Prediction of Near-Surface Temperatures Using a Non-Linear Machine Learning Post-Processing Model

Emy Alerskans[1, 3], Joachim Nyborg[2, 3], Morten Birk[3], Eigil Kaas[1, 4]

[1] The Niels Bohr Institute, University of Copenhagen
[2] Department of Computer Science, Aarhus University
[3] FieldSense A/S
[4] Danish Meteorological Institute



Motivation

- The weather has a significant impact on the growth, development and yield of crops, as well as the water and fertilisation need (Petr, 1991; Stigter et al., 2000; WMO, 2010).
- Accurate weather forecasts can aid in the planning of daily activities related to agricultural systems.
- The demand for more accurate and localized weather forecasts in relation to agricultural activities is the main motivation for this study.



Motivation

- Numerical Weather Prediction (NWP) models generally exhibit systematic biases in the forecasts, especially for near-surface variables such as 2 m air temperature.
- In order to correct these systematic errors, post-processing methods can be applied.
- In this study we have implemented a Transformer neural network model for post-processing 2 m air temperature forecasts from the Global Forecast System.
- The model has been trained on observational data from a dense network of private weather stations.
- Validation has been performed on independent data from the same network of private weather station, as well as on observational data from the Danish Meteorological Institute.



Observational Data



- Private Weather Stations (PWSs) from FieldSense.
- Measurements of e.g. 2m air temperature, relative humidity, wind speed, pressure and rainfall.
- Data from April, 2019 - December, 2020.
- Observations from agricultural sites, mainly in Denmark.



NWP data

Global Forecast System (GFS):

- Global model.
- Horizontal resolution: 0.25°.
- Vertical resolution: 64 layers.
- Forecast length: 16 days.
- Updated: 00, 06, 12, 18 UTC.
- Output: 1-hourly - 12-hourly.

Study specifications:

- Prediction range for post-processing: 2 days.
- Time period: April, 2019 - December, 2020.
- Closest grid cell used for each station location.



Post-Processing Method - The Transformer

- The Transformer (Vaswani et al., 2017).
- Has a sequence to sequence (Seq2Seq) architecture - transforms a given sequence of elements into another sequence.
- Uses an encoder-decoder structure.
- Encoder: Maps the input sequence to a higher dimensional vector.
- Decoder: Translates this vector to an output sequence.
- The Transformer model uses the so-called self-attention mechanism.
- Attention: Deciding, at each step, which other parts of the input sequence are important.



The Transformer - Attention

- Input matrix: nfeatures $\times$ ntime.
- nfeatures $=$ number of input features/variables.
- ntime $=$ length of time series
- Query: A set of vectors to calculate attention for.
- Key: A set of vectors to calculate attention against.
- Weights: Show how attended each query is against the keys.
- Multiplication of these weights by values gives the output.

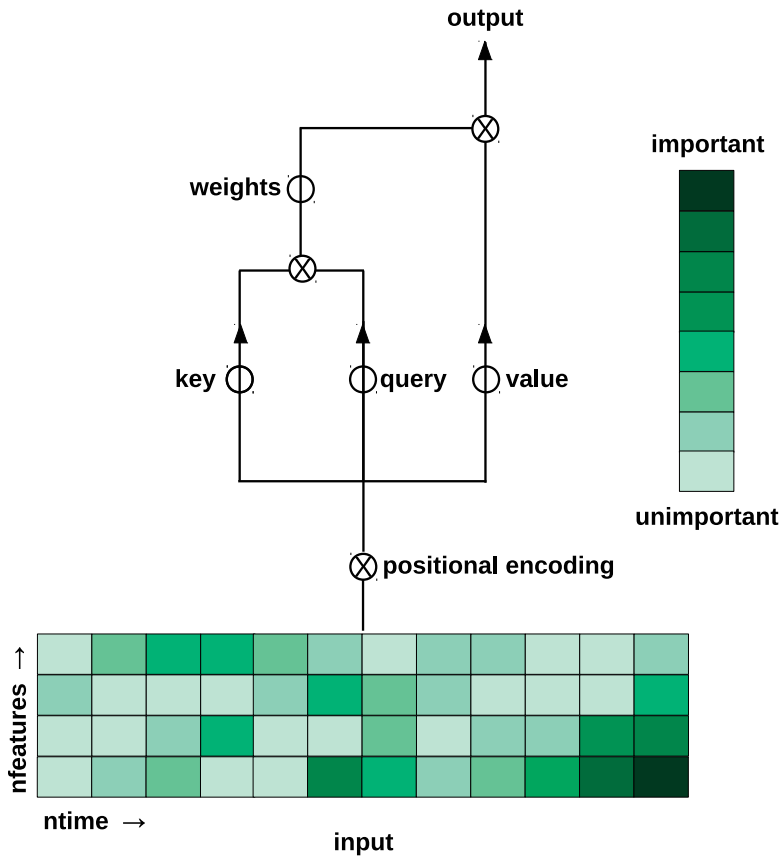

Figure: Flow diagram of the attention mechanism in the Transformer model. The attention mechanism decides which features and time steps that are important for the prediction.

The Transformer

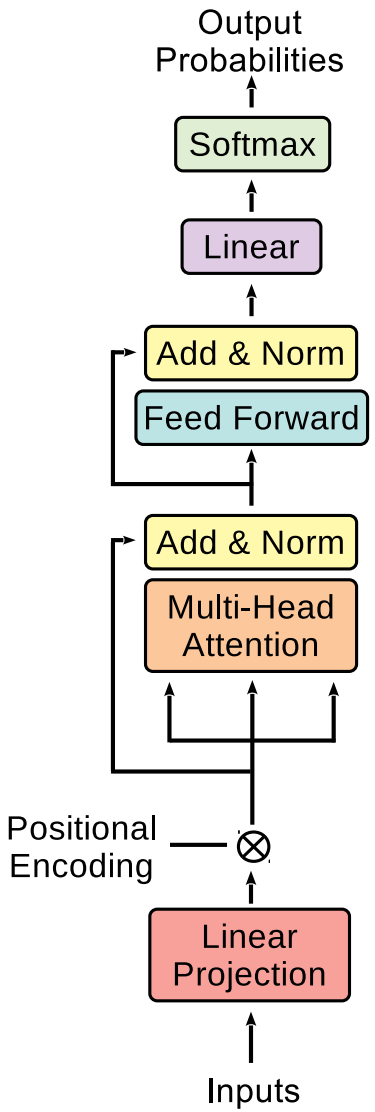

Figure: Flow diagram of the Transformer model used in this study. Adopted from Vaswani et al. (2017). The Transformer model used here is based on the minGPT model (https://github.com/karpathy/minGPT).

Input Features

Table: List of input features considered and which are included in the final model. * marks input features for which a cyclic encoding was used.

Variable	Variable description	Included	Variable	Variable description	Included
PWS t2m	observed 2m air temp.	x	NWP gh850	geopotential height of 850 hPa surface	x
NWP t2m	2m air temp.	x	NWP t500	temperature of 500 hPa surface	
NWP u10m	zonal 10m wind component		NWP t700	temperature of 700 hPa surface	
NWP v10m	meridional 10m wind component		NWP t850	temperature of 850 hPa surface	
NWP wspd10m	10m wind speed	x	NWP u500	zonal wind component at 500 hPa surface	
NWP wsdir10m*	10m wind direction	x	NWP u700	zonal wind component at 700 hPa surface	
NWP rh2m	2m rel. humidity	x	NWP u850	zonal wind component at 850 hPa surface	
NWP q2m	2m spec. humidity	x	NWP v500	meridional wind component at 500 hPa surface	
NWP td2m	2m dew point temp.	x	NWP v700	meridional wind component at 700 hPa surface	
NWP mslp	mean sea level pressure	x	NWP v850	meridional wind component at 850 hPa surface	
NWP lhtfl	latent heat flux		NWP w700	vertical velocity at 700 hPa surface	
NWP shtfl	sensible heat flux	x	NWP wspd500	wind speed at 500 hPa surface	
NWP nswrf	netto short wave radiation flux	x	NWP wspd700	wind speed at 700 hPa surface	
NWP nlwrf	netto long wave radiation flux		NWP wspd850	wind speed at 850 hPa surface	
NWP tcc	total cloud cover		NWP wsdir500*	wind direction at 500 hPa surface	
NWP tp	total precipitation		NWP wsdir700*	wind direction at 700 hPa surface	
NWP pwat	precipitable water	x	NWP wsdir850*	wind direction at 850 hPa surface	
NWP gh500	geopotential height of 500 hPa surface	x	doy*	day of year	
NWP gh700	geopotential height of 700 hPa surface	x	hod*	hour of day	



Datasets

- Training dataset - 2,507 entries (differing locations and time periods).
- Test dataset - 1,416 entries (locations and time periods).
- Validation dataset - 1,416 entries (locations and time periods).

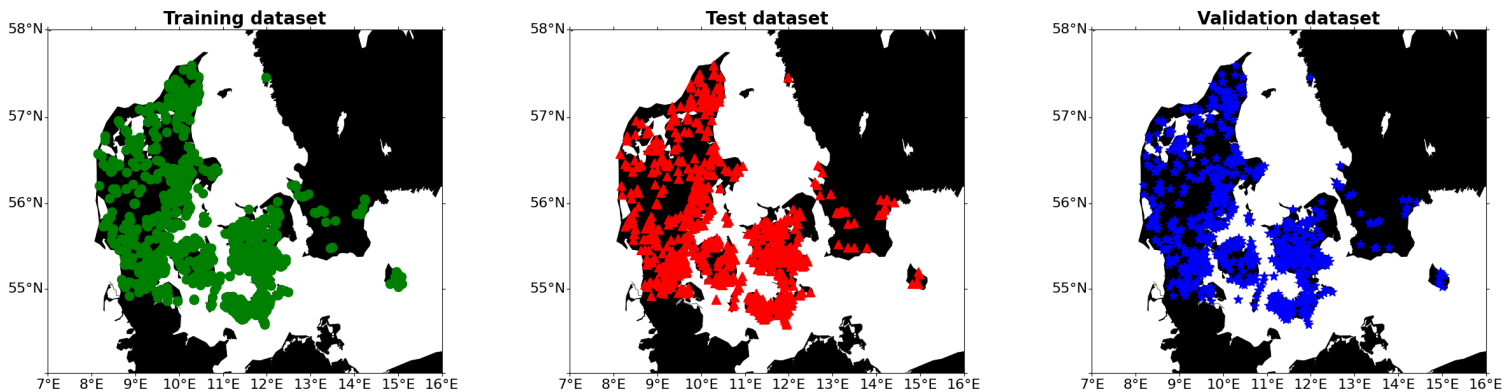

Figure: Spatial distribution of train (green), test (red) and validation (blue) datasets.

Results: Spatial Distribution of Bias

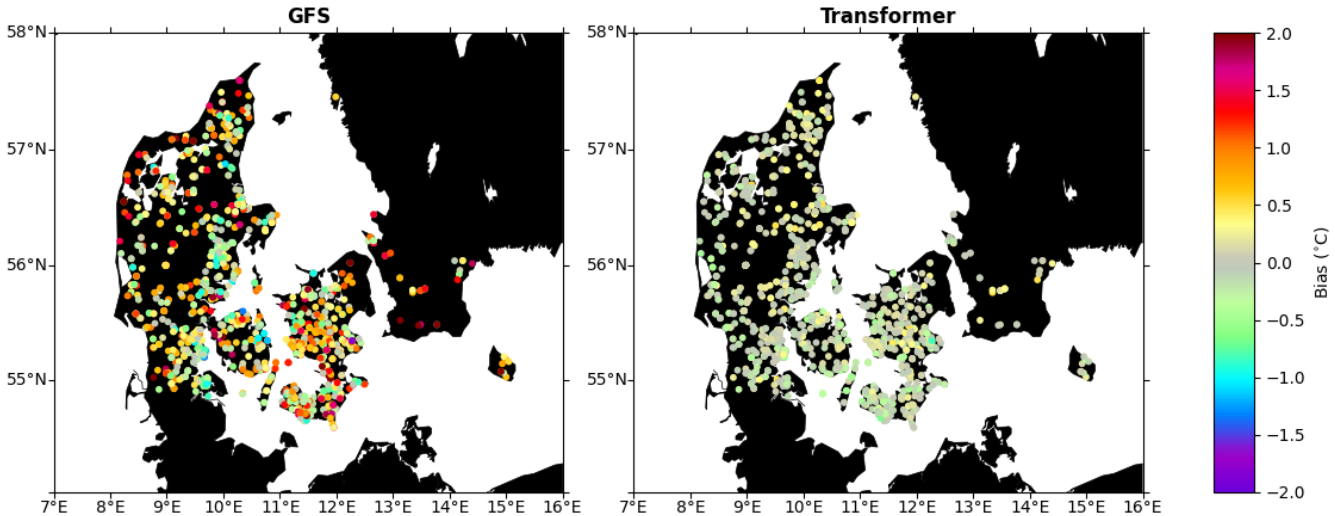

Figure: Spatial distribution of bias. The circles show the location and the colormap indicates the overall bias, calculated for the whole forecast period and all lead times.

- Transformer model geneally produces forecasts with smaller biases.

Results: Spatial Distribution of STD

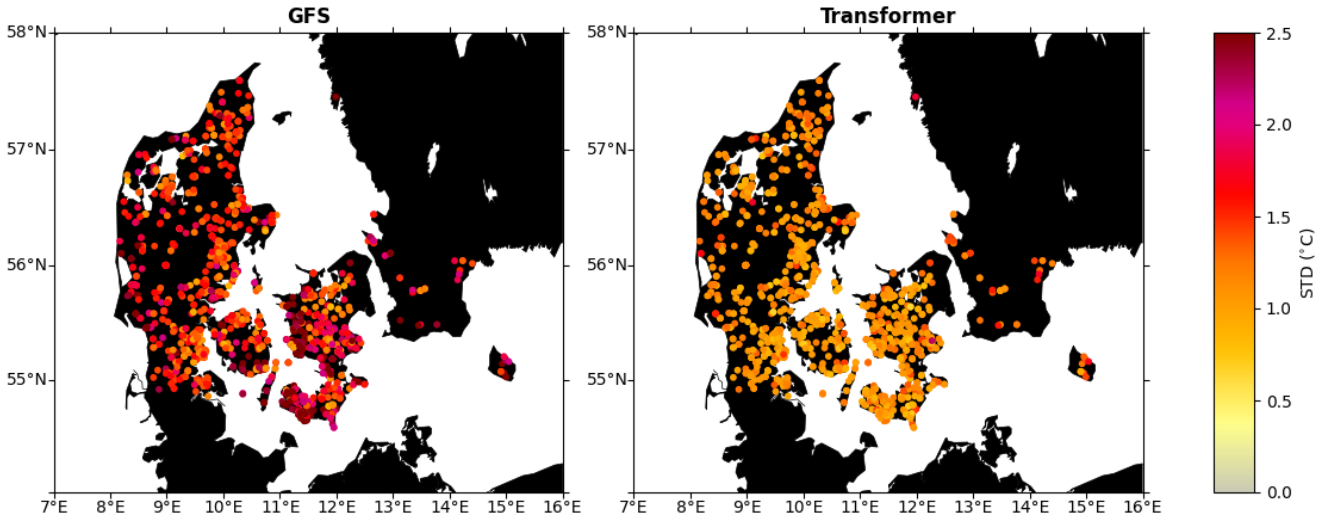

Figure: Spatial distribution of standard deviation. The circles show the location and the colormap indicates the overall standard deviation, calculated for the whole forecast period and all lead times.

- Generally higher standard deviation on the west and south coasts for the raw GFS forecast.
- Overall, the Transformer model produces forecasts with smaller standard deviations.

Results: Summary Statistics

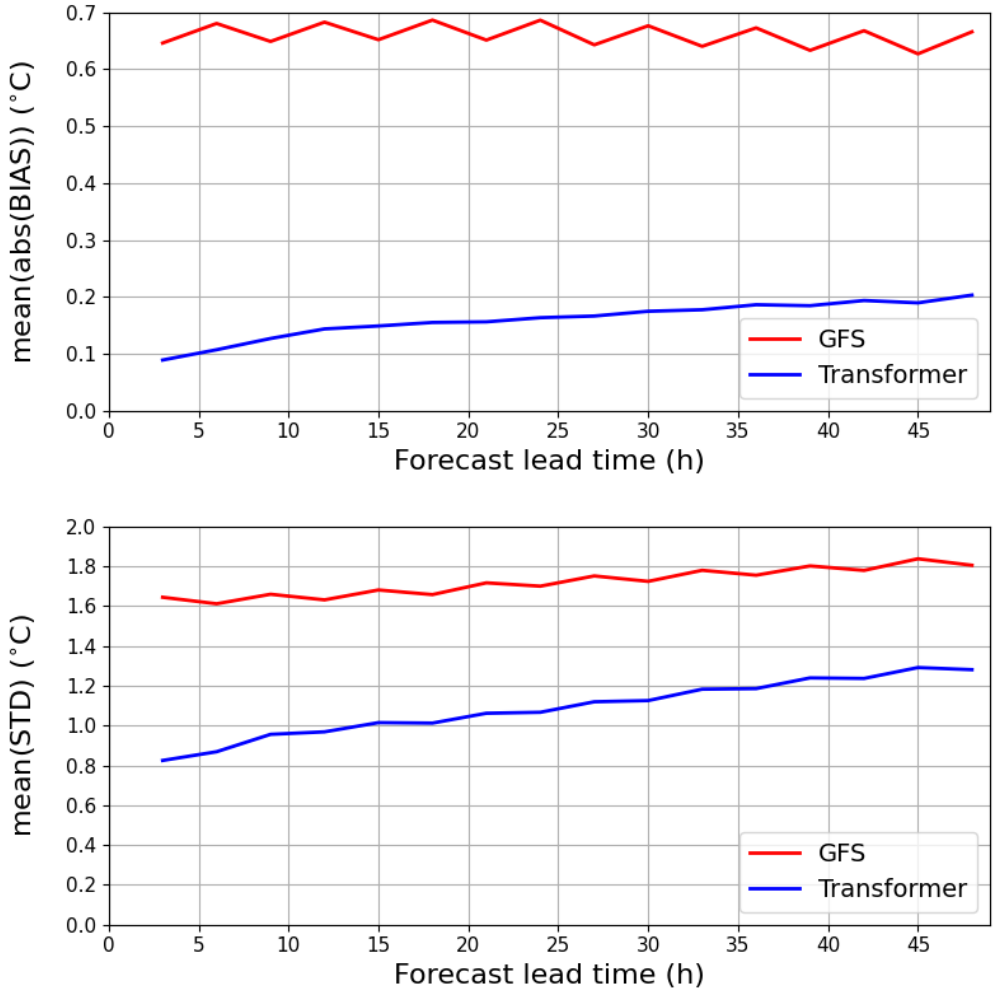

Figure: Summary statistics calculated based on all stations in the validation dataset for bias, in absolute values, (top) and standard deviation (bottom) for the raw GFS forecast data (red) and the Transformer predictions (blue).

Percentage of Improved Stations

Table: Percentage of locations in the validation dataset with improved statistics after post-processing of GFS forecast data, using the Transformer model, for each lead time. Each row show results for verification using the abs(bias), MAE, STD and RMSE skill score, respectively.

Metric	+3h	+6h	+9h	+12h	+15h	+18h	+21h	+24h	+27h	+30h	+33h	+36h	+39h	+42h	+45h
abs(bias)	91	89.1	88.8	87.7	88.1	86.7	88.3	85.4	87.5	83.7	85.5	83.3	85	81.5	83.9
MAE	100	99.9	99.9	99.7	99.9	99.8	99.9	99.9	99.7	99.4	99.6	99.2	99.6	99.2	99.4
STD	99.9	99.2	99.4	98.7	99.4	99.5	99.4	99.2	99.1	98.4	98.9	97.3	97.9	97.2	97.7
RMSE	100	99.7	99.9	99.4	99.9	99.9	99.8	99.8	99.9	99.3	99.7	98.7	99.2	98.6	99.2



Seasonal Bias: Summary Statistics

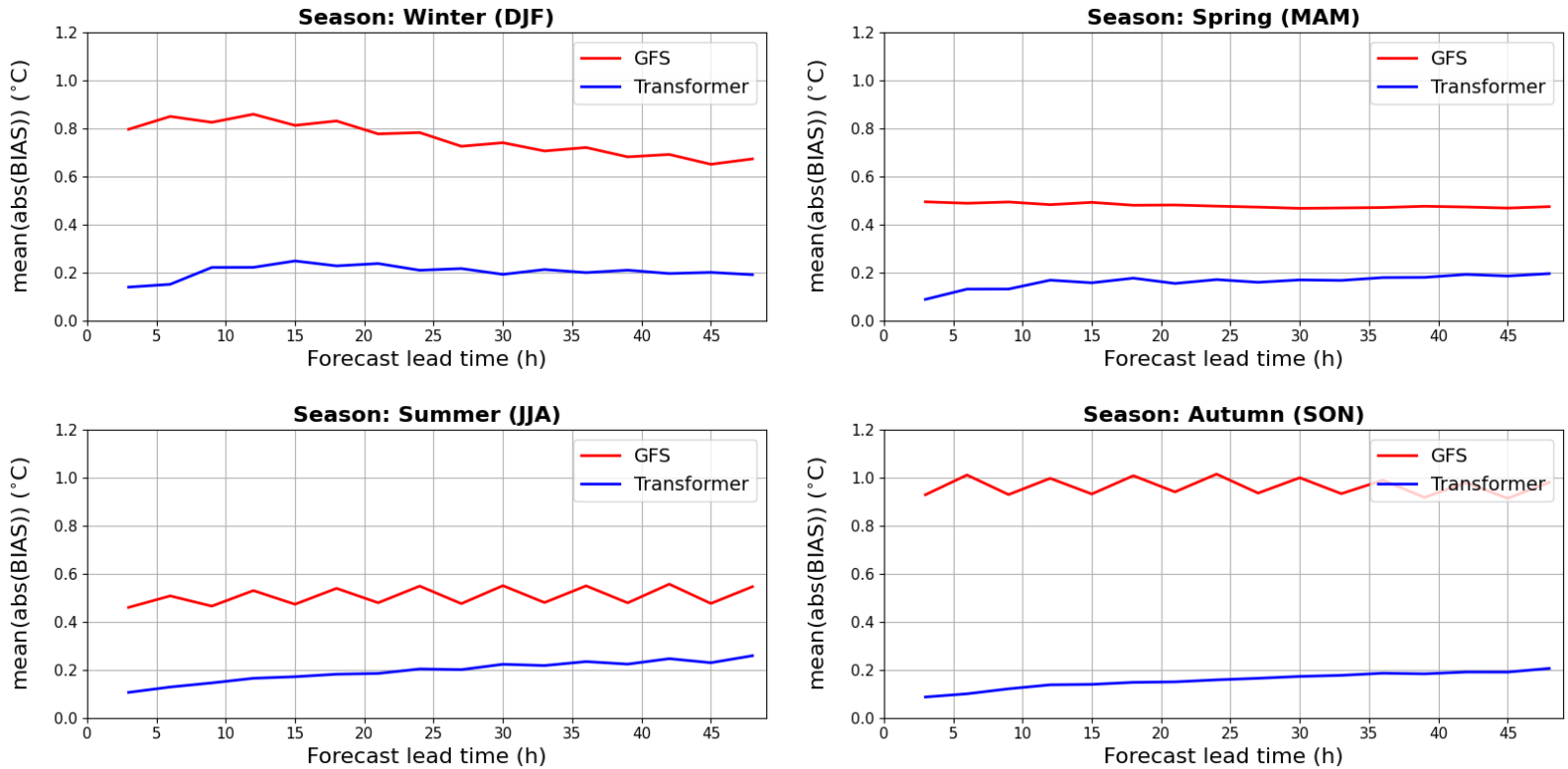

Figure: Seasonal summary statistics calculated based on all stations in the validation dataset for bias, in absolute values, for the raw GFS forecast data (red) and the Transformer predictions (blue). Results are shown for winter months December-February (DJF; top left), spring months March-May (MAM; top right), summer months June-August (JJA; bottom left), and autumn months September-November (SON; bottom right).

Seasonal STD: Summary Statistics

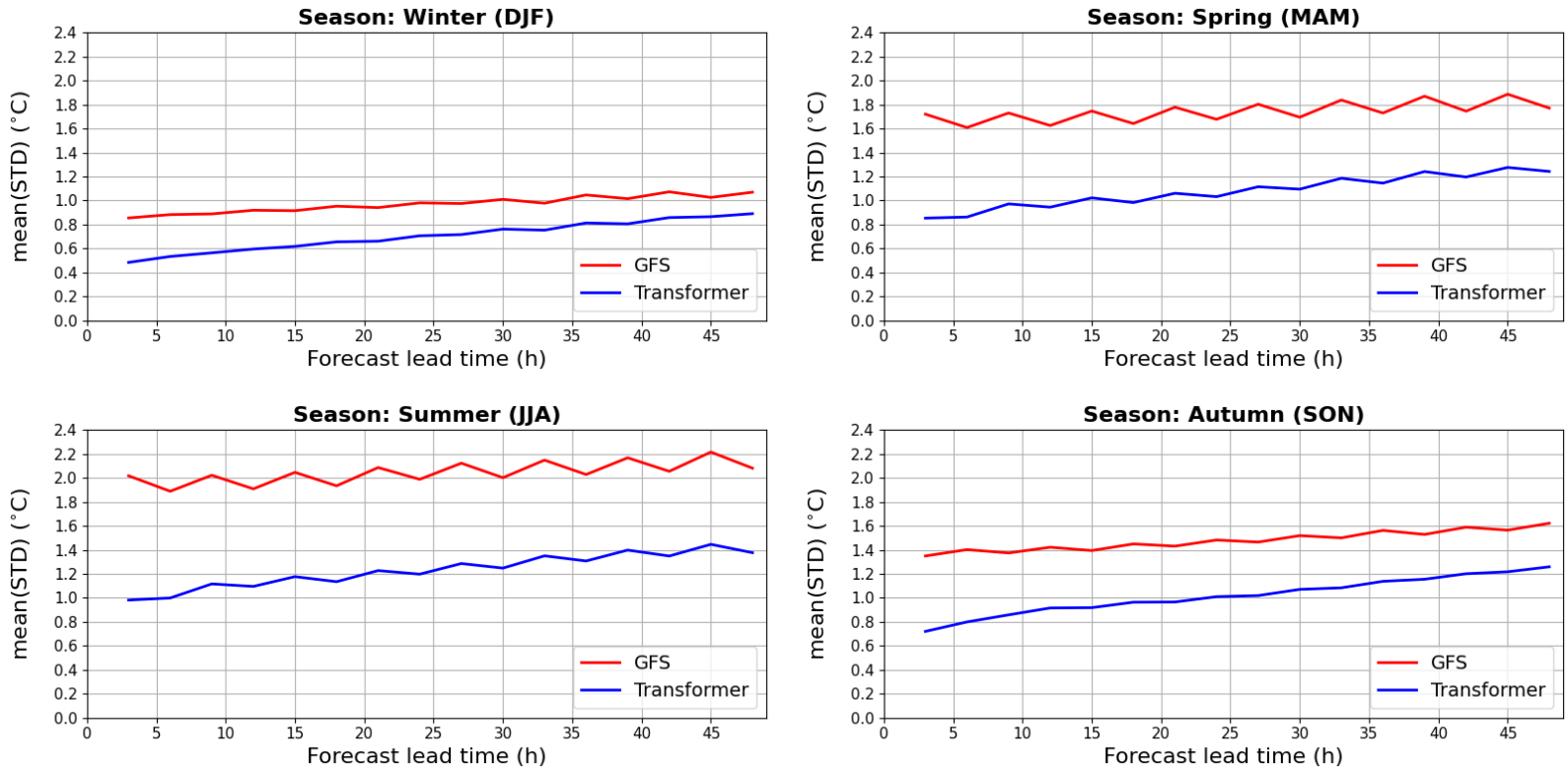

Figure: Seasonal summary statistics calculated based on all stations in the validation dataset for standard deviation for the raw GFS forecast data (red) and the Transformer predictions (blue). Results are shown for winter months December-February (DJF; top left), spring months March-May (MAM; top right), summer months June-August (JJA; bottom left), and autumn months September-November (SON; bottom right).

Validation on DMI Observations

- The Transformer model is applied to observational data from the Danish Meteorological Institute (DMI).
- Number of DMI stations in validation dataset: 50.
- Time period: April - December 2019.

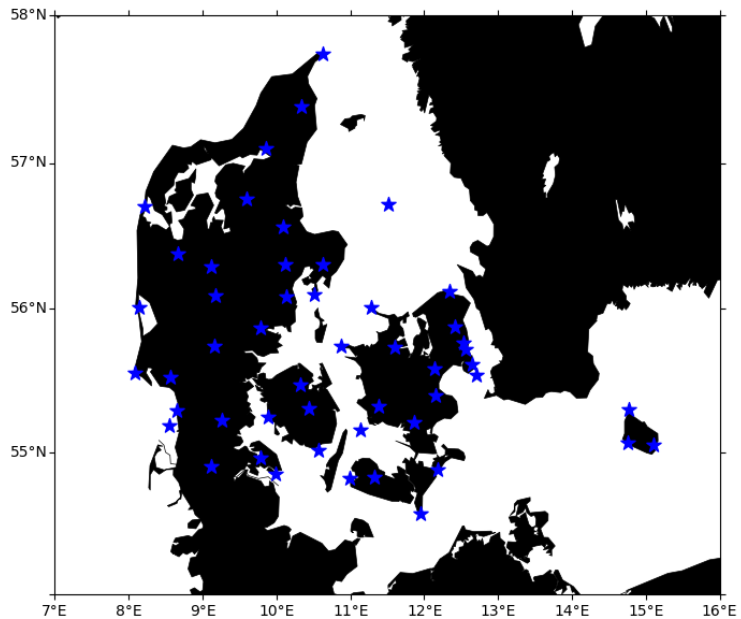

Figure: Location of the 50 weather stations from the Danish Meteorological Institute (DMI) used for validating the Transformer model.

Percentage of Improved Stations

Table: Percentage of DMI stations with improved statistics after post-processing of GFS forecast data for each lead time. Each row show results for verification using the abs(bias), MAE, STD and RMSE skill score, respectively.

Metric	+3h	+6h	+9h	+12h	+15h	+18h	+21h	+24h	+27h	+30h	+33h	+36h	+39h	+42h	+45h
abs(bias)	80	84	80	84	78	80	76	76	70	66	64	68	64	60	58
MAE	94	82	82	82	82	82	82	80	80	76	80	74	80	74	78
STD	90	80	78	78	80	80	80	78	78	76	78	72	78	72	78
RMSE	90	82	78	78	80	80	82	78	78	78	78	74	78	74	78

- More stations benefit from the post-processing for the shorter lead times, and less for the longer lead times.
- Fewer stations seem to benefit from using the Transformer model for post-processing, compared to the results shown previously for the private weather stations.



Spatial Distribution of Bias

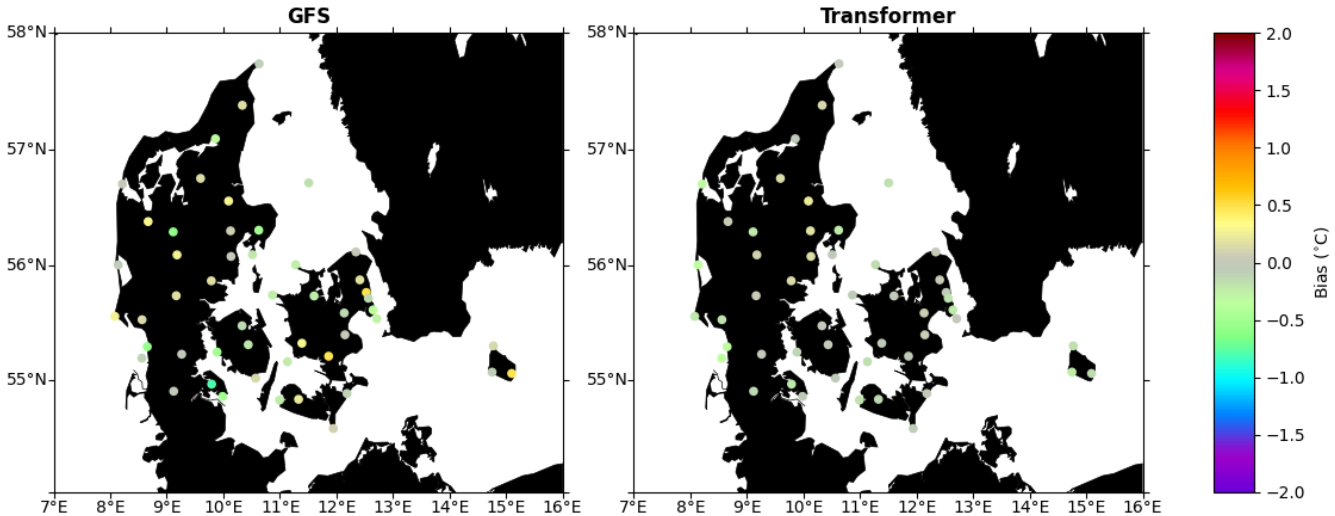

Figure: Spatial distribution of bias for the DMI stations. The circles show the location and the colormap indicates the overall bias, calculated for the whole forecast period and all lead times.

- No clear relationship between bias and spatial distribution.
- Might be because the station density is too low.

Spatial Distribution of STD

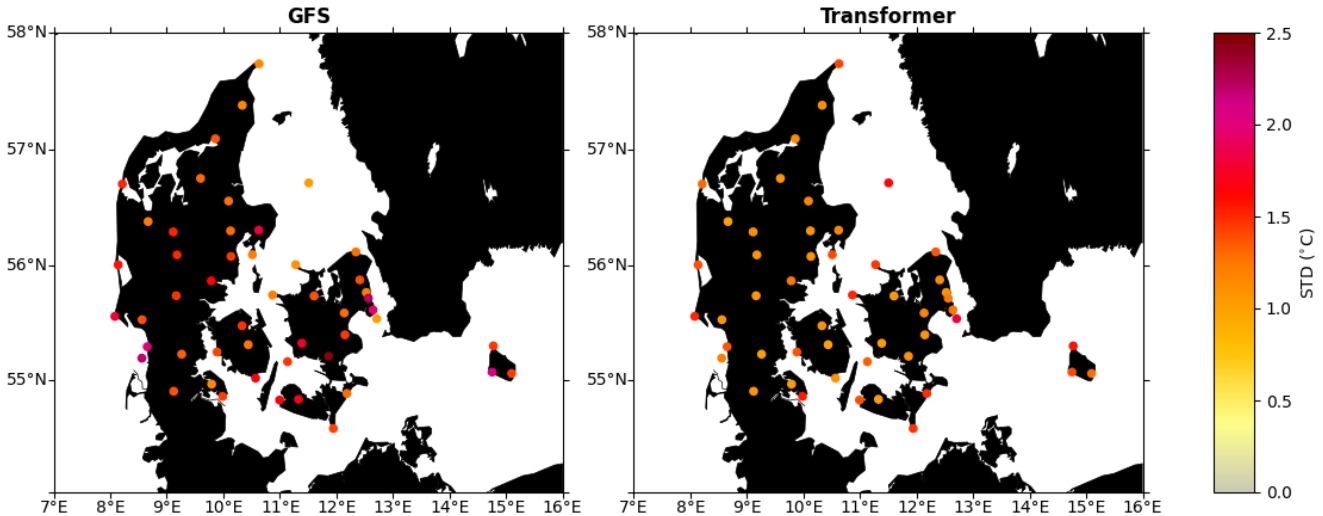

Figure: Spatial distribution of standard deviation for the DMI stations. The circles show the location and the colormap indicates the overall bias, calculated for the whole forecast period and all lead times.

Spatial Distribution of STD

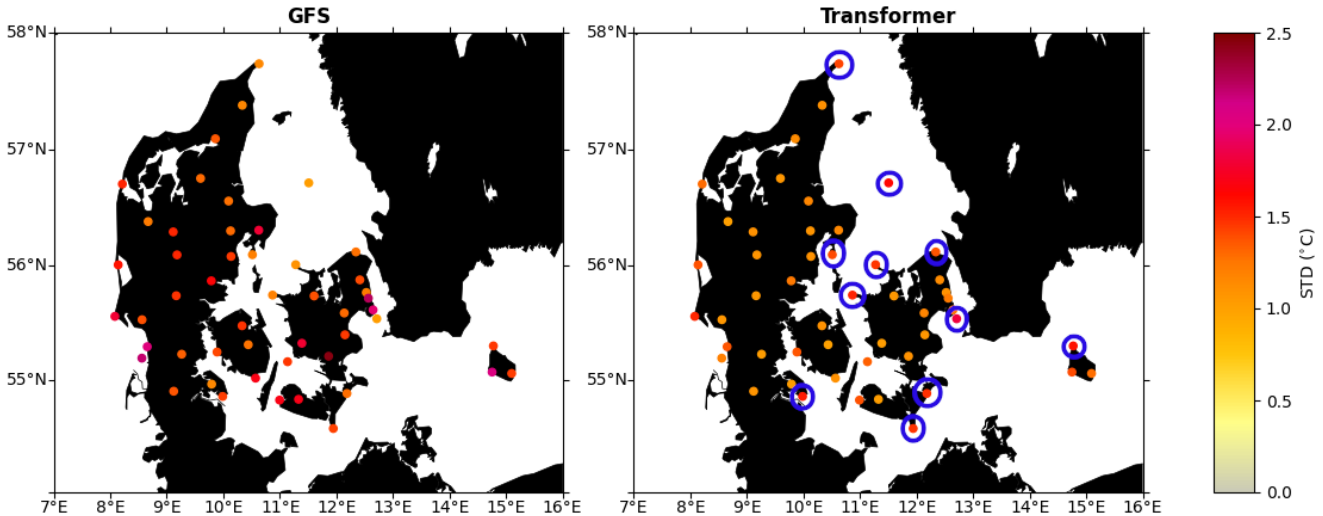

Figure: Spatial distribution of standard deviation for the DMI stations. The circles show the location and the colormap indicates the overall bias, calculated for the whole forecast period and all lead times.

- The Transformer model generally performs worse than the raw GFS forecast for stations close to the coast.
- Focus is on agricultural sites - Not many stations in the training dataset are located close to the coast.

Summary

- Possible spatial dependence of raw GFS forecast - removed when using the Transformer model to post-process the forecast data.
- Transformer model reduces the bias (for more than 80% of all locations) and standard deviation (for more than 90% of the locations) compared to the raw GFS forecast.
- Seasonal statistics show an improvement in bias that is most pronounced for autumn and winter months (September-February).
- For standard deviation, the improvement is most pronounced for spring and summer months (March - August).
- Validation on completely independent observational data from DMI shows that the model has not learned to correct sensor-specific systematic errors, but rather NWP model-specific systematic errors.
- Worse performance for stations located close to the coast - probably not enough of these in the training dataset since focus is on agricultural sites.



References

Petr, J. (1991). Weather and yield. *Developments in crop science*, 20:12–282.

Stigter, C., Sivakumar, M., and Rijks, D. (2000). Agrometeorology in the 21st century: workshop summary and recommendations on needs and perspectives. *Agricultural and Forest Meteorology*, 103(1-2):209–227.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need. *arXiv preprint arXiv:1706.03762*.

WMO (2010). *WMO/CAgM Guide to Agricultural Meteorological Practices (GAMP) WMO no. 134*. Worl Meteorological Organization, Geneva, Switzerland.

