Forecast Accuracy for Solar Resource in Tshwane using SolCast Data Services

Lawrence Pratt [1*], Manjunath Basappa Ayanna [2], Wisani Mkasi [3]

[1] CSIR - Meiring Naude Road, Pretoria, South Africa, +27 060 688 5929

2 CSIR - mbasappaayanna@csir.co.za

3 CSIR - hmkasi@csir.co.za

Abstract: This paper compares forecasted global horizontal irradiance (GHI) data from SolCast to GHI recorded at the CSIR meteorological station in Tshwane, South Africa. The study compares the SolCast forecast to the CSIR measured values and computes accuracy metrics for +1-hour, solar noon, +1-day, and +7-day ahead forecasts. The +1-hour ahead and solar noon forecasts are based on the average value for a single hour, while the +1-day and +7-day ahead are based on the insolation over 24 hours. Based on the limited data collected for this study, the normalised root mean squared deviation (nRMSD) is 28.9% for the +1-hour ahead forecast at the CSIR for forecasts made in the morning. This forecast error from the morning forecasts is nearly double the nRMSD reported in a study conducted by the SolCast team on forecasts across 33 sites worldwide for forecasts made at all hours of the day (16.6%). The high forecast error is likely due to the relatively low irradiance value forecasted and measured in the morning hours of the winter season captured in the study dataset. By contrast, the nRMSD for forecasts made for solar noon of the same day was 11.3%, which falls within the range of values reported by the SolCast team. The normalised mean bias deviations (nMBD) for the daily insolation for +1-day and the +7-day ahead forecasts were both 5.2% at the CSIR.

Keywords: solar resource, GHI, forecasting, SolCast

1. Introduction

Forecasting electricity generation capacity is critical for balancing the national grid. Grid operators must know how much electrical energy to expect from each asset to effectively deal with planned and unplanned maintenance without interruption to the supply of electricity required to meet demand. Conventional generators powered by fossil fuels, uranium, or plutonium are relatively easy to forecast compared to generators powered by renewable resources such as wind and solar. Setting aside any issues with the maintenance of the physical plants that might render the generator offline, the forecast for conventional generators depends primarily on the available stockpile of fuel. Renewable generators have no stockpile of fuel, so they rely on what comes from nature in the moment and in the immediate future. Fossil fuel is essentially chemically stored energy converted from photons that reached Earth millions of years ago from the sun. That photon energy is locked in materials that can be mined, traded, transported, stored, and burned with relative predictability compared to the photon energy that reaches Earth in real-time and must be used immediately for energy conversion or storage. On the other hand, the supply of fossil fuels is subject to price fluctuations, political decisions, or environmental concerns that could impact supply but that would never impact the supply of solar and wind resources. As more renewable energy connects to the national grid, improved forecasting for renewable energy generators will become more critical for Eskom to properly manage and maintain the grid.

The Council for Scientific and Industrial Research (CSIR) is a national research council based in Tshwane, South Africa. The CSIR Solar PV team has operated a dual-axis tracker-based weather station (Kipp and Zonen Solys 2) since 2016 on the rooftop of Building 34 (-25.75020, 28.27857), which includes pyranometers and pyrheliometers measuring global horizontal irradiance (GHI), direct normal irradiance (DNI) and diffuse horizontal irradiance (DHI). SolCast is a DNV Company offering historical and forecast data for solar resources nearly worldwide on a commercial basis, while limited subscriptions are available to researchers for free. SolCast provides historical data and forecasts focused explicitly on the solar resources available anytime, anywhere, and globally. With an accurate forecast for the solar resource at a given PV plant, the electricity generation at that PV plant can be predicted accurately, given the strong linear relationship between the available solar resource in the plane of the PV array and the output of the PV array.

Concerning historical solar resource data, SolCast claims a 15.99 % error (nRMSD) for hourly GHI when comparing typical meteorological year (TMY) data with ground station data across 207 sites worldwide [1]. The nRMSD drops to 12.07 % for the 37 sites in Africa. Recent publications confirmed the best-in-class performance of SolCast historical data at multiple sites in South Africa for daily GHI estimates [2] and hourly GHI estimates [3]. However, the forecast accuracy has not been validated independently for South African locations.

A SolCast report [4] on forecast accuracy at 33 sites across the globe, including one ground-based measurement station in South Africa, but the precise location is unknown. The SolCast report claims a 16.6% error (nRMSD) for a +1-hour forecast of GHI averaged across all 33 sites worldwide, shown in Figure 1, ranging from 6.0% to 36.6%, depending on the site. The analysis was based on the differences between predicted values from SolCast and the measured data from the 33 high-quality measurement stations. The report notes the impact of including nocturnal zero values on the accuracy metrics for hourly production forecasts and hourly solar resource forecasts. Specifically, when nocturnal zero values are included in the hourly forecasts, the average hourly forecast error is reduced by nearly 50% because the data includes zero values. However, nighttime forecasts are useless for solar generators and are best ignored when quantifying forecast accuracy. The metrics for daily insolation values are no different when nocturnal zero values are included because the daily insolation values are the same with or without nocturnal values.

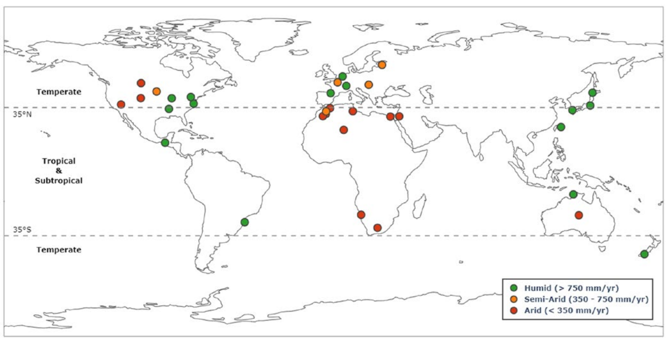

Figure 1 SolCast measurement stations used to determine the forecast accuracy [4]

2. Works Cited

DNV published a study on the historical time series data from SolCast versus surface irradiance measurements [1] to build confidence in the market regarding the dataset's quality. DNV, a global energy consulting firm, acquired SolCast in 2023. Historical solar resource data accuracy is critical for making accurate forecasts about future lifetime electricity production, a key component in every business case model for a PV plant. The DNV study evaluated 207 sites globally with ground-measurement stations. According to that study, the Solcast hourly GHI data exhibited a mean bias of +0.33% for all sites and +0.05% for the highest-quality GHI sites. The standard deviation for the mean bias was 2.47% for all sites, inferring that the mean bias between the modelled GHI and measured values for GHI should fall within +/- 4.94% of the mean bias for 95% of the sites anywhere on the planet. This is a relatively small error when developing performance models and estimating lifetime production on an annual basis.

Recent publications by researchers in South Africa, including the South Africa Weather Service, confirmed the excellent performance of SolCast historical data at multiple sites in South Africa for daily GHI estimates [2] and hourly GHI estimates [3]. These two reports from a team of South African researchers were also cited in the DNV report [1]. Among the five data sources evaluated, Table 7 [3] shows SolCast historical data was the best choice for 11 out of 13 stations in South Africa. CAMS, CMSAF, SolCast, ERA5, and MERRA2 were the five data sources. PVGIS, a free source for weather data, includes SARAH from 2005 to 2016 and SARAH2 from 2005 to 2020, which are also incorporated into CMSAF. ERA5 is also available on the PVGIS website from 2005 to 2020. The researchers concluded that the mean bias error for SolCast hourly GHI ranged from -3% to + 4% across the 13 sites evaluated in South Africa, averaged over the year, which is slightly better than the statistics reported by DNV for a larger sample of 207 measurement stations worldwide (+/- 4.94%).

The SolCast team published the results of a study that analysed forecasting errors to enable users to estimate Solcast real-time and forecast accuracy for their site(s) before subscription or integration effort [4]. The analysis of forecasts was conducted at 33 sites worldwide, including one in South Africa. The study compared GHI and power generation for real-time, +1-hour, +2-hour, and +24-hour ahead forecasts based on SolCast, the Smart Persistence Model, and GFS. The nRMSD for the +1-hour forecasts from SolCast averaged 16.6% across all 33 sites, with a range of 6.0% to 36.6% across sites. For the seven arid sites in the study, including the one site in South Africa, the nRMSD was 20.5%, ranging from 10.4% to 26.7%, which indicates a higher bias for this subset given the mean, but a more precise forecast given the narrower range across sites. Statistics for the 24-hour insolation forecasts were not provided.

The manuscript must be written in English. The standard font for the manuscript (Normal style) is Times New Roman for the text and Symbol for special characters. Body text should be in Times New Roman, 10 pt, justified. The text should be in a two-column format for the body of the document.

3. Methodology

Figure 2 shows the methodology for assessing the accuracy of SolCast forecasts for GHI at the CSIR weather station. The data preparation, visualisation, and analysis were all carried out using standard Python libraries, including pandas, numpy, and matplotlib. First, the dataset was assembled by merging the forecast data downloaded weekly from the SolCast website and the corresponding data from the CSIR weather station. The forecast data was adjusted from UTC to the local time zone (UTC +2) before the merge so that clear sky irradiance profiles from both sources lined up at sunrise and sunset. SolCast records the time stamp at the end of a sampling interval, while the CSIR records the time stamp at the beginning of a sampling interval. The data was aggregated appropriately for +1-hour, +1-day, and +7-day ahead forecasts. The data was plotted and reviewed to detect any visual anomalies, outliers, or mismatches in time between the forecast and the measured data. Finally, the metrics were computed from the datasets and interpreted.

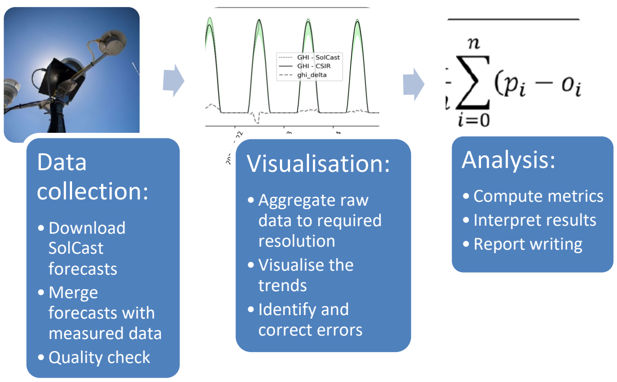

Figure 2: Methodology followed for assessing the accuracy of SolCast forecasts for GHI at the CSIR weather station

SolCast operates a global cloud tracking and forecasting system, using near-real-time satellite imagery from 11 weather satellites and weather data from 7 numerical weather forecast (NWP) models [4]. The Solcast irradiance algorithm consists of a cloud model, a clear sky model, a separation model, and a transposition model [1]. The inputs include data from geostationary satellite images, atmospheric pressure, water vapour and albedo, and topographic data. The weather data comes from ERA5. These inputs are used to model irradiance and PV power output at a given orientation, and the results are distributed via the Solcast Application Programming Interface (API), which enables automated, synchronous data requests for any point on Earth.

Several metrics were used to evaluate the accuracy of the forecasts by the SolCast team [4] and they are computed for this study, as well. The equations for mean absolute difference (MAD), mean bias difference (MBD), and root mean squared difference (RMSD) follow, in which p represents the predicted value from SolCast and o represents the observed ground-based measurement from the CSIR. These metrics express the differences in units of Wh/m2.

$$MAD = \frac{1}{n}\sum_{i=0}^{n}|p_i - o_i| \quad (1)$$

$$MBD = \frac{1}{n}\sum_{i=0}^{n}(p_i - o_i) \quad (2)$$

$$RMSD = \sqrt{\frac{1}{n}\sum_{i=0}^{n}(p_i - o_i)^2} \quad (3)$$

The following equations normalise the metrics by the average of the observed values and multiply by 100 to express the differences as a percentage relative to the average observations.

$$nMAD = \frac{MAD}{\bar{o}} * 100 \quad (4)$$

$$nMBD = \frac{MBD}{\bar{o}} * 100 \quad (5)$$

$$nRMSD = \frac{RMSD}{\bar{o}} * 100 \quad (6)$$

4. Results and Discussion

In this work, we compared a series of 7-day forecasts from SolCast against the ground-based measurement data from the CSIR met station for the same forecast period using the metrics defined in the SolCast data validation report [1] to quantify forecast accuracy. We calculated the +1-hour accuracy metrics, the +1-day, and +7-day forecasts to determine a reasonable expectation for forecast accuracies that may be required from independent power producers operating in South Africa.

Figure 3 shows the SolCast half-hourly forecast data and the corresponding CSIR measured data for one week during the study period. This forecast dataset was downloaded from SolCast on 2024-05-20 in the morning, so the first forecast data point is only one hour into the future. In contrast, the last forecast data point is seven days into the future. The overall result for one week of GHI data plotted below indicates a MAD of 11.7 W/m2 for the hourly data, including nighttime forecasts. The MAD for daylight hours only over the same period is 24.2 W, approximately double the estimate when the nighttime values were included, simply because the daytime-only data is approximately one-half of the total observations in the dataset. The sum of the predicted GHI and the measured GHI for this period differ by 1373 Wh/m2, i.e. the forecasted insolation was 4.4% higher than the measured insolation. However, the hourly forecast error is likely minimised for this time of year in Tshwane during the dry season, when most days have clear skies, compared to the rainy season when the hourly irradiance is less predictable. No data from the rainy season was available for analysis, but the forecast errors are expected to be higher during the rainy season.

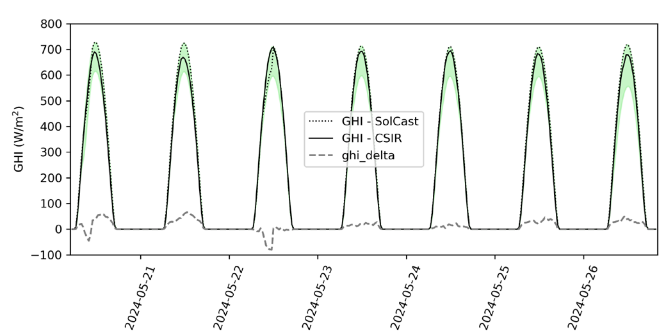

Figure 3: Half-hourly GHI predicted from SolCast (dotted line) versus measured by the CSIR (solid line) and the forecast interval from SolCast (green). The grey dashed line represents the difference between the forecast and the measurement.

Figure 4 shows the self-consistency across multiple forecasts for a given day in the study period. The points represent the forecasted GHI for a given hour from 20 different forecasts. The first forecast for the hourly GHI on August 13, 2024, shown in the graph, was made on August 8th at 19:00 hours. The green band represents the 80% forecast interval for each hour from SolCast. SolCast includes the 80% forecast interval with every forecast dataset, but only the forecast interval from August 7 at 19:00 hours was included in this graph. Most of the forecasted values from subsequent forecasts fall within the green band. In theory, 80% of the hourly forecasted points should fall within the green band.

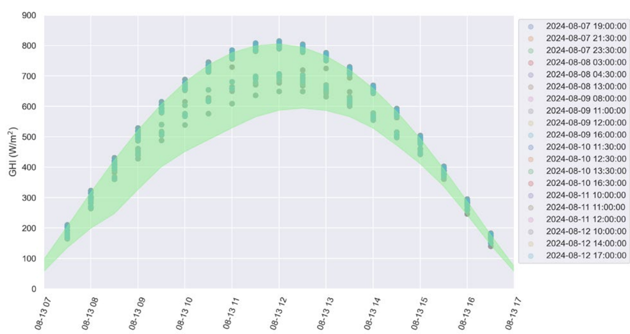

Figure 4 Forecast self-consistency for August 13, 2024, showing the forecasted data points for each hour from 20 different forecasts made between August 7 and August 12. The green band is the 80% forecast interval generated by the first (oldest) forecast.

Figure 5 shows the hourly insolation values from the +1-hour ahead SolCast forecast and the corresponding CSIR measured data over ten weeks. All the forecasts were downloaded every Monday or Tuesday morning from 20th May 2024 to 19 August 2024 between 7:30 AM and 12:00 PM, so the +1-hour ahead forecasts also correspond to morning values. The green band represents the SolCast upper and lower forecast intervals expressed in Wh/m2, and the grey reference lines were drawn every 50 Wh/m2 (10% relative to 500). The forecasts were quite precise in this timeframe, so the forecast intervals represented by the green band were narrow. The forecasts are within 10% of the measured values for most points in the graph. However, the +1-hour ahead forecast can be far off from the measured insolation by the mere passing of a stray cloud. For instance, the forecast for 2024-06-03 was off by approximately 70% because the SolCast forecast anticipated clear skies until 10:00 AM, but the clouds rolled in by 07:00 AM. Similarly, SolCast forecasted a clear sky all day on July 1st, but clouds rolled in at 11:00 AM, and the +1-hour forecast on that day was for 11:20 AM. Note that the SolCast forecasts were downloaded once weekly, so only one data point per week from the mornings contributes to this analysis. Ideally, the +1-hour ahead includes forecasts for all hours in the day over the study period.

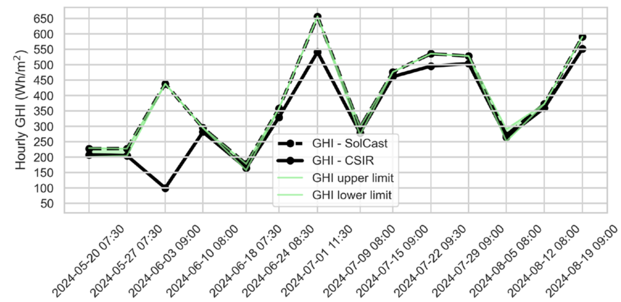

Figure 5: Hourly insolation values from the one hour ahead SolCast forecast and the corresponding CSIR measured data with a green band representing the SolCast upper and lower forecast interval

Figure 6 shows the daily insolation values from the +1-day ahead SolCast forecast and the corresponding CSIR measured data. The green band represents the daily sum of the SolCast upper and lower forecast intervals for each hour in Wh/m2. The grey reference lines represent a 400 Wh/m2 (10% relative to 4000) change. The forecast interval band is wider for the +1-day ahead forecast (15-45%) compared to the +1-hour ahead (< 5%) in this sample. Interestingly, the predicted value is often the same as the upper forecast limit, which is reasonable given the upper bound on the actual clear sky irradiance. On 2024-06-25, SolCast predicted a clear sky day from sunrise to sunset with a peak hourly insolation of 659 Wh/m2. In comparison, the measured data reflected overcast skies with a peak hourly insolation of 549 Wh/m2 at 1:30 p.m., resulting in a relatively large difference between the predicted and actual values.

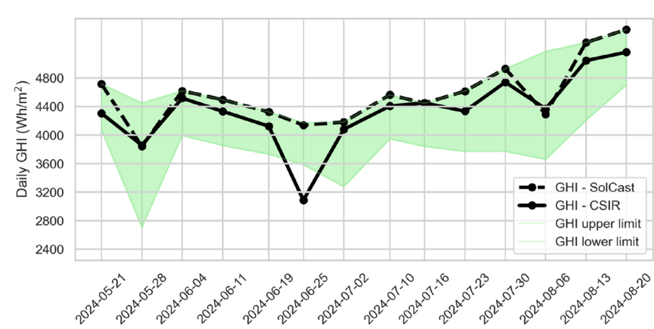

Figure 6: Daily insolation values from the +1-day ahead SolCast forecast and the corresponding CSIR measured data with a green band representing the SolCast upper and lower forecast interval summed up from the hourly values

Figure 7 shows the daily insolation values from the +7-day ahead SolCast forecast and the corresponding CSIR measured data. The green band represents the SolCast upper and lower forecast interval, and the grey lines represent a 400 Wh/m2 change in the value, or 10% relative to 4000 Wh/m2. The forecast interval band is wider for the +7-day ahead forecast (15-60%) than the +1-day ahead (15-45%). The one-day ahead forecasts and the +7-day ahead forecasts occur on different days, so a direct comparison of the points in Figure 6 and Figure 7 is not justified. On 2024-06-02, SolCast predicted a clear sky day from sunrise to sunset with a peak hourly insolation of 690 Wh/m2. The measured data reflected overcast skies with a peak hourly insolation of 622 Wh/m2 at 11:00 in the morning, resulting in a relatively large difference between the predicted and actual values. Note that two forecast values were excluded from the analysis because of missing data.

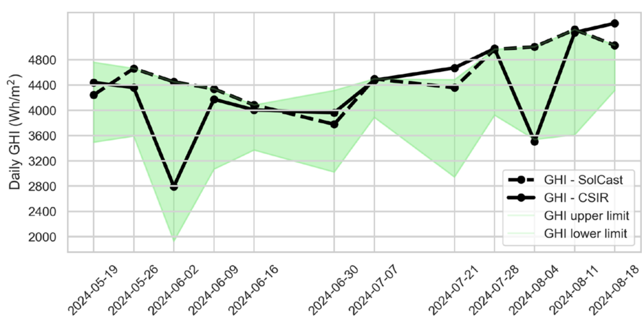

Figure 7: Daily insolation values from the +7-day ahead SolCast forecast and the corresponding CSIR measured data with a green band representing the SolCast upper and lower forecast interval

Table 1 shows the accuracy metrics for +1-hour, solar noon, +1-day, and +7-day ahead forecasts from SolCast compared to measured data from the CSIR weather station. The nRMSD for the +1-hour ahead forecast is 28.9%, which falls far outside the 16.6% average reported by the SolCast Team for daytime-only forecasts across all 33 sites. The nRMSD reported in the SolCast study varied from 6% to 36.6% across the 33 sites, and the estimate for the CSIR falls just inside that reported range. We speculate that the higher nRMSD reported for the CSIR site is due to the limited dataset gathered exclusively from morning hours in the winter season. All the +1-hour data is based on forecasts during the mornings over the study period rather than all hours of the day due to the manual data collection plan. If the forecasts for all daylight hours were included, the average of the measured values would be higher, and the calculated RMSD normalised by the average value (nRMSD) would be smaller because it would be normalised by a higher mean value, assuming the differences in hourly forecasts remained relatively constant. To test this point, Table 1 also includes metrics for the forecasts at solar noon on the same day the forecast data was downloaded. These forecasts predict irradiance 3 to 5 hours into the future, yet they have a smaller nRMSD compared to the +1-hour ahead forecasts. Specifically, the nRMSD for the solar noon forecasts is 11.3%, below the average of 16.6% nRMSD reported by the SolCast Team, and well below the 28.9% from the +1-hour forecasts for morning hours.

Table 1 Accuracy metrics for SolCast forecasts compared to measured data from the CSIR weather station

Metric	+1-hour ahead	Solar Noon	+1-day ahead	+7-day ahead	Units
Data aggregation	Average over one hour		Sum over 24 hours		N/A
Count	14	14	14	12	points
Average of measured	338.8	671.3	4 343.8	4 331.4	Wh/m^2
MBD	50.4	35.3	223.8	226.2	Wh/m^2
MAD	50.4	50.7	237.3	402.0	Wh/m^2
RMSD	97.9	75.6	345.8	671.6	Wh/m^2
nMBD	14.9	5.3	5.2	5.2	%
nMAD	14.9	7.6	5.5	9.3	%
nRMSD	28.9	11.3	8.0	15.5	%

The +1-hour ahead forecast might assist grid operators in planning for meeting imminent peak demand. Unfortunately, peak demand typically occurs in the mornings and evenings when PV generators produce relatively little electricity, and the accuracy of forecasts for morning hours is not great (nRMSE = 28.9%). Forecasting +1-hour ahead in the morning and evening hours is associated with a high degree of error.

By contrast, the +1-day and +7-day ahead forecasts for daily insolation might be more beneficial to grid operators when planning short-term maintenance projects. The nMBD indicates the forecasts for daily GHI values are 5.2% higher than observed for both the +1-day and +7-day ahead forecasts, on average. The bias of the forecast is also an important consideration. Grid operators may prefer to underestimate future production rather than overestimate it to reduce the risk of unmet demand. The nMBD reported in Table 1 indicates that the forecasted GHI is higher than the measured GHI for both forecast horizons. Based on this analysis, the forecast accuracy for daily insolation, as reflected in the nMBD, is approximately 5-6%, meaning the grid operator can forecast the available solar resource, i.e., available fuel, to a reasonable degree for a given 24-hour period in the future.

5. Conclusion

This study compared forecasted GHI from SolCast to GHI recorded at the CSIR meteorological station in Tshwane, South Africa. We compared the SolCast forecast and the CSIR measured values and computed accuracy metrics for the +1-hour, +1-day, and +7-day ahead forecasts. The +1-hour ahead forecast was based on the average value for a single hour, while the +1-day and +7-day ahead were based on the insolation over 24 hours. Based on the limited data collected for this study, the normalised root mean squared deviation (nRMSD) for the +1-hour ahead forecast from morning forecasts during the winter season is 28.9% at the CSIR, which is nearly double the average nRMSD reported in a study conducted by the SolCast Team on forecasts across 33 sites worldwide but still within the reported range across all sites. The normalised mean bias deviations (nMBD) for the daily insolation for +1-day and the +7-day ahead forecasts were 5.2%. We conclude that the nRMSD for +1-hour ahead forecast of hourly irradiance is too large to be of practical use for grid planning. However, the nMBD for the forecasts of the +1-day and +7-day ahead daily insolation is sufficiently small to be of practical use for predicting electricity production from PV generators in the future.

Acknowledgements (Heading 5 style)

We thank the CSIR for supporting this research and SolCast for permitting free access to a limited dataset for academic research.

References

[1] "DNV__Review_and_validation_of_Solcast_Historical_and_TMY_data.pdf." 2024. [Online]. Available: https://www.dnv.com/Publications/review-and-validation-of-solcast-historical-and-tmy-data

[2] B. Mabasa, M. D. Lysko, and S. J. Moloi, "Comparison of Satellite-Based and Ångström–Prescott Estimated Global Horizontal Irradiance under Different Cloud Cover Conditions in South African Locations," *Solar*, vol. 2, no. 3, pp. 354–374, Aug. 2022, doi: 10.3390/solar2030021.

[3] B. Mabasa, M. D. Lysko, and S. J. Moloi, "Validating Hourly Satellite Based and Reanalysis Based Global Horizontal Irradiance Datasets over South Africa," *Geomatics*, vol. 1, no. 4, pp. 429–449, Nov. 2021, doi: 10.3390/geomatics1040025.

[4] The SolCast Team, SolCast Live and Forecast Accuracy. Accessed: Jul. 26, 2024. [Online]. Available: https://solcast.com/forecast-accuracy