Data Assimilation using ERA5, ASOS, and the U-STN model for Weather Forecasting over the UK

Wenqi Wang
Department of Earth Science and Engineering
Imperial College London
wenqi.wang21@imperial.ac.uk

Jacob Bieker
Open Climate Fix
jacob@openclimatefix.org

César Quilodrán-Casas
Department of Earth Sciences and Engineering
Department of Computing
Imperial College London
c.quilodran@imperial.ac.uk

Abstract

In recent years, the convergence of data-driven machine learning models with Data Assimilation (DA) offers a promising avenue for enhancing weather forecasting. This study delves into this emerging trend, presenting our methodologies and outcomes. We harnessed the UK's local ERA5 850 hPa temperature data and refined the U-STN12 global weather forecasting model, tailoring its predictions to the UK's climate nuances. From the ASOS network, we sourced T2m data, representing ground observations across the UK. We employed the advanced kriging method with a polynomial drift term for consistent spatial resolution. Furthermore, Gaussian noise was superimposed on the ERA5 T850 data, setting the stage for ensuing multi-time step virtual observations. Probing into the assimilation impacts, the ASOS T2m data was integrated with the ERA5 T850 dataset. Our insights reveal that while global forecast models can adapt to specific regions, incorporating atmospheric data in DA significantly bolsters model accuracy. Conversely, the direct assimilation of surface temperature data tends to mitigate this enhancement, tempering the model's predictive prowess.

1 Introduction

Numerical Weather Prediction (NWP) has progressed over the decades, establishing itself as a cornerstone in weather forecasting [19]. Its efficacy, demonstrated across diverse environmental scenarios, provides probabilistic forecasts [13]. Among various NWP models, the ECMWF IFS (Integrated Forecasting System) is consistently recognised for its most accurate representation of atmospheric conditions in designated geographical areas [16]. While it maintains a lead over machine learning (ML) models, the rise of ML-driven models — with their refined designs and increasing resolutions — cannot be overlooked [15]. These models leverage historical atmospheric data, aspiring to match or surpass IFS predictions while conserving computational resources and boosting processing speeds. This includes resolutions from 5.625 degree in the WeatherBench Convolution Neural Network(CNN) series [14] to 1.0 degree in Keisler's Graph Neural Network (GNN) model [8] and 0.25 degree in GraphCast [10]. Even though ML model developers do not anticipate replacing NWP entirely, the strides made in ML suggest its capacity to refine existing forecasting approaches [10]. In recent years, the integration of data-driven ML models and Data Assimilation (DA) techniques has attracted substantial attention due to its potential to enhance

model performance [3]. DA serves as an intermediary, merging observational data with model outputs to hone initial backgrounds, thereby fostering improved forecast results. In the domain of atmospheric research, 850 hPa temperature (T850) and 500 hPa geopotential height (Z500) are fundamental datasets. Their significance stems from their representation of large-scale circulation in the troposphere, which inherently affects near-surface weather conditions and extremities [4]. The objective of weather forecasting is to predict the future state of the atmosphere for a designated time [14]. However, this forecasting is confined to the atmosphere's prediction horizon, estimated to be approximately two weeks [17].

In this study, we not only used data from ERA5 and the ASOS network but also incorporated global meteorological data from the U-STN model. Machine learning models excel at handling heterogeneous data from multiple sources. They can automatically identify correlations between different datasets and effectively integrate this information, which is crucial for improving data assimilation (DA) outcomes. Weather systems are highly complex and nonlinear. Machine learning models, especially deep learning models, are adept at capturing this complexity, providing a finer understanding compared to traditional numerical weather prediction models.One key advantage of machine learning models is their ability to adapt and respond to changes in new data. This means that as we collect more meteorological data and real-time observations, the models can continuously self-optimize and adjust, thereby enhancing long-term prediction accuracy.

In this study, we adapted a global weather forecasting model [4] employing a U-Net enhanced with a deep spatial transformer (U-STN) and focused on a single data variable. This model was specifically retrained for the UK region. To explore the optimisation effects of DA on predictive outcomes, we incorporated the sigma-point Ensemble Kalman Filter (SPEnKF) algorithm [1], particularly leveraging surface data from ground observation stations. A primary model was trained using distinct 12-hour time interval on ERA5 dataset: U-STN12. In our study, four different data typologies were utilized for DA: (1) Employing $\sigma_{\text{obs}} = \sigma_z$ and $\sigma_{\text{obs}} = 0.5\sigma_z$, the T850 data was augmented with Gaussian noise at two levels, simulating observational data with noise coefficients of 1 and 0.5, respectively. This methodology is aligned with standard practices in DA and facilitates the examination of DA's impact [4]; (2) Utilizing $\sigma_{\text{obs}} = \sigma_z$ and $\sigma_{\text{obs}} = 0.5\sigma_z$, virtual observations were generated on T850 via U-STN12 model, offering a closer approximation to actual observations than Gaussian noise; (3) Observational data were directly simulated using data from ASOS observation stations; (4)The ERA5 T2m data served as a surrogate for simulated observational data. Theoretically, its data distribution mirrors that of the ASOS data, thereby mitigating potential alterations in data distribution characteristics during the processing of ASOS data. Our primary objective is to find the influence of incorporating ground observational data in DA processes for atmospheric dataset.

2 Methods

2.1 Datasets

2.1.1 ERA5

The ECMWF's ERA5 reanalysis dataset integrates NWP forecast models with contemporaneous observations using 4D-Var DA [6]. We collected T850 hourly data from 1940 to present [2] and T2m hourly data from 1940 to present, encompassing the timeframe from 1979 to 2022. The 2022 annual T850 temperature pattern is shown in 1 (b). For more details of the data and the processing procedures, see Appendix 4

2.1.2 ASOS Dataset

Recognising the potential correlations with surface observational data, we incorporated 2-meter temperature (T2m) data from ASOS via the Iowa Environmental Mesonet [7], which delivers hourly surface observations spanning the period from 1979 to 2022 and encapsulating readings from 112 observation stations within the UK. This system predominantly supports aviation weather forecasting and further weather forecasting research [11]. The 2022 annual T2m temperature and distribution of observation stations are depicted in 1 (a). Our aim with the ASOS data was to align its structure with the ERA5 data, as visualised in Figure 1.
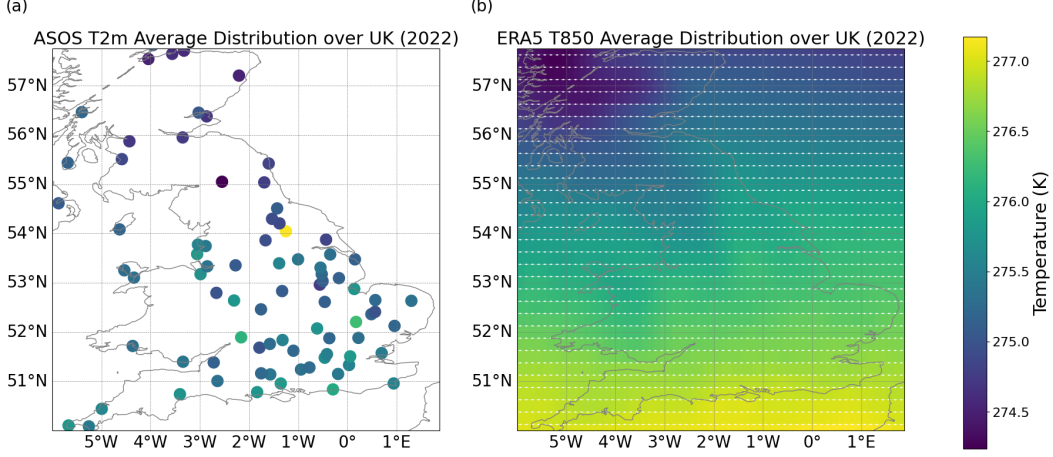
Figure 1: (a) Geographical distribution of ASOS surface observation stations and annual average T2m values between latitudes 50° to 57.75° and longitudes -6° to 1.875°(2022). (b) Annual average T850 temperature pattern from ERA5 within the same latitudinal and longitudinal bounds (2022).

2.2 Model

We employed a U-Net-based machine learning model enhanced with a deep spatial transformer [4]. Using the Adam optimiser [9] and a mean squared error (MSE) loss function, the model's parameters were aligned with those specified in [4]. Although our primary goal is the exploration of DA from ground observation stations, promising results may prompt further hyperparameter optimisation specific to the UK or other regions. Our prediction approach is autoregressive: a prediction at time $T(t + \Delta t)$ becomes the input for $T(t + 2\Delta t)$. Depending on the forecast's intended duration, Δt might vary. In this work, we set Δt at 12h for model named U-STN12. The specific model code and implementation are available on GitHub: `https://anonymous.4open.science/r/ TacklingClimateChangeAIforNeurIPS2023_2_02DF/src/models`.

2.3 Data assimilation

We adopted the SPEnKF approach [18]. The SPEnKF, unlike the EnKF, employs an unscented transformation that utilizes sigma points, which are a predefined set of points used to estimate the mean and variance of a nonlinear function. By processing these sigma points through the nonlinear function, it becomes feasible to gauge the output's mean and covariance without relying on perturbation-based linearisation [1]. The specific assimilation code and implementation are available on GitHub: `https://anonymous.4open.science/r/ TacklingClimateChangeAIforNeurIPS2023_2_02DF/src/assimilation`.

3 Results

We employed U-STN12 as our ML model and SPEnKF as the DA algorithm. The prediction's initial value is derived from a noisy observation. Every 24 hours, we assimilate this noisy observation to refine the prediction. By introducing various noise levels, Gaussian noise $N(0, \sigma_{\text{obs}})$ is superimposed onto the ERA5 T850 data. We then analysed the root mean square error (RMSE) between the predicted mean and the noisy data in the T850 domain over the initial 120 hours across 50 random conditions. The result is shown in Figure 2. Our attention is primarily drawn to the similarities and disparities in trends before and after the DA juncture. Given that our model commences training from the 12 lead time, the RMSE for the initial 12 hours exhibits an upward trend due to lack of training. Between the 12 and 48 hours lead time, there is a consistent reduction in RMSE, underscoring the model's effective predictive performance. Notably, a marked decrease in RMSE is observed at the 24 hours lead time point, corresponding with the introduction of DA. This decrease is followed by a gradual increase in RMSE over time, likely a consequence of accumulating losses. A scenario where $\sigma_{obs} = \sigma_T$ results in a larger RMSE, suggesting that increased noise levels detrimentally
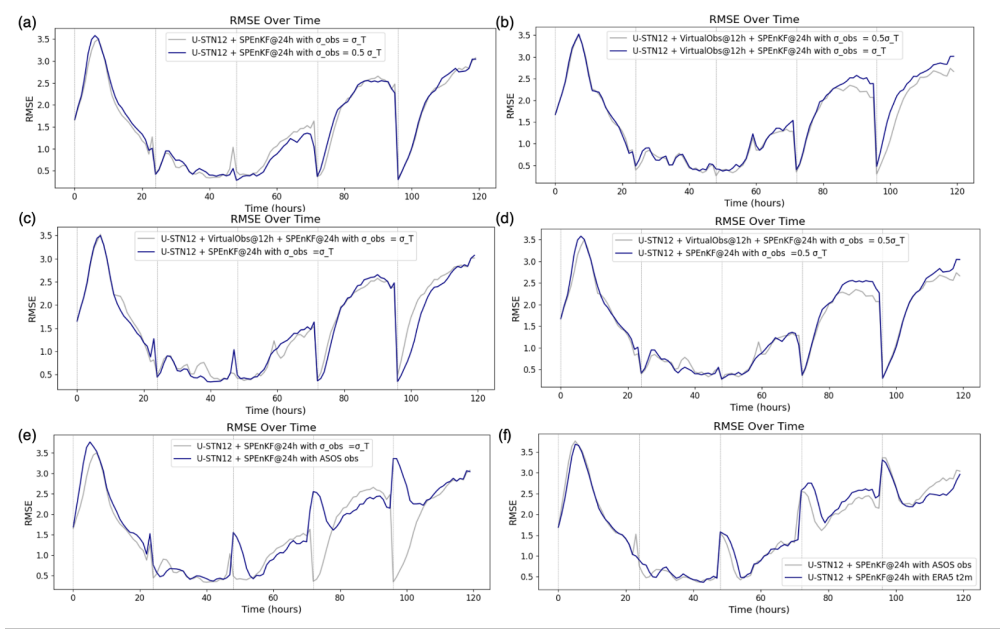
Figure 2: Root Mean Square Error (RMSE) Comparison for the U-STN12-SPEnKF Model Across Different DA Sources.

impact prediction accuracy. Nonetheless, the RMSE trends remain consistent between $\sigma_{obs} = \sigma_T$ and $\sigma_{obs} = 0.5\sigma_T$, attesting to the model's robustness.

As illustrated in Figure2 (b), We experimented with using the model's training data at a time step $t = 12$ hours as simulated observations, as an alternative to Gaussian noise simulated observations over a 24-hour DA span. Both $\sigma_{obs} = 0.5\sigma_T$ and $\sigma_{obs} = \sigma_T$ exhibit consistent performance trends. Notably, with every 12-hour increment when virtual observations are introduced, there is a discernible reduction in RMSE.

In Figure2 (c) and (d), when contrasting this with the RMSE trend without the addition of virtual observations, it is evident that introducing these virtual observations results in a more gradual ascent in the model's RMSE. When juxtaposed with the RMSE trajectory without virtual observations, the early stages, characterised by high model prediction accuracy, did not manifest any pronounced benefits from the virtual measurements. However, past the 48-hour mark, as the model's predictive efficacy weakens, the RMSE resulting from the inclusion of virtual observations was notably lower than that from Gaussian noise simulation.

As illustrated in Figure 2 (e), the integration of `T2m` data from the `ASOS` ground observation station into the DA process impairs its efficacy. A significant RMSE spike at the assimilation juncture manifestly evidences this degradation. As Figure 2 reveals, the original `ASOS` ground observation stations exhibit a sparse and non-uniform distribution. When interpolated to match the `ERA5` dataset resolution, inevitable errors emerge. To validate that discrepancies between ground and atmospheric data diminish DA performance, we similarly integrated `ERA5 T2m` data into the assimilation process. As shown in Figure2 (f), this integration mirrors the previous trend, with RMSE elevating at assimilation points, confirming the detrimental impact on DA. However, it is salient that the RMSE when utilising `ASOS` ground data is less than that from the `ERA5 T2m` data. This observation paves the way for further optimisation of the interpolation technique and exploration of ground-atmospheric observation correlations.

4 Conclusion

Adapting a global ML model to regional predictions is viable. However, regional-specific retuning of hyperparameters or model architecture is essential. The SPEnKF assimilation method, when applied with model-based noisy or multi-time step virtual observations, augments prediction accuracy. Given the inherent discrepancies and spatial irregularities of ASOS observational data, especially within the UK context, the direct assimilation of surface observation data with atmospheric T850 appears inadvisable. Based on our findings, we propose three refinements: (1) Data Sources and Interpolation: Acquire denser datasets and implement advanced interpolation and preprocessing techniques to better align T2m values with T850; (2) Hyperparameter Tuning: Optimise hyperparameters in our models, including evaluating the number of autoregressive time steps, to enhance performance; and (3) Incorporation of Multi-Layer Pressure Levels: Integrate ground observation data with models employing multi-layer pressure levels, such as Graph Neural Networks (GNNs), to reduce interpolation errors and improve prediction accuracy. Through these enhancements, we aim to uncover patterns in using ground observation data for DA. Beyond the structural and data-based refinements previously mentioned, incorporating regional domain knowledge can dramatically improve the performance and adaptability. Establishing a feedback system where the predictions of the model are continuously compared against actual observations. Any discrepancies can be used as learning points for the model, enabling it to self-correct and adapt to the specificities of the region over time.

References

[1] Sigma-Point Kalman Filter Data Assimilation Methods for Strongly Nonlinear Systems. *Journal of the Atmospheric Sciences*, 66(2), 2009.

[2] B. Bell, H. Hersbach, A. Simmons, P. Berrisford, P. Dahlgren, A. Horányi, J. Muñoz-Sabater, J. Nicolas, R. Radu, D. Schepers, C. Soci, S. Villaume, J.-R. Bidlot, L. Haimberger, J. Woollen, C. Buontempo, and J.-N. Thépaut. The ERA5 global reanalysis: Preliminary extension to 1950. *Quarterly Journal of the Royal Meteorological Society*, 147(741):4186–4227, 2021.

[3] Mark Buehner, Ron McTaggart-Cowan, and Sylvain Heilliette. An Ensemble Kalman Filter for Numerical Weather Prediction Based on Variational Data Assimilation: VarEnKF. *Monthly Weather Review*, 145(2):617–635, February 2017.

[4] Ashesh Chattopadhyay, Mustafa Mustafa, Pedram Hassanzadeh, Eviatar Bach, and Karthik Kashinath. Towards physics-inspired data-driven weather forecasting: integrating data assimilation with a deep spatial-transformer-based U-NET in a case study with ERA5. *Geoscientific Model Development*, 15(5):2221–2237, March 2022.

[5] J. Zhuang et al. pangeo-data/xesmf: v0.8, Sep. 01 2023.

[6] Hersbach H., Bell, B., P. Berrisford, G. Biavati, A. Horányi, and J. Muñoz Sabater. ERA5 hourly data on pressure levels from 1940 to present. *Copernicus Climate Change Service (C3S) Climate Data Store (CDS)*.

[7] Daryl Herzmann. Iem:: Download asos/awos/metar data. `https://mesonet.agron.iastate.edu/request/download.phtml`. Accessed: Sep. 03, 2023.

[8] Ryan Keisler. Forecasting Global Weather with Graph Neural Networks, February 2022. arXiv:2202.07575 [physics].

[9] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. *3rd International Conference for Learning Representations*, Jan 2015.

[10] Remi Lam et al. GraphCast: Learning skillful medium-range global weather forecasting, December 2022.

[11] Richie Lengel. Everything You Need to Know about AWOS and ASOS, June 2018.

[12] S. Müller and L. Schüler. Geostat-framework/gstools: v1.5.0 "nifty neon". Zenodo, Jun. 15 2023.

[13] Tim Palmer. The ECMWF Ensemble Prediction System: Looking Back (more than) 25 Years and Projecting Forward 25 Years. *Quarterly Journal of the Royal Meteorological Society*, 145(S1):12–24, 2019.

[14] Stephan Rasp, Peter D. Dueben, Sebastian Scher, Jonathan A. Weyn, Soukayna Mouatadid, and Nils Thuerey. WeatherBench: A benchmark dataset for data-driven weather forecasting. *Journal of Advances in Modeling Earth Systems*, 12(11), November 2020.

[15] Stephan Rasp, Stephan Hoyer, Alexander Merose, Ian Langmore, Peter Battaglia, Tyler Russel, Alvaro Sanchez-Gonzalez, Vivian Yang, Rob Carver, Shreya Agrawal, et al. Weatherbench 2: A benchmark for the next generation of data-driven global weather models. *arXiv preprint arXiv:2308.15560*, 2023.

[16] Stephan Rasp and Nils Thuerey. Data-driven medium-range weather prediction with a Resnet pretrained on climate simulations: A new model for WeatherBench. *Journal of Advances in Modeling Earth Systems*, 13(2):e2020MS002405, February 2021.

[17] S. Scher and G. Messori. Generalization properties of feed-forward neural networks trained on lorenz systems. *Nonlinear Processes in Geophysics*, 26(4):381–399, 2019.

[18] Rudolph F. van der Merwe, Eric A. Wan, and Simon J. Julier. Sigma-Point Kalman Filters for Nonlinear Estimation and Sensor-Fusion–Applications to Integrated Navigation. In *Proceedings of the AIAA Guidance, Navigation & Control Conference*, volume 3, Aug 2004.

[19] Jonathan Weyn, Dale Durran, and Rich Caruana. Can machines learn to predict weather? using deep learning to predict gridded 500-hpa geopotential height from historical weather data. *Journal of Advances in Modeling Earth Systems*, 11, 08 2019.

Appendix

Datasets

ERA5

We utilised the ERA5 dataset spanning the years 1979 to 2020 for training, which comprises approximately 367,920 samples. The 2021 dataset, containing around 8,760 samples, was designated for validation, while the 2022 dataset, with a similar count of roughly 8,760 samples, was allocated for testing. Both ERA5 datasets encompass a geographical expanse with latitude ranging from 49°to 58°and longitude from -8°to +2°.

The inception year of 1979 was strategically chosen, reflecting a pronounced enhancement in data availability and quality post this year. This period correlates with significant advancements in ancillary research domains, such as climate change studies [6].

To ensure integrity in the preprocessing of the downloaded raw data, efforts were directed to retain its intrinsic properties. This strategy aids subsequent comparisons with prevailing global weather forecast models [4] and the WeatherBench [14]and WeatherBench2 [15]. The preprocessing steps involved:

(1) Cropping: The focus area was delimited to latitudes between 57.75°and 50.0°, and longitudes between -6° and 1.875°. (2) Remeshing: The tool xsemf [5]was employed to realize a grid configuration of 32x64. (3) Handling Null Values: In instances of null data points, a 3-hour sliding window was applied, imputing the average value. Consequently, the annual datasets were re-gridded, conforming to the dimensions (time, latitude, longitude) of 8760×32×64, with a slight adjustment to 8784×32×64 for leap years.

ASOS Dataset

We adopted the ASOS dataset by the following preprocessing steps: (1) Hourly Normalisation: Recognizing the varied reporting schedules of ASOS stations, we standardized the hourly reporting times. Specifically: (a)For timestamps within the first 30 minutes of an hour (e.g., 01:20:00), if a full-hour data point precedes it, we omit the data. If not, it's shifted to the preceding full hour. (b)For timestamps beyond the 30-minute mark (e.g., 01:50:00), if a full-hour data point follows, the data is omitted; otherwise, it's shifted to the subsequent full hour.(2)Region Filtering: We retained data within the bounds of latitude [50, 58] and longitude [-6, +2]. (3) Kriging Interpolation: Given the uneven station density, we used kriging interpolation via gstools [12] to ensure spatial consistency. Polynomial drift terms were incorporated to capture trends along latitude and longitude, and their interplay. The formula for drift is:

$$\text{drift} = F(1, \text{lat}, \text{lon}, \text{lat}^2, \text{lon}^2, \text{lat} \times \text{lon}) \tag{1}$$

where 1 is the baseline constant term, the lat term represents the linear variation with latitude, the lon term indicates the linear variation with longitude, lat^2 denotes the quadratic trend with latitude, lon^2 signifies the quadratic trend with longitude, and $\text{lat} \times \text{lon}$ term captures the interaction of latitude and longitude.

In meteorology, spatial data often exhibit non-stationarity, meaning that the statistical characteristics of the data (such as mean and variance) vary across space. For example, the climate in the UK is influenced by various factors, including topography and proximity to the ocean, which can lead to different trends in variables like temperature and rainfall across space. The Kriging method with polynomial drift effectively handles this non-stationarity by introducing a polynomial trend term to capture these spatial variations.

The polynomial drift term allows the model to consider trends in variables as they vary geographically. This is particularly important when dealing with variables like temperature, which may change with latitude, altitude, and other geographic factors. By modeling these trends, we can more accurately estimate meteorological conditions at unobserved points.