Meteorological Applications
The Royal Meteorological Society Journal for applied meteorologists, forecasters and users of meteorological services

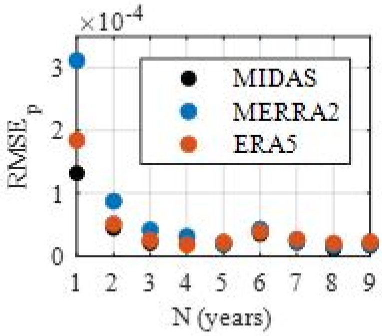
RMetS
Royal Meteorological Society

Effect of sample timespan on wind speed and insolation modelling with MERRA2 and ERA5 reanalyses evaluated against UK MIDAS observations

Journal:	*Meteorological Applications*
Manuscript ID	Draft
Wiley - Manuscript type:	Research Article
Keywords:	Evaluation < Modelling, Statistical Models < Modelling, Resolution < Verification, Meteorological applications < Miscellaneous
Manuscript keywords:	

Bartle, R.S., Kerkhoff, M.A., Carter, Z., Macleod, A.*.

Effect of sample timespan on wind speed and insolation modelling with MERRA2 and ERA5 reanalyses evaluated against UK MIDAS observations



For MIDAS, MERRA2, and ERA5 windspeed data, and for full data probability density functions across locations, increasing the number of years *N* used to calculate the parameter generally decreased mean and root-mean-squared errors compared to the parameter calculated by the full timespan of data until around the fifth year; thereafter, the error magnitude was small and no significant relationship between *N* and error was observed.

Effect of sample timespan on wind speed and insolation modelling with MERRA2 and ERA5 reanalyses evaluated against UK MIDAS observations

Roy S. Bartle [1], Maria A. Kerkhoff [2], Zak Carter [1], Alasdair Macleod [1*]

[1] Department of Engineering, University of the Highlands and Islands – Outer Hebrides, Isle of Lewis, HS2 0XR, UK.

[2] Chilliwack, British Columbia, Canada.

* Corresponding author: Dr Alasdair Macleod (alasdair.macleod@uhi.ac.uk).

Keywords:

MERRA2, ERA5, reanalysis, wind, solar, energy systems, timespan, variability

Abstract

Historical windspeed and solar energy data is required for many activities in the physical sciences including energy systems modelling. Reanalysis models provide such data where physical observations are unreliable or inextant, but their accuracy needs to be understood within geographical bounds. By evaluating the MERRA2 and ERA5 reanalysis models against MIDAS observations, we found that both adequately modelled UK windspeed and insolation, with ERA5 generally returning better accuracy. Additionally, we investigated the effect of sample timespan on probability density function (pdf) accuracy of windspeed and insolation using Weibull fit of hourly data, and P50 and P90 aggregated data methodologies. For the P90 parameters, there was a consistent negative correlation between the number of years modelled and the error between the parameter calculated using the modelled years and the same parameter using the full ten years' timespan of data. For the other methodologies, error metrics decreased until around 5 years and did not reduce significantly thereafter.

1 Introduction

Historical meteorological data is required to observe climatic trends [1], understand historical energy use [2], and forecast future energy consumption [3]. Renewable energy engineers need historical wind and solar data to design and optimise renewable energy systems with energy storage whilst minimising system cost and environmental impact [4].

Much work on temporal aspects of wind and solar modelling is related to the selection of timespan resolution [5]–[7]. Pryor et al. summarise previous work on interannual variability (IAV) and, by simulation of 101,761 grid cell locations of 12 x 12 km in the central and eastern USA, show that the modelled IAV is lower than the IAV typically assumed in energy modelling for large windfarms [8]. Mora et al. show the effect of mean windspeed uncertainty on wind farm financing [9]. Relatively little work has been done on the effect that timespan selection has on the accuracy of the resulting probability density function (pdf) compared to the pdf generated from a long timespan dataset. Kubik et al. noted the computational expense of evaluating the long-term wind variability for a region with dynamical downscaling [10]. Lee et al. argue that the use of hourly data does not differentiate amongst geographic regions with precision, whilst annual mean data typically contains too few datapoints, and so monthly data is preferred [11]. They then considered the effects of varying the sample timespan on monthly windspeed data and found that 10 ± 3 years of monthly data was required to achieve 90% confidence [12]. Chen et al. [13] investigated probabilistic modelling of windspeed from measurements and MEPS reanalysis at five locations in the Norwegian Arctic; they concluded that the Weibull and Nakagami distributions best modelled the pdfs of the sites, and both they and Safari and Gasore [14] hypothesised that multiple years' data would be beneficial for improving the pdf accuracy and predicting extreme windspeed events. Huang et al. observed that, for any time scale of windspeed data, the pdf in any consecutive sub-period is similar to the pdf for the total time period [7].

For energy systems modelling activities, a fitted pdf can be used directly, or additional statistical parameters can be obtained for more complex analysis, typically if the underlying dataset is limited or incomplete. The P50 and P90 methodologies are examples of the latter approach in which the mean is obtained and probability analysis performed on it; the P50 value has a 50% probability of being exceeded (i.e. the 50th percentile value on the pdf), whilst the P90 has a 90% probability of being exceeded (i.e. the 10th percentile value in the pdf) and is a more conservative measure for energy systems modelling. P50 and P90 values can be calculated for the input energy source (e.g. [9]) or for the energy generated by the system (e.g. [8]). By analysing 62 year windspeed records of 60 Canadian weather stations, and randomly permuting the annual values from each location, Bodini et al. observed that no improvement in annual energy P50 error was gained by using more than 4 to 5 years of data, whilst using datasets exceeding 20 years actually increased P50 error [15].

At many locations, historical wind and insolation observations are not available; therefore, alternative methods of obtaining such data must be used. Reanalysis models provide one such method. Reanalysis models combine real-world observations with theoretical analytical models, often also used for forecasting, to reproduce historical meteorological data [16]. They provide an understanding of meteorological parameters at locations for which observational data is unavailable or unreliable. Along with many other physical science applications, within the energy sector reanalysis has been used for large- [17] and small- [18] scale energy systems modelling, and modelling the relationship between climatic variation and energy [19]. However, there is a need to ensure that reanalysis model accuracy is better understood and described [20]. The MERRA2 and ERA5 reanalyses are two recent and popular models with global range. The Modern-Era Retrospective Analysis for Research and Applications version 2 (MERRA2) is developed by the Global Modelling and Assimilation Office (GMAO), which supports the US's NASA activities [21]. ERA5 is developed by the Copernicus Climate Change Service (CCCS) supported by the European Commission and part of the European Centre for Medium-Range Weather Forecasts (ECMWF) [22].

The performance of both MERRA2 and ERA5 have been evaluated against physical observations of windspeed for Sweden [23]; Brazil [24]; towers distributed across the world [25]; France [26]; offshore East Coast US [27]; and a multi-country study involving Brazil, New Zealand, the US, and South Africa [28]. MERRA2 and ERA5 have been compared against 10 m wind and shortwave solar data from seven sites in Ireland and found that ERA5 generally returned more accurate results [29]. Studies comparing both solar reanalysis models with readings at 57 BSRN stations across the world [30], and six Indonesian stations [31], also reported better agreement with ERA5 than MERRA2. Relatively little research has been published on reanalysis model performance for the UK. Sharp et al. reported that the CFSR reanalysis model returns a correlation coefficient of 0.81 for the 355 UK onshore locations considered [32]. Cannon et al. evaluated the original MERRA reanalysis hourly wind model against readings from every UK MIDAS station and found a mean correlation coefficient of 0.73 [33]. Similarly, 10 m height ERA5 windspeed data has been compared against observations across the UK and Europe and returned correlation coefficients between 0.6 and 0.85 [34]. To our knowledge, no publications address solar energy reanalysis model performance in the UK.

This paper investigates the timespan required for accurate modelling of hourly windspeed and insolation data, and then evaluates the accuracy of MERRA2 and ERA5 reanalysis models for use in the UK. Section 2 explains the methods employed; Section 3.1 shows the relationship between number of years of data and pdf accuracy; Section 3.2 provides accuracy metrics for MERRA2 and ERA5 in the UK; and Section 4 extracts the main conclusions from this work.

2 Method

Eight UK Met Office stations at which wind and solar data was available for ten calendar years from 1st January 2010 to 31st December 2019 were used in the comparison. These reflect the population distribution of the UK, are shown in Table 1, and mapped in Figure 1. They span from Dyce, Aberdeenshire in the north to Liscombe, Somerset in the south. For each location, comparable hourly data for wind speed and short-wave insolation was downloaded from the UK Met Office observations (MIDAS) [35], and MERRA2 and ERA5 reanalysis sets, as detailed in Table 2.

The MERRA2 datasets downloaded were the u and v wind components at 10 m height [36], and surface net downward shortwave flux [37]. MERRA2 solar datapoints are timestamped at the half-hour (e.g. 5:30, 6:30); to compare them with the other datasets, it is necessary to resolve them onto on-hour timestamps [30]. The ERA5 datasets downloaded were the u and v wind components at 10 m height and the surface solar downward radiation from the "hourly data on single levels" model [38]. For both reanalysis models, the u and v wind velocity components were resolved into a scalar windspeed w_{mod} by

$$w_{\text{mod}} = \sqrt{u^2 + v^2} \; . \qquad (2)$$

Data from the leap days in the period, 29th February 2012 and 2016, was excluded from the analysis; datapoints returning error values were similarly excluded from the averages for the dataset. Further information about each dataset, along with data availability, is shown in Table 2.

2.1 Timespan selection for modelling

To evaluate the effect of the timespan selection on the resulting probability density function (pdf), for each location a pdf of the full 10-year dataset was compared with pdfs of pseudorandom permutations of $N = [1{:}9]$ years of data. The Weibull distribution, for which the general function is

$$f(x) = \left(\frac{\beta}{\alpha}\right)\left(\frac{x}{\alpha}\right)^{\beta-1} \exp\left[-\left(\frac{x}{\alpha}\right)^{\beta}\right], \tag{3}$$

is widely used to model both windspeed [39]–[41] and insolation [42]–[44]. The parameters α and β were obtained using the maximum likelihood method in which the equations

$$\alpha = \left[\frac{1}{n}\sum_{i=1}^{n} x_i^{\beta}\right]^{1/\beta}, \text{ and} \tag{4}$$

$$\beta = \frac{n}{\frac{1}{\alpha}\sum_{i=1}^{n} x_i^{\beta} \log x_i - \sum_{i=1}^{n} \log x_i}, \tag{5}$$

are solved simultaneously. The pseudorandom permutations of N years were generated by making a string of "cards" with length N containing pseudorandom integers using the MATLAB *randperm* function; these pseudorandom integers corresponded to each of the 10 years under evaluation and were used to recall the wind and insolation data for those years (e.g. for $N = 3$, the pseudorandom string [1,7,8] recalls the data for the years 2010, 2016, and 2017). For each location and N years combination, $n = 320$ repeats of pseudorandom permutations were performed. The Weibull pdf values were computed at 1000 evenly-spaced points between 0.1 and the maxima. Mean error and RMSE between the full 10 years dataset and the N years sample were calculated as

$$\varepsilon = \frac{1}{n}\sum_{i=1}^{n}\left(x_{i,10} - x_{i,N}\right), \tag{6}$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_{i,10} - x_{i,N}\right)^2}, \tag{7}$$

where x_{10} and x_N refer to the 10 years and N years datasets respectively.

The effect of sample timespan on P50 and P90 methodologies was also evaluated. For each location, N, and random permutation, the mean was obtained, the normally-distributed pdf was fitted, the P50 and P90 windspeed and insolation values read, and thus ε and RMSE obtained from eqns. (6) and (7).

2.2 Comparison of reanalysis models

The accuracy of the reanalysis models was compared with the MIDAS observations by the mean error ε, the root-mean-squared-error RMSE, and the coefficient of determination r^2:

$$\varepsilon = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - \hat{x}_i\right), \tag{8}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \hat{x}_i\right)^2}\ \text{, and} \tag{9}$$

$$r^2 = 1 - \frac{\sum_{i=1}^{n}\left(x_i - \hat{x}_i\right)^2}{\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}, \tag{10}$$

where x is the observed variable, $\bar{x}$ is the mean observed variable, and $\hat{x}$ is the modelled variable.

3 Results and discussion

3.1 Timespan selection for energy modelling

The Weibull pdfs for each dataset and location is shown in Figure 2 and the corresponding mean probability error in Figure 3. The use of a single year's data is shown to underpredict both windspeed and insolation across the MIDAS and reanalyses datasets. This trend towards underprediction continues whilst $N < 5$ for both MIDAS and MERRA2 insolation. Figure 4 shows that RMSE generally decreases whilst $N < 5$ and thereafter does not change significantly. Figures Figure 5(a-c)Figure 7 (a, c) also clearly display this trend for the P50 error parameters and substantiate Bodini et al.'s findings that there was negligible advantage in increasing N beyond ~5 years [15]. The P90 error results in Figure 6 and Figure 7 (b, d) are quite different: for the full range of years tested, the P90 resulting from N years was lower than the P90 for the full 10 years, with both ε and RMSE decreasing with increasing *N*; furthermore, for each *N*, the absolute values of the P90 errors are considerably larger than the P50 errors.

The computational effort, in terms of power and time, required to model meteorological data in energy systems increases linearly with the timespan of the data. Our results suggest that, when producing Weibull pdfs for direct modelling or the P50 methodology for aggregated modelling, there is negligible advantage in using timespans in excess of five years. However, where a conservative aggregated methodology such as P90 is employed, for the 10 year timespan tested in this paper, increasing timespan reduces the magnitude of the underestimate. Nonetheless, as methodologies requiring the download pre-aggregated data require significantly less computational effort than methodologies requiring the download of high temporal (e.g. hourly) resolution data, P90 calculation with a full 10 years' annual mean data is still significantly quicker than producing a pdf of hourly data for 5 years. The reanalysis models displayed the same general trends as the MIDAS observations, suggesting that reanalysis data may be suitable for analysis of timespan effects when observational data is unavailable or incomplete.

The implications of these findings on the modelling and optimisation of energy systems is not explored in this paper; we hope to evaluate the effect of meteorological modelling method on the modelling of small- and large- scale energy systems in a subsequent paper. Neither has this present work tested the part of those hypotheses that suggest that extreme wind events would be better modelled with multiple years' data; a proper study of extreme weather event probabilities requires dataset timespans much larger than the 10 years considered here [33], [45].

3.2 Comparison of reanalysis models

Unlike MIDAS and MERRA2 insolation data, ERA5 did not return a null value of 0 W m^{-2} when such might be expected, for example during the night, but instead returned a value of -528 μW m^{-2} at these times. The implication, that the earth surface emits shortwave radiation during the night, is correct as visible light (e.g. from streetlights) is emitted upward from the surface of the earth [46]. However, unlike the uniform negative insolation value returned across all locations in ERA5, in reality nocturnal shortwave radiation emissions vary across locations. This question is not captured by the MIDAS data or modelled in MERRA2 and therefore introduces a miniscule source of systematic error into the comparisons.

Figure 8 and Figure 9, and Table 3, show the relationships between the reanalysis data and MIDAS observations for wind and solar for the year 2014. Considered across locations, both reanalysis models slightly overpredicted windspeed, although in the case of ERA5 the mean error was negligible; showing very similar results to the Irish study of Clarke et al. [29]. MERRA2 tended to overpredict across all windspeeds, whilst ERA5 tended to underpredict at higher windspeeds. Jourdier observed similarly that ERA5 underpredicted across all terrain types, whilst MERRA2 considerably overpredicted in flat terrain and underpredicted in mountainous terrain [26]. Our analysis does not indicate a significant relationship between predictive performance and terrain although it is acknowledged that none of the MIDAS locations used are in mountainous terrain comparable to the Alpine sites investigated by Jourdier. As previously reported [23]–[28], ERA5 consistently outperformed MERRA2 in windspeed predictive accuracy metrics.

Figure 9 shows that both reanalysis models underpredict at high insolation levels, which corroborates the Indonesian study of Sianturi et al. [31] and the multinational study of Yang and Bright [30]. Again, ERA5 generally outperformed MERRA2 as has previously been observed [30], [31]. The lower MERRA2 ME at some locations is explained by the higher MERRA2 RMSE at all locations and by Figure 9: the error distribution was more uniform with MERRA2, whilst ERA5 systematically underpredicted although the absolute magnitude of these errors was lower.

4 Conclusion

Evaluation of the effect of timespan, and comparison of MERRA2 and ERA5 reanalysis against UK MIDAS observations, for windspeed and insolation was performed with the following main conclusions:

- For each dataset, for both windspeed and insolation, and for full data pdf and P50 methodologies, increasing the number of years used to calculate the parameter generally decreased mean and RMS errors compared to the parameter calculated by the full timespan of data until around the fifth year; thereafter, the error magnitude was small and no significant relationship between N and error was observed.
- Whilst both windspeed and insolation P90s were consistently lower for each N than for the full timespan of data, increasing N consistently decreased errors.

- Both MERRA2 and ERA5 returned good correlation with MIDAS observations, with ERA5 slightly more accurate overall for both windspeed and insolation. ERA5 reanalysis can therefore be recommended for meteorological modelling in the UK.

Declarations

This work was funded by the European Union Regional Development Fund – Interreg – Northern Periphery and Arctic Programme (project number 366, SW-GROW – Innovations for Seaweed Producers in the Northern Periphery Area project; http://sw-grow.interreg-npa.eu/).

The authors have no conflicts of interest to declare.

References

[1] Y. Lai and D. A. Dzombak, "Use of Historical Data to Assess Regional Climate Change," *J. Clim.*, vol. 32, no. 14, pp. 4299–4320, Jul. 2019, doi: 10.1175/JCLI-D-18-0630.1.

[2] M. De Rosa, V. Bianco, F. Scarpa, and L. A. Tagliafico, "Historical trends and current state of heating and cooling degree days in Italy," *Energy Convers. Manag.*, vol. 90, pp. 323–335, Jan. 2015, doi: 10.1016/j.enconman.2014.11.022.

[3] I. Staffell and S. Pfenninger, "The increasing impact of weather on electricity supply and demand," *Energy*, vol. 145, pp. 65–78, 2018, doi: 10.1016/j.energy.2017.12.051.

[4] V. Khare, S. Nema, and P. Baredar, "Solar-wind hybrid renewable energy system: A review," *Renew. Sustain. Energy Rev.*, vol. 58, pp. 23–33, 2016, doi: 10.1016/j.rser.2015.12.223.

[5] B.-M. Hodge, M. Hummon, and K. Orwig, "Solar Ramping Distributions over Multiple Timescales and Weather Patterns," Golden, CO, 2011. [Online]. Available: https://www.nrel.gov/docs/fy15osti/52735.pdf.

[6] L. Luo, C. Fang, Y. Yang, C. Li, F. Li, and P. Zhang, "Study on probability distribution of photovoltaic power fluctuations at multi-time scales," 2018, doi: 10.34890/441.

[7] J. T. Huang, Y. Q. Liu, and S. Han, "The time scale study of wind speed probability distribution," in *2nd IET Renewable Power Generation Conference (RPG 2013)*, 2013,

pp. 3.17-3.17, doi: 10.1049/cp.2013.1824.

[8] S. C. Pryor, T. J. Shepherd, and R. J. Barthelmie, "Interannual variability of wind climates and wind turbine annual energy production," *Wind Energy Sci.*, vol. 3, no. 2, pp. 651–665, Sep. 2018, doi: 10.5194/wes-3-651-2018.

[9] E. B. Mora, J. Spelling, A. H. van der Weijde, and E. M. Pavageau, "The effects of mean wind speed uncertainty on project finance debt sizing for offshore wind farms," *Appl. Energy*, vol. 252, no. June, 2019, doi: 10.1016/j.apenergy.2019.113419.

[10] M. L. Kubik, D. J. Brayshaw, P. J. Coker, and J. F. Barlow, "Exploring the role of reanalysis data in simulating regional wind generation variability over Northern Ireland," *Renew. Energy*, vol. 57, pp. 558–561, Sep. 2013, doi: 10.1016/j.renene.2013.02.012.

[11] J. C.-Y. Lee, M. J. Fields, J. K. Lundquist, and M. Lunacek, "Determining variabilities of non-Gaussian wind-speed distributions using different metrics and timescales," *J. Phys. Conf. Ser.*, vol. 1037, p. 072038, Jun. 2018, doi: 10.1088/1742-6596/1037/7/072038.

[12] J. C. Y. Lee, M. J. Fields, and J. K. Lundquist, "Assessing variability of wind speed: comparison and validation of 27 methodologies," *Wind Energy Sci.*, vol. 3, no. 2, pp. 845–868, Nov. 2018, doi: 10.5194/wes-3-845-2018.

[13] H. Chen, Y. Birkelund, S. N. Anfinsen, R. Staupe-Delgado, and F. Yuan, "Assessing probabilistic modelling for wind speed from numerical weather prediction model and observation in the Arctic," *Sci. Rep.*, vol. 11, no. 1, p. 7613, Dec. 2021, doi: 10.1038/s41598-021-87299-4.

[14] B. Safari and J. Gasore, "A statistical investigation of wind characteristics and wind energy potential based on the Weibull and Rayleigh models in Rwanda," *Renew. Energy*, vol. 35, no. 12, pp. 2874–2880, Dec. 2010, doi: 10.1016/j.renene.2010.04.032.

[15] N. Bodini, J. K. Lundquist, D. Zardi, and M. Handschy, "Year-to-year correlation, record length, and overconfidence in wind resource assessment," *Wind Energy Sci.*, vol. 1, no. 2, pp. 115–128, 2016, doi: 10.5194/wes-1-115-2016.

[16] L. C. Slivinski, "Historical Reanalysis: What, How, and Why?," *J. Adv. Model. Earth Syst.*, vol. 10, no. 8, pp. 1736–1739, Aug. 2018, doi: 10.1029/2018MS001434.

[17] J. Olauson and M. Bergkvist, "Modelling the Swedish wind power production using MERRA reanalysis data," *Renew. Energy*, vol. 76, pp. 717–725, Apr. 2015, doi: 10.1016/j.renene.2014.11.085.

[18] L. Ramirez Camargo, K. Gruber, and F. Nitsch, "Assessing variables of regional reanalysis data sets relevant for modelling small-scale renewable energy systems," *Renew. Energy*, vol. 133, pp. 1468–1478, 2019, doi: 10.1016/j.renene.2018.09.015.

[19] H. C. Bloomfield *et al.*, "The Importance of Weather and Climate to Energy Systems: A Workshop on Next Generation Challenges in Energy–Climate Modeling," *Bull. Am. Meteorol. Soc.*, vol. 102, no. 1, pp. E159–E167, Jan. 2021, doi: 10.1175/BAMS-D-20-0256.1.

[20] W. S. Parker, "Reanalyses and Observations: What's the Difference?," *Bull. Am. Meteorol. Soc.*, vol. 97, no. 9, pp. 1565–1572, Sep. 2016, doi: 10.1175/BAMS-D-14-00226.1.

[21] R. Gelaro *et al.*, "The Modern-Era Retrospective Analysis for Research and Applications, Version 2 (MERRA-2)," *J. Clim.*, vol. 30, no. 14, pp. 5419–5454, Jul. 2017, doi: 10.1175/JCLI-D-16-0758.1.

[22] H. Hersbach *et al.*, "The ERA5 global reanalysis," *Q. J. R. Meteorol. Soc.*, vol. 146, no. 730, pp. 1999–2049, Jul. 2020, doi: 10.1002/qj.3803.

[23] J. Olauson, "ERA5: The new champion of wind power modelling?," *Renew. Energy*, vol. 126, pp. 322–331, Oct. 2018, doi: 10.1016/j.renene.2018.03.056.

[24] J. Santos *et al.*, "Wind Speed Evaluation of MERRA-2, ERA-Interim and ERA-5 Reanalysis Data at a Wind Farm Located in Brazil," in *Proceedings of the ISES Solar World Congress 2019*, 2019, pp. 1–10, doi: 10.18086/swc.2019.45.10.

[25] J. Ramon, L. Lledó, V. Torralba, A. Soret, and F. J. Doblas-Reyes, "What global reanalysis best represents near-surface winds?," *Q. J. R. Meteorol. Soc.*, vol. 145, no. 724, pp. 3236–3251, Oct. 2019, doi: 10.1002/qj.3616.

[26] B. Jourdier, "Evaluation of ERA5, MERRA-2, COSMO-REA6, NEWA and AROME to simulate wind power production over France," *Adv. Sci. Res.*, vol. 17, pp. 63–77, Jun. 2020, doi: 10.5194/asr-17-63-2020.

[27] L. M. Sheridan, R. Krishnamurthy, and B. J. Gaudet, "Assessment of Model Hub

Height Wind Speed Performance Using DOE Lidar Buoy Data," Alexandria, VA, 2021. [Online]. Available: https://www.osti.gov/servlets/purl/1779495.

[28] K. Gruber, P. Regner, S. Wehrle, M. Zeyringer, and J. Schmidt, "Towards global validation of wind power simulations: A multi-country assessment of wind power simulation from MERRA-2 and ERA-5 reanalyses bias-corrected with the global wind atlas," *Energy*, vol. 238, p. 121520, 2022, doi: 10.1016/j.energy.2021.121520.

[29] E. D. Clarke, S. Griffin, F. McDermott, J. M. Correia, and C. Sweeney, "Which reanalysis dataset should we use for renewable energy analysis in Ireland?," *Atmosphere (Basel).*, vol. 12, no. 5, pp. 1–13, 2021, doi: 10.3390/atmos12050624.

[30] D. Yang and J. M. Bright, "Worldwide validation of 8 satellite-derived and reanalysis solar radiation products: A preliminary evaluation and overall metrics for hourly data over 27 years," *Sol. Energy*, vol. 210, no. February, pp. 3–19, 2020, doi: 10.1016/j.solener.2020.04.016.

[31] Y. Sianturi, Marjuki, and K. Sartika, "Evaluation of ERA5 and MERRA2 reanalyses to estimate solar irradiance using ground observations over Indonesia region," 2020, doi: 10.1063/5.0000854.

[32] E. Sharp, P. Dodds, M. Barrett, and C. Spataru, "Evaluating the accuracy of CFSR reanalysis hourly wind speed forecasts for the UK, using in situ measurements and geographical information," *Renew. Energy*, vol. 77, pp. 527–538, May 2015, doi: 10.1016/j.renene.2014.12.025.

[33] D. J. Cannon, D. J. Brayshaw, J. Methven, P. J. Coker, and D. Lenaghan, "Using reanalysis data to quantify extreme wind power generation statistics: A 33 year case study in Great Britain," *Renew. Energy*, vol. 75, pp. 767–778, 2015, doi: 10.1016/j.renene.2014.10.024.

[34] M. O. Molina, C. Gutiérrez, and E. Sánchez, "Comparison of ERA5 surface wind speed climatologies over Europe with observations from the HadISD dataset," *Int. J. Climatol.*, vol. 41, no. 10, pp. 4864–4878, Aug. 2021, doi: 10.1002/joc.7103.

[35] Met Office, "Met Office MIDAS Open: UK Land Surface Stations Data (1853-current)," Exeter, 2019. [Online]. Available: http://catalogue.ceda.ac.uk/uuid/dbd451271eb04662beade68da43546e1.

[36] Global Modeling and Assimilation Office (GMAO), “MERRA-2 inst1_2d_asm_Nx: 2d, 1-Hourly, Instantaneous, Single-Level, Assimilation, Single-Level Diagnostics V5.12.4,” Greenbelt, MD, 2015. doi: 10.5067/3Z173KIE2TPD.

[37] Global Modeling and Assimilation Office (GMAO), “MERRA-2 tavg1_2d_rad_Nx: 2d, 1-Hourly,Time-Averaged, Single-Level, Assimilation, Radiation Diagnostics V5.12.4,” Greenbelt, MD, 2015. doi: 10.5067/Q9QMY5PBNV1T.

[38] H. Hersbach *et al.*, “ERA5 hourly data on single levels from 1979 to present,” Reading, 2018. doi: 10.24381/cds.adbb2d47.

[39] G. J. Bowden, P. R. Barker, V. O. Shestopal, and J. W. Twidell, “The Weibull Distribution Function and Wind Power Statistics,” *Wind Eng.*, vol. 7, no. 2, pp. 85–98, 1983, [Online]. Available: https://www.jstor.org/stable/43749036.

[40] J. V. Seguro and T. W. Lambert, “Modern estimation of the parameters of the Weibull wind speed distribution for wind energy analysis,” *J. Wind Eng. Ind. Aerodyn.*, vol. 85, no. 1, pp. 75–84, Mar. 2000, doi: 10.1016/S0167-6105(99)00122-1.

[41] K. Mohammadi, O. Alavi, A. Mostafaeipour, N. Goudarzi, and M. Jalilvand, “Assessing different parameters estimation methods of Weibull distribution to compute wind power density,” *Energy Convers. Manag.*, vol. 108, pp. 322–335, Jan. 2016, doi: 10.1016/j.enconman.2015.11.015.

[42] G. Tina and S. Gagliano, “Probabilistic analysis of weather data for a hybrid solar/wind energy system,” *Int. J. Energy Res.*, vol. 35, no. 3, pp. 221–232, Mar. 2011, doi: 10.1002/er.1686.

[43] C. M. Fernández Peruchena *et al.*, “A statistical characterization of the long-term solar resource: Towards risk assessment for solar power projects,” *Sol. Energy*, vol. 123, pp. 29–39, Jan. 2016, doi: 10.1016/j.solener.2015.10.051.

[44] O. M. Kam, S. Noël, H. Ramenah, P. Kasser, and C. Tanougast, “Comparative Weibull distribution methods for reliable global solar irradiance assessment in France areas,” *Renew. Energy*, vol. 165, pp. 194–210, Mar. 2021, doi: 10.1016/j.renene.2020.10.151.

[45] F. E. L. Otto, S. Philip, S. Kew, S. Li, A. King, and H. Cullen, “Attributing high-impact extreme events across timescales—a case study of four different types of events,” *Clim. Change*, vol. 149, no. 3–4, pp. 399–412, Aug. 2018, doi:

10.1007/s10584-018-2258-3.

[46] X. Li, Y. Zhou, M. Zhao, and X. Zhao, “A harmonized global nighttime light dataset 1992–2018,” *Sci. Data*, vol. 7, no. 1, p. 168, Dec. 2020, doi: 10.1038/s41597-020-0510-y.

Nomenclature

N	number of years	
n	number of datapoints	
r	correlation coefficient	
RMSE	root-mean-squared error	
S	insolation	W m^{-2}
u	wind velocity	m s^{-1}
v	wind velocity	m s^{-1}
w	windspeed	m s^{-1}
x	placeholder variable	
α	Weibull parameter	
β	Weibull parameter	
ε	error	

Tables

Table 1: MIDAS locations under comparison.

N°$_L$	Location	Coordinates
1	Dyce, Aberdeenshire	57.205° N, 2.205° W
2	Gogarbank, Midlothian	55.928° N, 3.344° W
3	Aldergrove, Antrim	54.664° N, 6.225° W
4	Leeming, North Yorkshire	54.297° N, 1.533° W
5	Valley, Gwynedd	53.253° N, 4.536° W
6	Sutton Bonnington, Nottinghamshire	52.836° N, 1.251° W
7	Heathrow, Greater London	51.479° N, 0.451° W
8	Liscombe, Somerset	51.087° N, 3.609° W

Table 2: Dataset information and data availability.

	MIDAS		MERRA2		ERA5	
Organisation	UK Met Office		NASA GMAO		ECMWF: Copernicus	
Ref.	[35]		[36]	[37]	[38]	
Spatial resolution	point		0.5° x 0.625°		0.1° x 0.1°	
Data availability (%)						
	w	***S***	***w***	***S***	***w***	***S***
mean	**99.0**	**99.3**	**97.0**	**100**	**100**	**100**
Location 1	95.2	99.2	97.7	100	100	100
2	99.9	99.2	97.3	100	100	100
3	99.7	99.3	95.2	99.9	100	100
4	98.8	98.9	96.2	100	100	100
5	99.8	99.7	100	100	100	100
6	99.8	99.8	95.3	100	100	100
7	99.3	99.2	98.2	99.8	100	100
9	99.5	98.8	96.3	100	100	100

Table 3: Error metrics of MERRA2 (M2) and ERA5 (E5) reanalysis models for windspeed and insolation, each compared with MIDAS data.

Metric	**Variable**	**Model**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**mean**
ε	***w*** **(m s^{-1})**	**M2**	1.07	1.49	1.27	1.24	.829	1.57	.822	.859	**1.14**
		E5	0.524	.104	-.204	5.41x10^{-3}	-.930	1.06	-3.07x10^{-2}	.458	**.123**
	S (W m^{-2})	**M2**	-1.90	-2.27	-5.59	-2.65	-4.05	-5.07	-3.53	2.06	**-2.88**
		E5	6.32	-2.36	1.08	-3.04	-2.63	-1.83	-3.03	4.10	**-1.74x10^{-2}**
RMSE	***w*** **(m s^{-1})**	**M2**	2.29x10^{-2}	2.37x10^{-2}	2.09x10^{-2}	2.31x10^{-2}	2.35x10^{-2}	2.32x10^{-2}	1.66x10^{-2}	1.92x10^{-2}	**2.17x10^{-2}**
		E5	1.62x10^{-2}	1.31x10^{-2}	1.27x10^{-2}	1.48x10^{-2}	1.98x10^{-2}	1.72x10^{-2}	1.21x10^{-2}	1.62x10^{-2}	**1.50x10^{-2}**
	S (W m^{-2})	**M2**	.866	.783	.785	.821	.827	.788	.829	.866	**.821**
		E5	.746	.747	.726	.750	.713	.706	.731	.746	**.733**
r^2	***w*** **(-)**	**M2**	.680	.699	.717	.622	.720	.680	.716	.680	**.689**
		E5	.693	.774	.752	.703	.840	.691	.733	.693	**.735**
	S **(-)**	**M2**	.835	.828	.837	.828	.869	.849	.856	.834	**.842**
		E5	.877	.843	.860	.856	.903	.879	.887	.876	**.873**

Figures

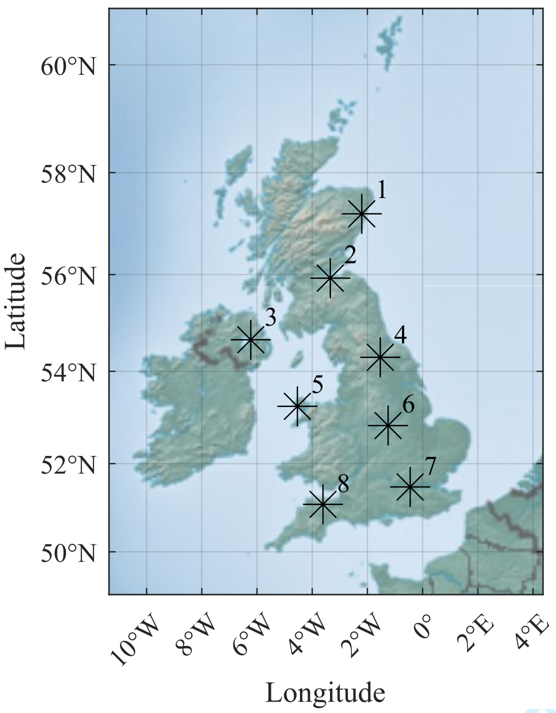


Figure 1: Location of MIDAS sites selected for data comparison.

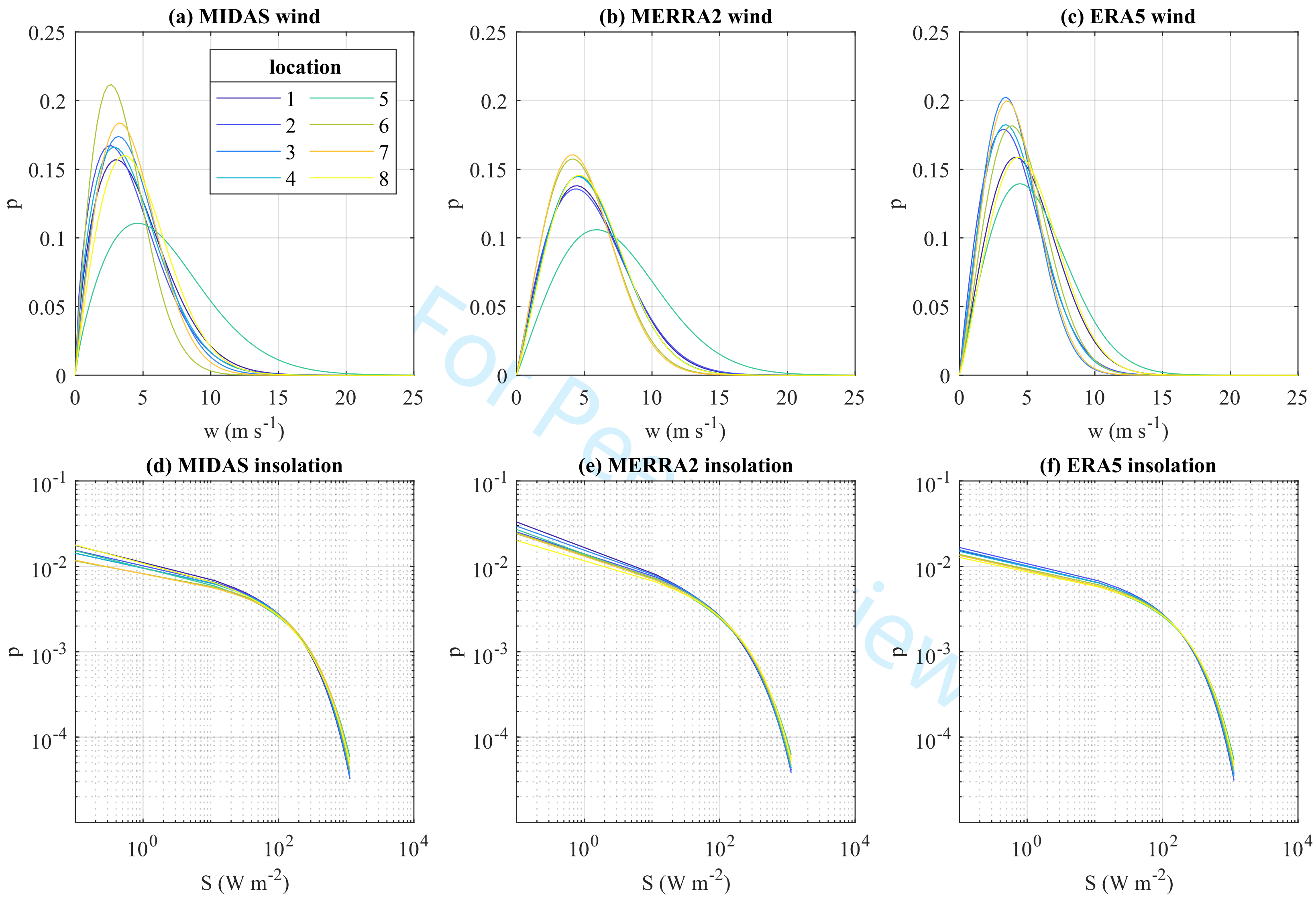


Figure 2: Weibull pdfs for datasets (a) to (f) for each location with the full data timespan (N = 10 years).

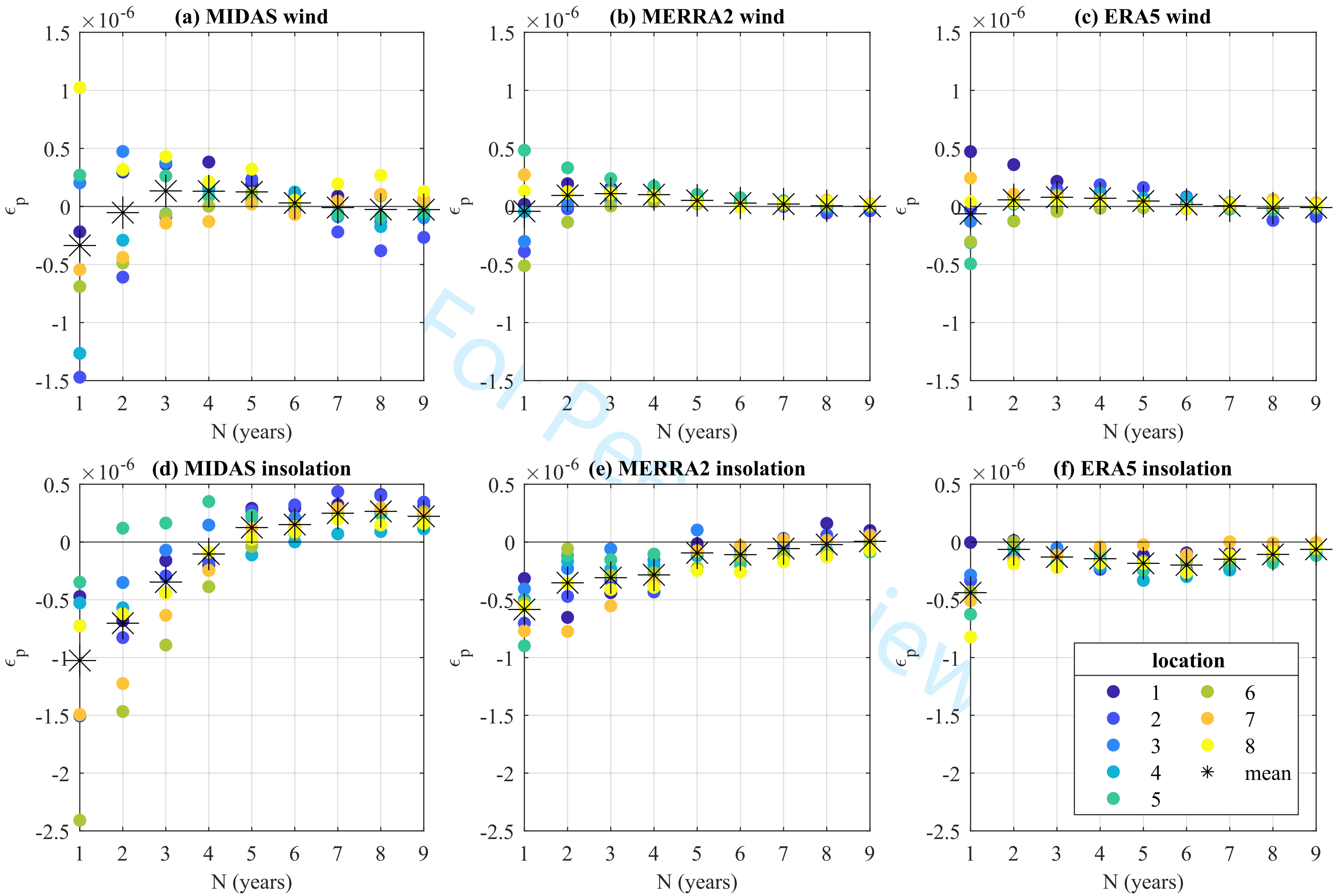


Figure 3: For datasets (a) to (f), mean probability error ε_p from the pdf of 320 random permutations of N years against the pdf generated by N = 10 years, showing each location and overall means across locations.

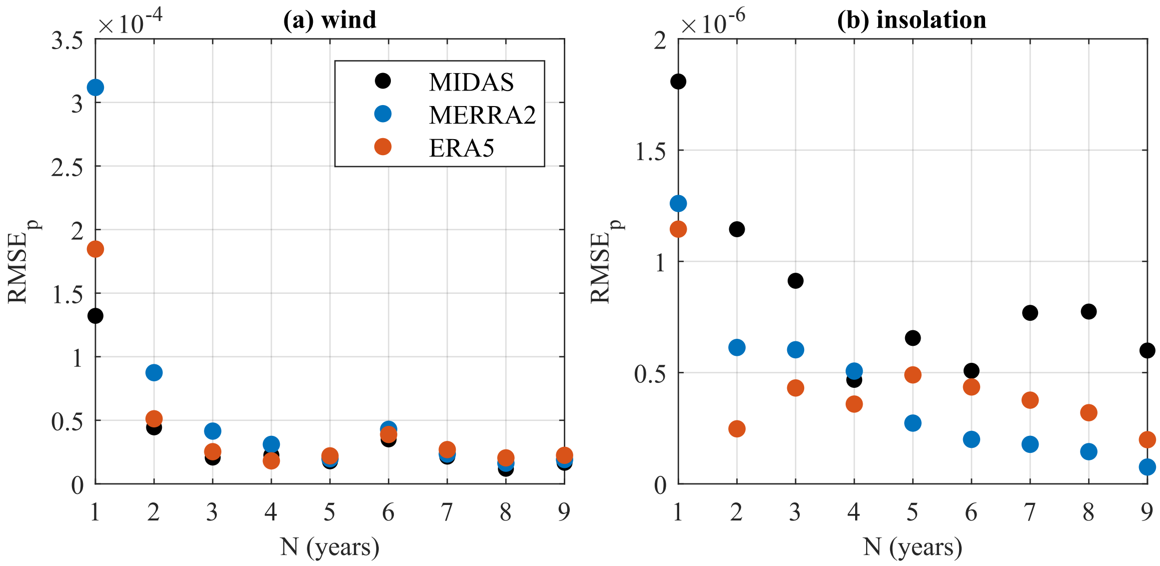


Figure 4: Mean RMSE across locations for (a) wind, showing negligible change where $N > 4$; and (b) insolation.

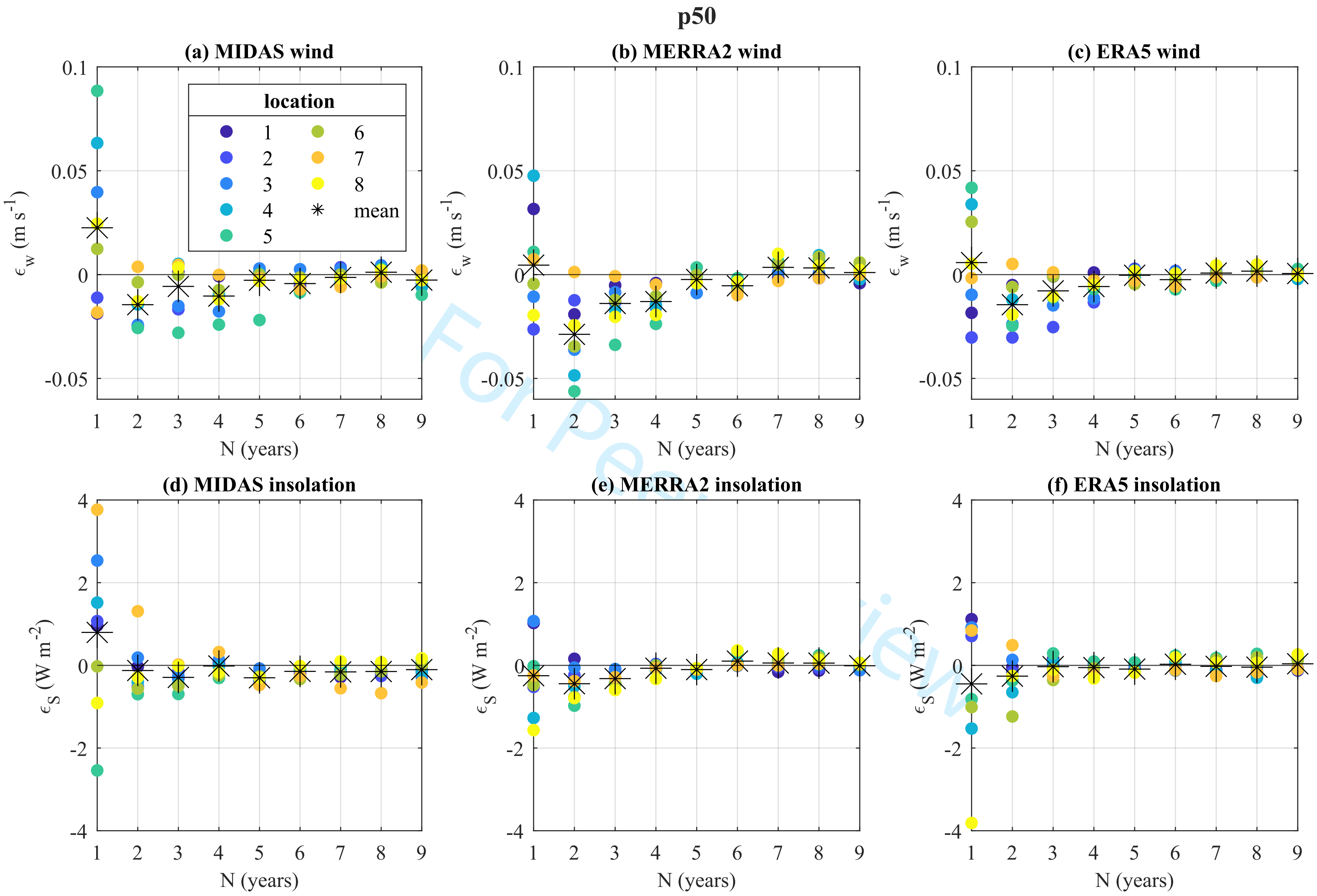


Figure 5: Mean p50 (a-c) windspeed error ε_w and (d-f) insolation error ε_S from random permutations of N years measured against $N = 10$ years, showing each location and overall means across locations.

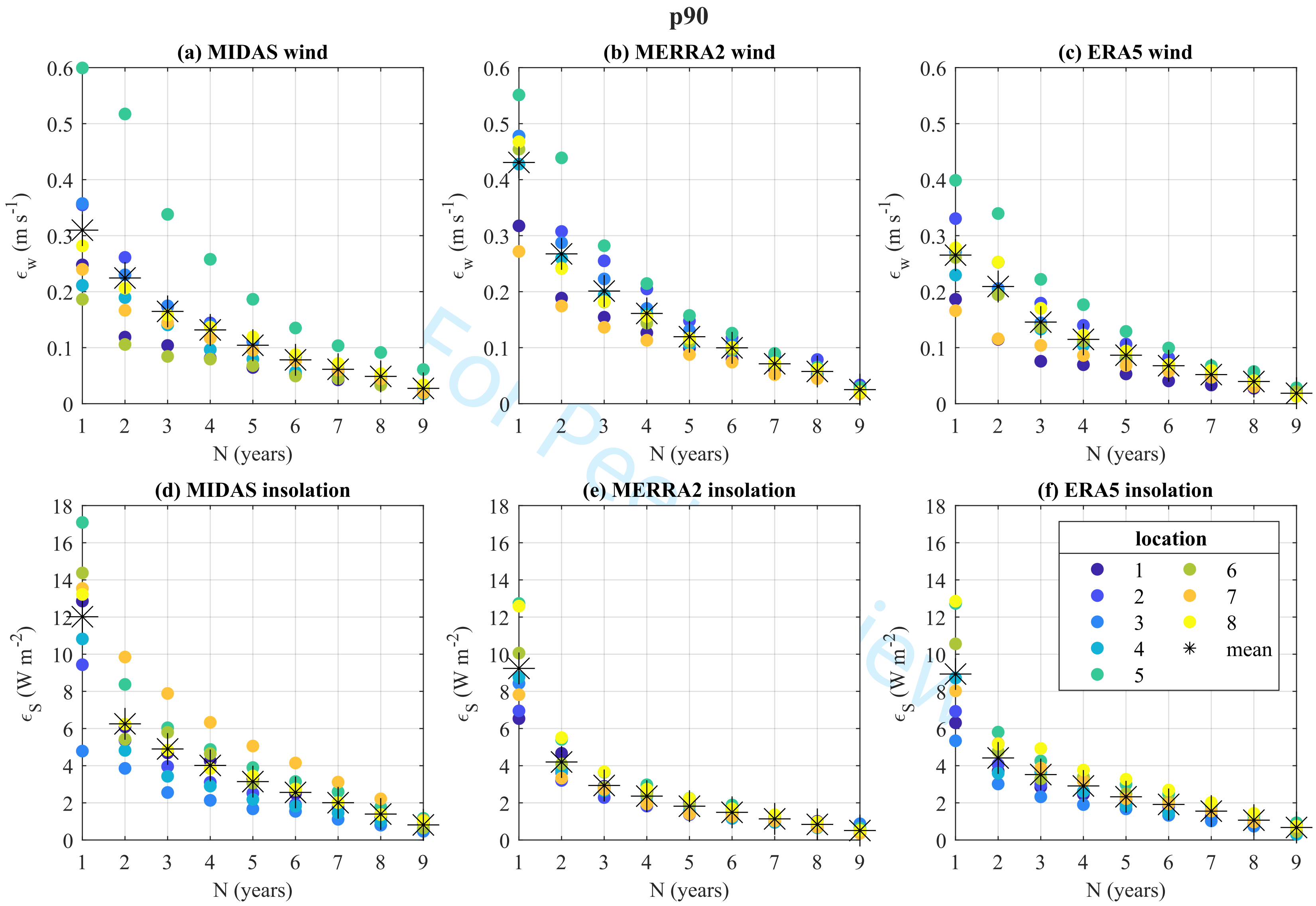


Figure 6: Mean p90 (a-c) windspeed error ε_w and (d-f) insolation error ε_S from random permutations of N years measured against $N = 10$ years. showing each location and overall means across locations.

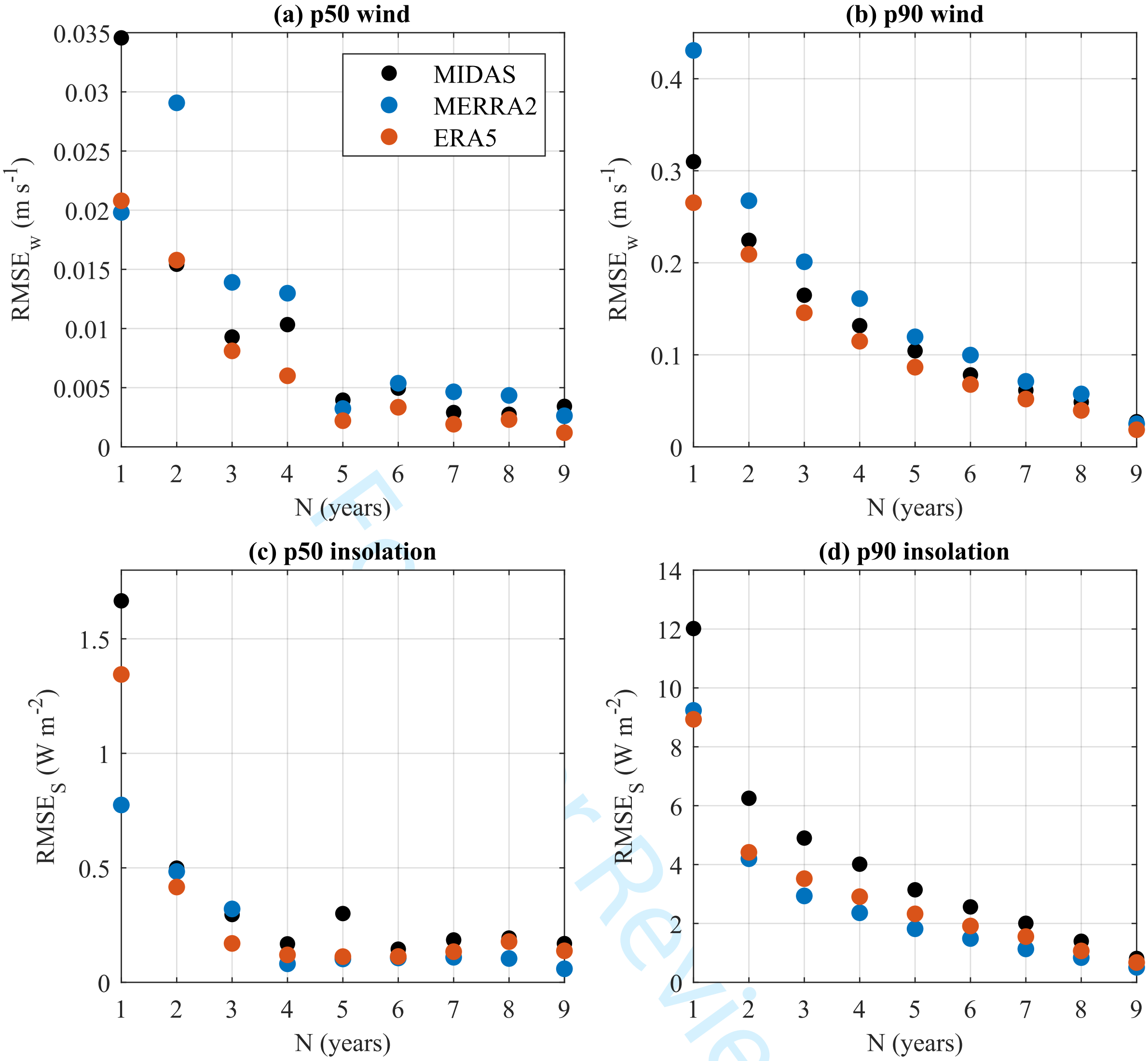


Figure 7: Mean RMSE across locations for p50 and p90 windspeed and insolation.

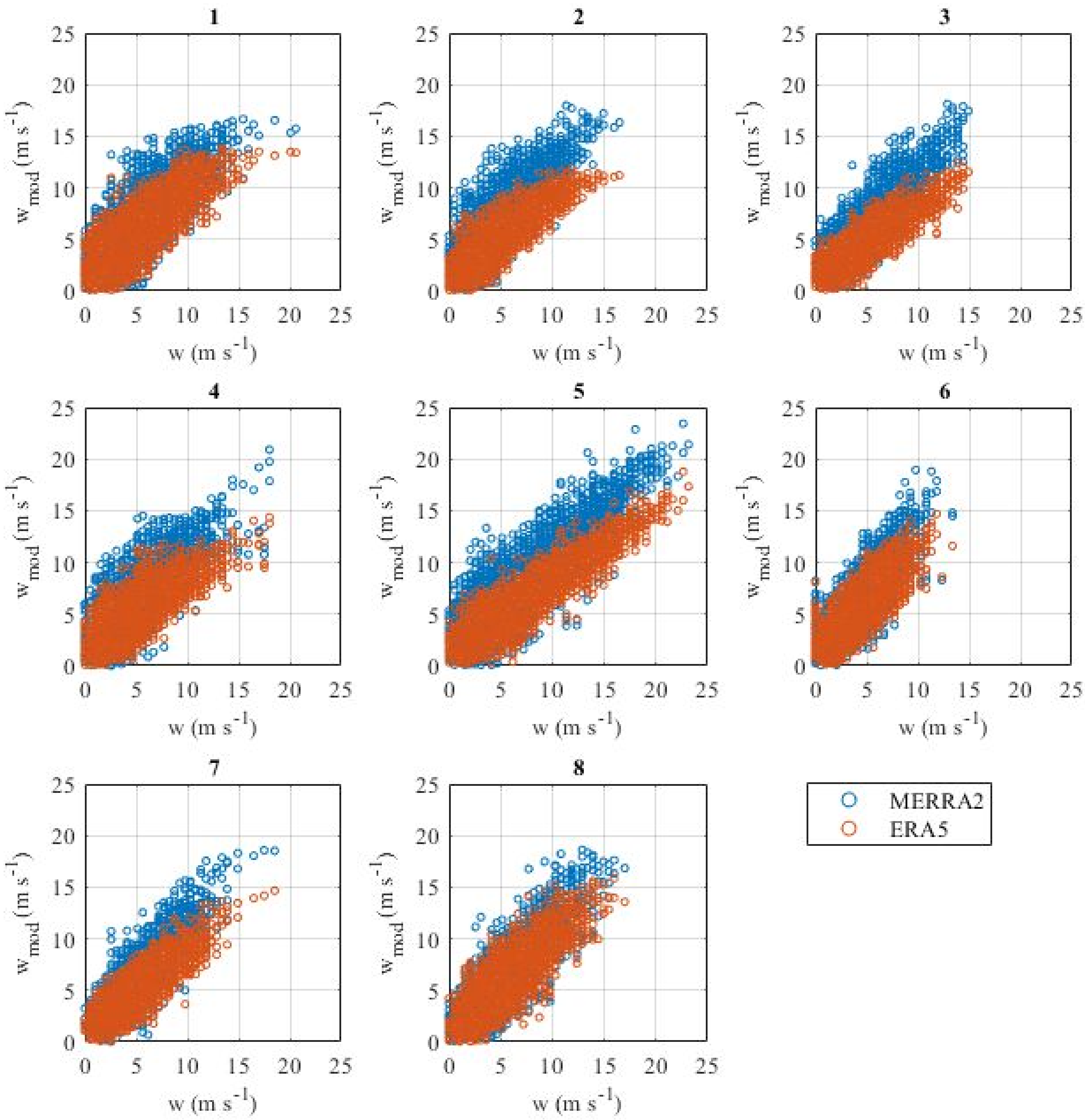


Figure 8: Correlation of wind MERRA2 and ERA5 reanalysis data compared to MIDAS observations at each location.

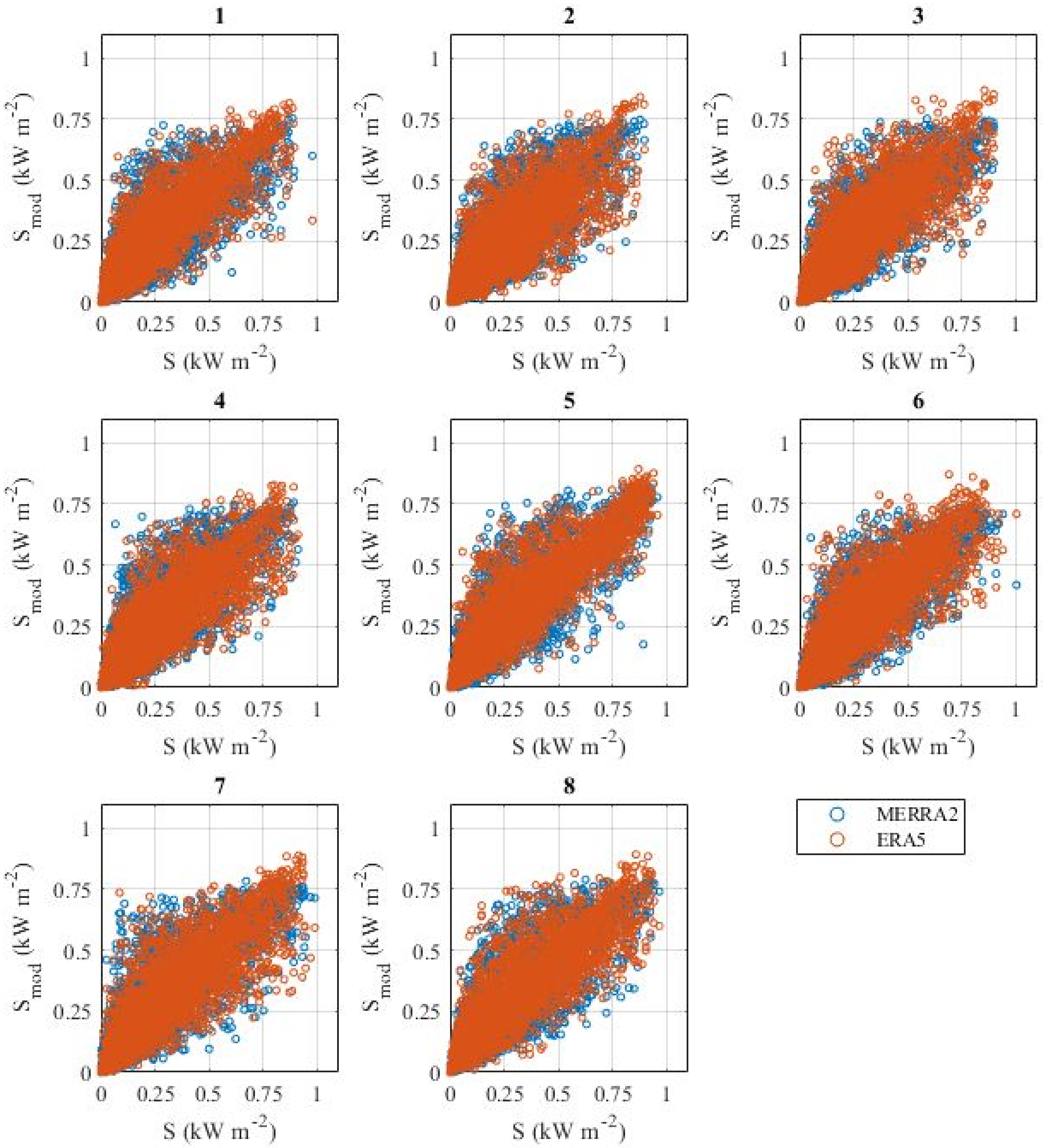


Figure 9: Correlation of solar MERRA2 and ERA5 reanalysis data compared to MIDAS observations at each location.