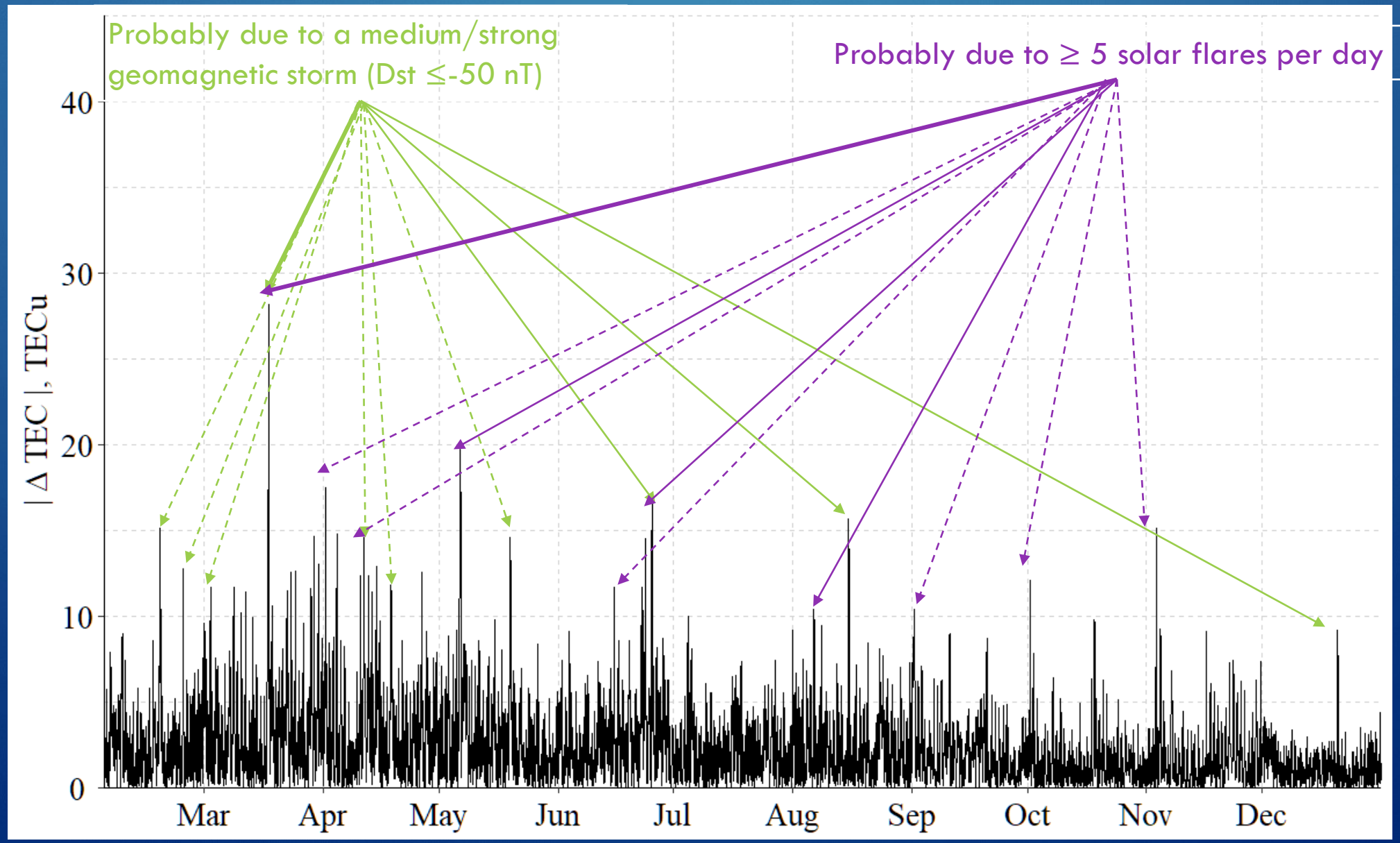
	Daily mean TEC	EOF1	EOF2
NN (layers & nodes)	(14,7,3) (14,6,4) (14,4,2)	(6,4)	(10,8,4)
Predictors	7 predictors	3 predictors	5 predictors
Best predictors	MgII Dst N. C flares ap AE XR Daily mean TEC	AE MgII Dst	AE Dst F10.7 v EOF2
r	**0.93**	**0.31**	**0.39**
MAE/RMSE (TECu)	**1.73/2.40**	**0.03/0.04**	**0.13/0.17**

PERFORMANCE OF THE PCA-NN MODEL

PCA-NN MODEL: FORECASTS FOR 1h TEC SERIES

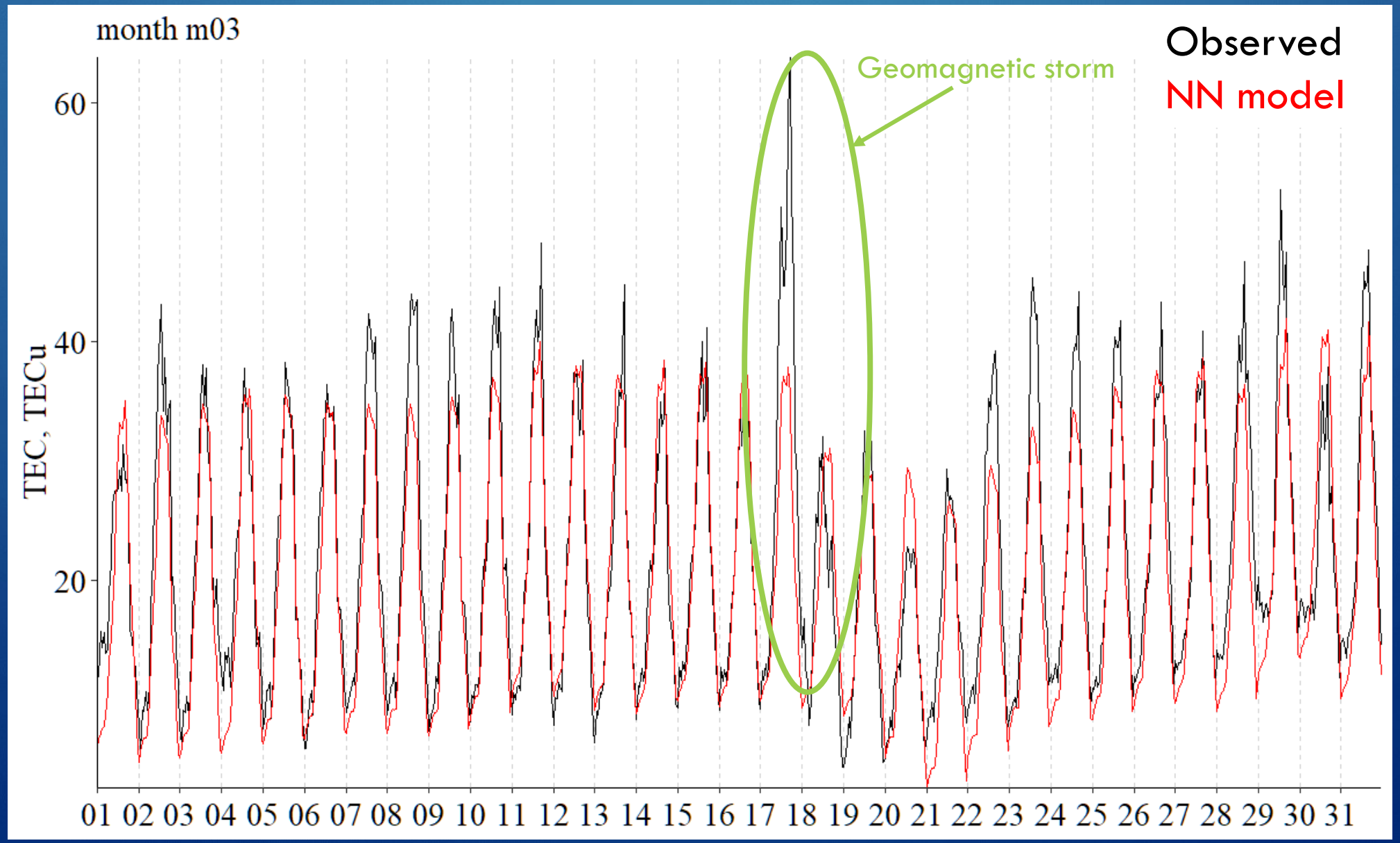
- Forecast errors are higher during days of geomagnetic storms or days with high number of the solar flares

COMPARISON OF PC-NN FORECASTS AND OBSERVATIONS: 1h TEC: absolute errors

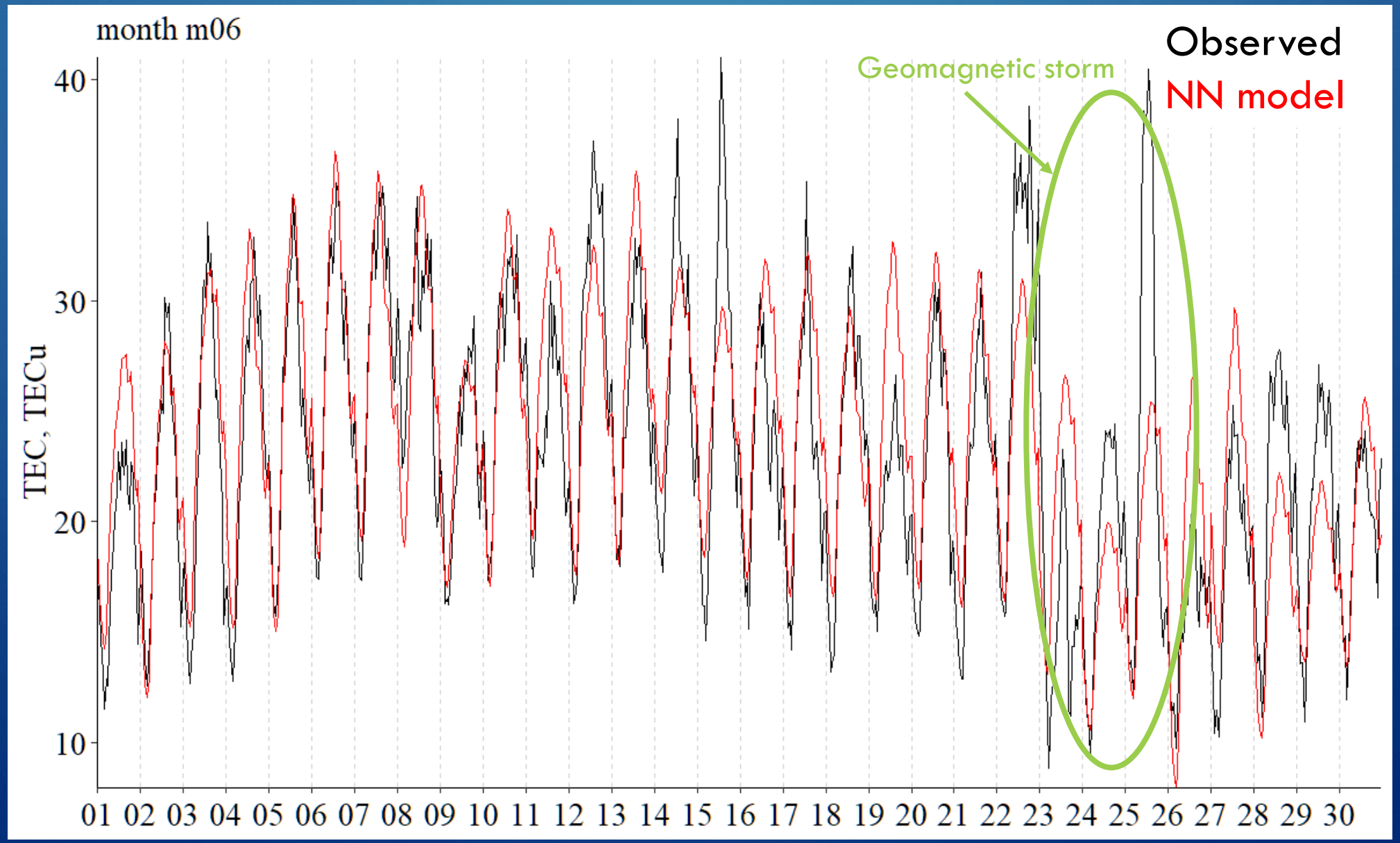


EXAMPLES FOR SOME MONTHS OF 2015

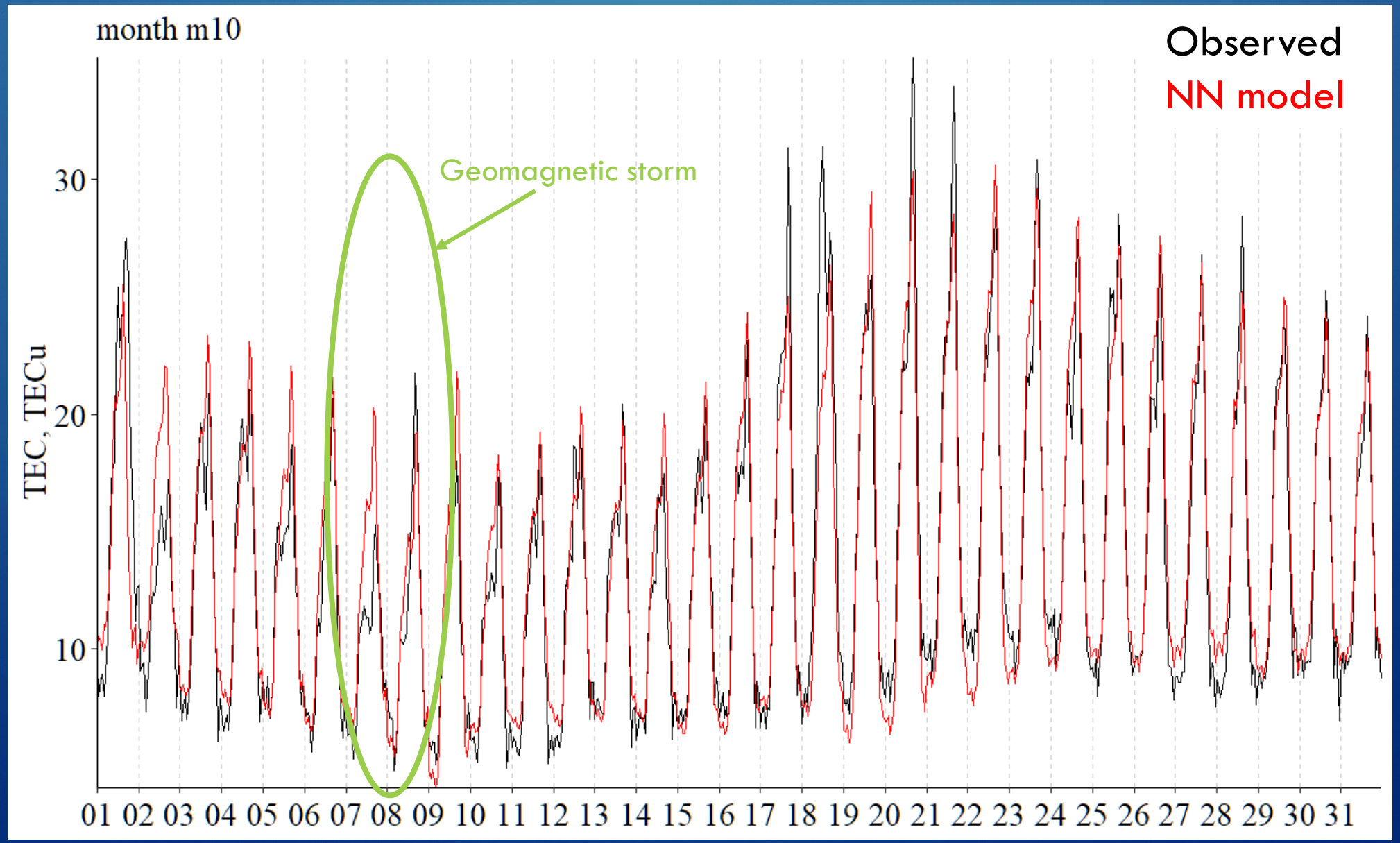
COMPARISON OF PC-NN FORECASTS AND OBSERVATIONS: 1h TEC, March 2015



COMPARISON OF PC-NN FORECASTS AND OBSERVATIONS: 1h TEC, June 2015



COMPARISON OF PC-NN FORECASTS AND OBSERVATIONS: 1h TEC, October 2015

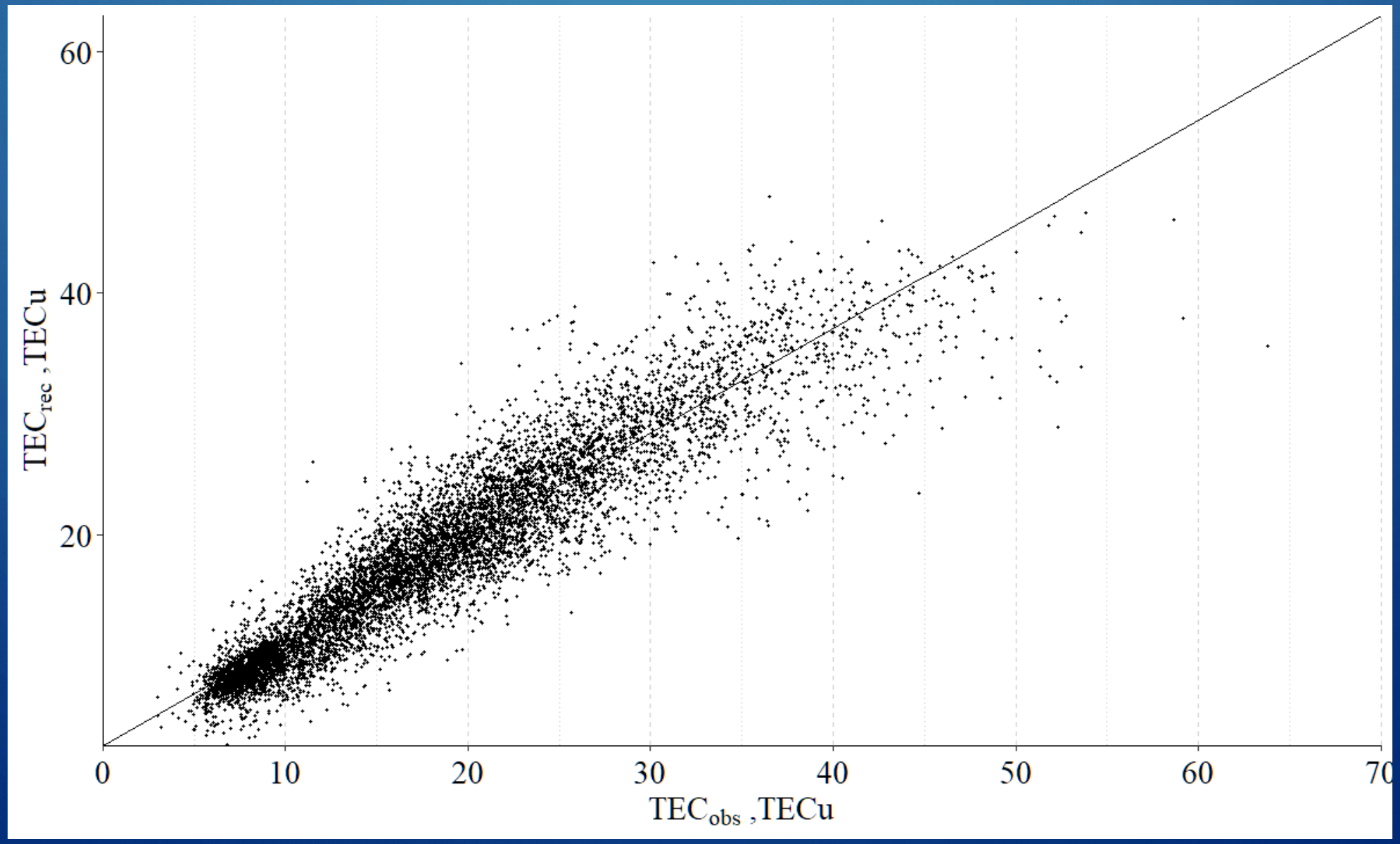


COMPARISON OF TEC_{OBS} AND TEC_{REC}

COMPARISON OF TEC_{OBS} AND TEC_{REC}

- Correlation between the observed (TEC_{obs}) and the forecasted (TEC_{rec}) series:
  - Daily mean TEC: r = 0.93
  - 1h TEC: r = 0.93
  - 1h TEC – daily mean TEC: r = 0.94
- PCA-NN forecasting skills:
  - MAE
    - Daily mean TEC: 1.7 TECu
    - 1h TEC: r = 2.5 TECu
    - 1h TEC – daily mean TEC: r = 1.7 TECu
  - RMSE
    - Daily mean TEC: 2.2 TECu
    - 1h TEC: r = 3.45 TECu
    - 1h TEC – daily mean TEC: r = 2.4 TECu
  - MaxE
    - Daily mean TEC: 9.5 TECu
    - 1h TEC: r = 28.3 TECu
    - 1h TEC – daily mean TEC: r = 18.9 TECu

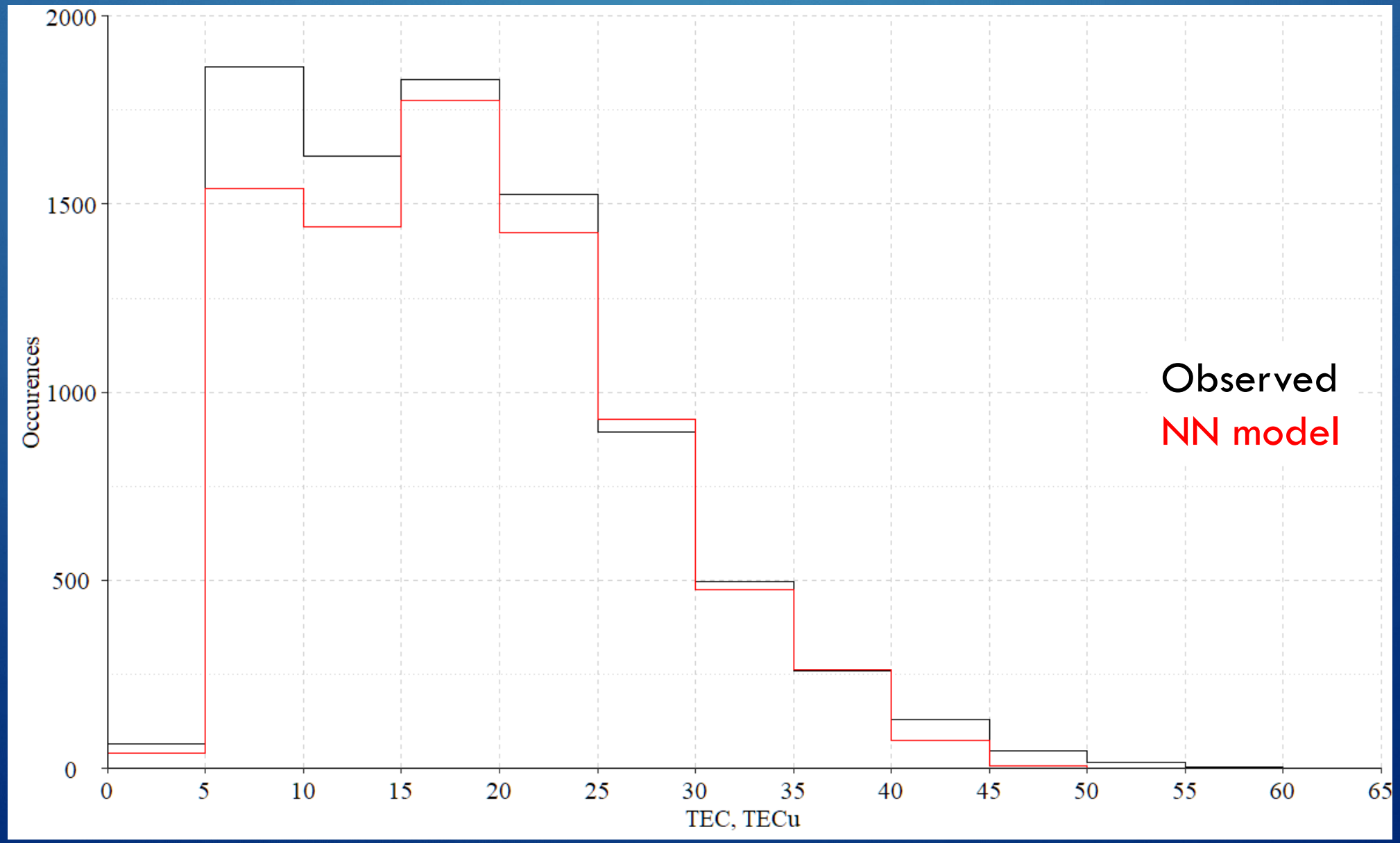
COMPARISON OF PC-NN FORECASTS AND OBSERVATIONS: 1h TEC: reconstructed (forecasted) vs observed



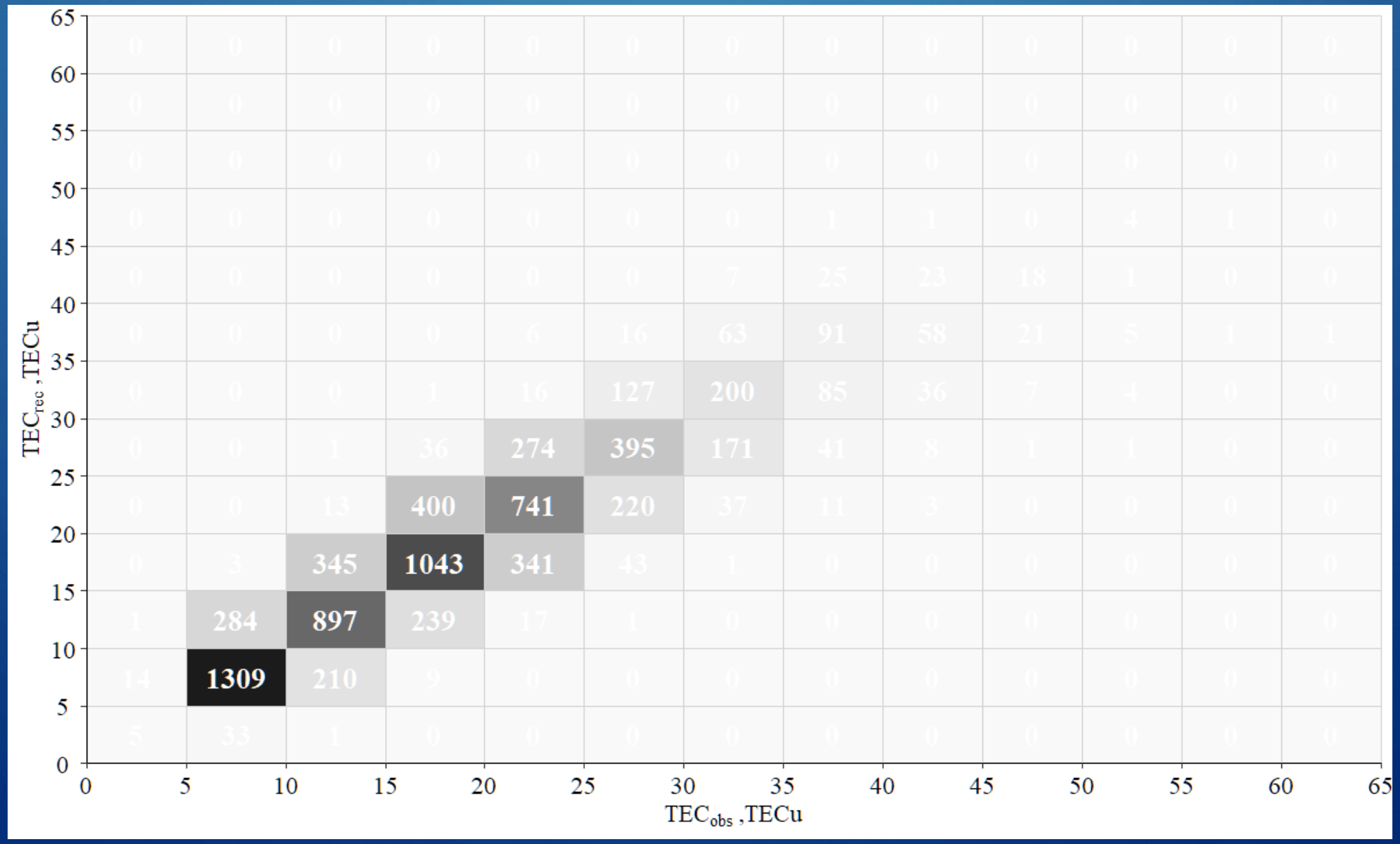
COMPARISON OF TEC_{OBS} AND TEC_{REC}

- Occurrences of certain TEC values are about the same for the observed and forecasted series
- In most cases the differences between TEC_{obs} and TEC_{rec} do not exceed 5 TECu, however:
  - For the nights & quiet days (TEC ≤ ~20 TECu) the PCA-NN model over-estimates TEC by ~5 TECu more frequently than under-estimates TEC by ~5 TECu
  - For more disturbed days (TEC > ~20 TECu) the PCA-NN model under-estimates TEC by ~5 TECu more frequently than over-estimates TEC by ~5 TECu

COMPARISON OF PC-NN FORECASTS AND OBSERVATIONS: 1h TEC: occurrences for the forecasted & observed values



COMPARISON OF PC-NN FORECASTS AND OBSERVATIONS: 1h TEC: occurrences for the forecasted vs observed values



CONCLUSIONS

- SW predictors
  - Best sets of SW predictors for PCA-NN models:
    - Daily mean TEC – 7 predictors (including daily mean TEC)
    - EOF1 – 3 predictors
    - EOF2 – 5 predictors (including EOF2)
  - Most important predictors:
    - Dst, AE, solar UV proxies – used for all TEC parameters
      - daily mean TEC & EOF1 ⇔ Mg II; EOF2 ⇔ F10.7
  - The use of correlated predictors may improve the prediction quality
  - Solar wind coupling functions do not improve NN forecasts
- NN with small number of layers perform better (only 2 or 3 layers for 6 to 14 input SW series)
- Highest errors are observed for days with strong geomagnetic storms (Dst ≤ -50 nT) and/or many solar flares (≥ 5 per day)

ACKNOWLEDGEMENTS

- This research is supported through the projects
  - "**PRIME**: Portuguese Regional Ionosphere Model " (EXPL/CTAMET/0677/2021), FCT
- IA is supported through the projects
  - UIDB/04434/2020 and UIDP/04434/2020 , FCT

1 2 9 0 UNIVERSIDADE D COIMBRA

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

Cofinanciado por:

COMPETE 2020

PORTUGAL 2020

UNIÃO EUROPEIA
Fundo Europeu
de Desenvolvimento Regional