# Probabilistic forecasting of solar radiation

Dr Adrian Grantham

School of Information Technology and Mathematical Sciences
School of Engineering

7 September 2017

University of South Australia

# Acknowledgements

Funding:

Collaborators:

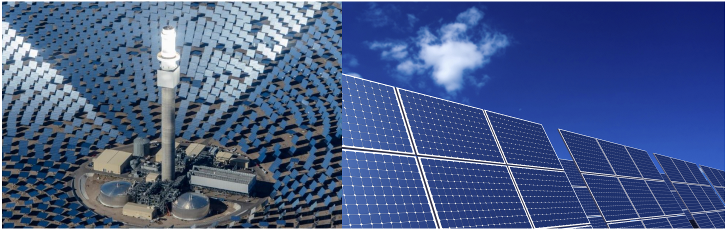
- Professor John Boland (UniSA)
- Associate Professor Yulia Gel (University of Texas)

# Motivation

If we are supplying energy into the electricity grid, it is important that we have knowledge of the expected output from the farm to the grid.


# Motivation

If we are supplying energy into the electricity grid, it is important that we have knowledge of the expected output from the farm to the grid.



A *probabilistic forecast*:

- provides information about all expected outcomes
- allows one to both assess a wide range of uncertainties and facilitate **decision making**.

# Overview

Our approach to developing a probabilistic forecast:

1. develop a point forecast
2. develop a probabilistic forecast (prediction intervals).

# Overview

Our approach to developing a probabilistic forecast:

1. develop a point forecast
2. develop a probabilistic forecast (prediction intervals).

Today we will look at:

1. p1i1 methods
2. p2i2 methods
3. results and performance
4. synthetic sequences of solar radiation.

# Data

Locations and Köppen-Geige climate classification system:

- Adelaide: Mediterranean
- Darwin: tropical
- Mildura: semi-arid.

Each data set consists of 10 years of hourly global horizontal irradiation (GHI) values (8 years in-sample, 2 years out-of-sample).

# Point forecast i1

The hourly global horizontal irradiation (GHI) $I_t$ for Mildura is given as

$$I_t = F_t + A_t + Z_t,$$

where $F_t$ is a seasonal component, $A_t$ is an autoregressive component (a linear combination of previous time steps), and $Z_t$ is a noise such that $EZ_t = 0$, $EZ_t Z_l = 0$ if $t \neq l$ and $EZ_t^2 = \sigma_t^2$. That is, $Z_t$ may be heteroscedastic.

# Point forecast i1

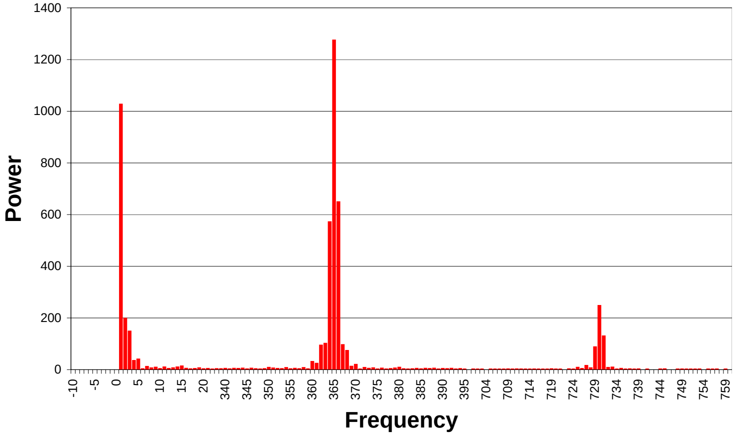



Figure 1: Power spectrum of hourly global horizontal irradiation (GHI) for Mildura. The last three significant frequencies are not shown.

# Point forecast i1

The Fourier component $F_t$ of $I_t$ is given as

$$
\begin{aligned}
F_t \quad = \quad & \alpha_0 + \alpha_1 \times \cos\frac{2\pi t}{8760} + \beta_1 \times \sin\frac{2\pi t}{8760} + \\
& \alpha_2 \times \cos\frac{4\pi t}{8760} + \beta_2 \times \sin\frac{4\pi t}{8760} + \\
& \sum_{i=3}^{11}\sum_{n=1}^{3}\sum_{m=-1}^{1}[\alpha_i \times \cos\frac{2\pi(365n+m)t}{8760} + \\
& \beta_i \times \sin\frac{2\pi(365n+m)t}{8760}]
\end{aligned}
$$

Note that in examples we have tested, the amount of the variance explained by the Fourier Series is approximately 80-85%.

# Point forecast i1

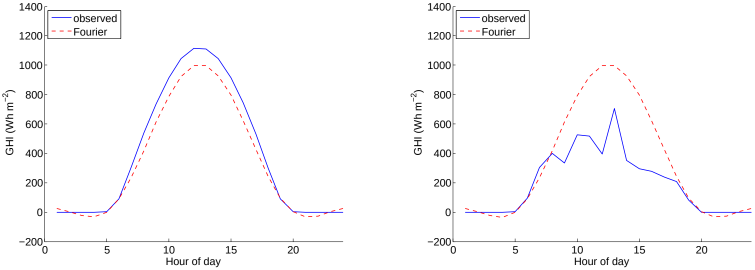



Figure 2: Mildura: observed vs. Fourier global horizontal irradiation (GHI) for a clear sky day on January 2, 2003 (left panel) and a cloudy day on January 1, 2003 (right panel).

# Point forecast i1

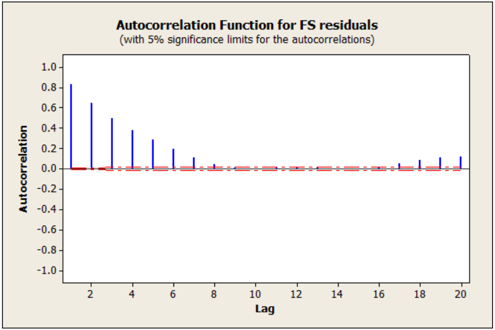



Figure 3: Plotted autocorrelation function (ACF) of the global horizontal irradiation (GHI) deseasoned residuals of Mildura, with 5% significance limits shown in red.

# Point forecast i1

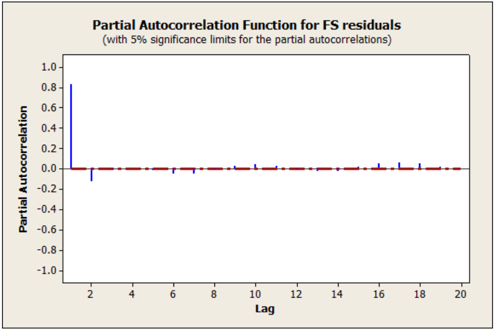



Figure 4: Plotted partial autocorrelation function (PACF) of the global horizontal irradiation (GHI) deseasoned residuals of Mildura, with 5% significance limits shown in red.

# Point forecast i1

The ACF decaying slowly, while the PACF has significant spikes at lags one and two indicating an autoregressive model of order 2, AR(2).

However, after trying to overfit the model with an AR(3) process, it is found that the AR(3) process showed slightly better performance and the *p*-values for all three coefficients are significant at the 5% level.

# Point forecast i1

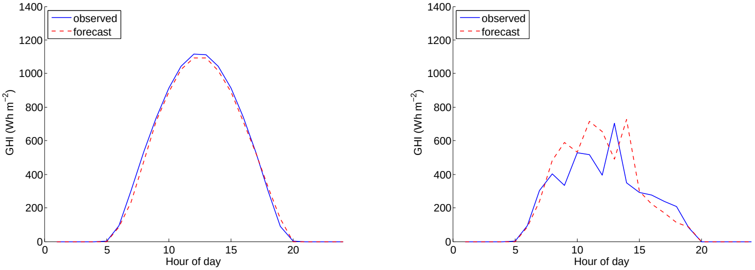



Figure 5: Observed vs. forecast global horizontal irradiation (GHI) for a clear sky day on January 2, 2003 (left panel) and a cloudy day on January 1, 2003 (right panel).

# Probabilistic forecast i1

We assume the hourly errors are heteroscedastic and that this is driven by a specific sun position.
So we place the hourly daytime errors into a 2-dimensional matrix according to sun elevation and sun hour angle.
We do this to take care of the systematic variation in variance in the GHI time series.

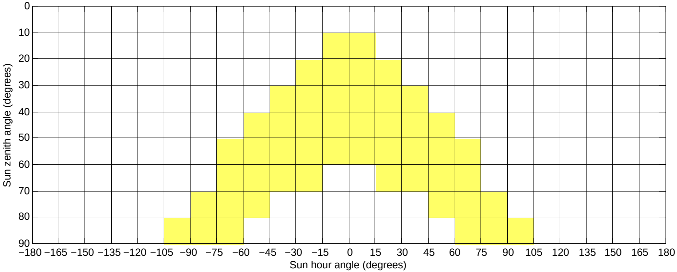



Figure 6: Sun map: $Z_t$ binning matrix

# Probabilistic forecast i1

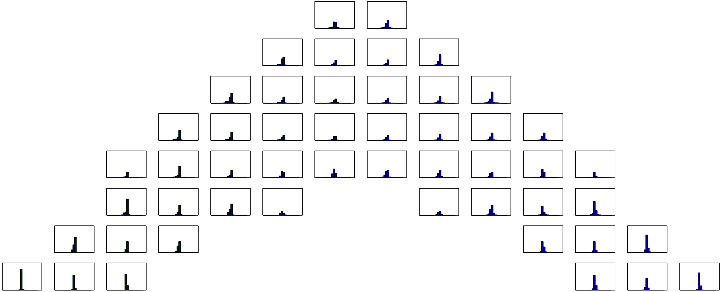


Figure 7: Histograms of errors in each bin of the two-dimensional array, in respect to sun position (as described above), for Mildura.

# Probabilistic forecast i1

**Algorithm 1:** Algorithm for generating (100-$\alpha$) prediction intervals using the simplified method.

**Data:** Out-of-sample hourly daytime forecasting model $I_t$ with length $N$, and the two-dimensional array of errors binned according to sun position $B_{i,j}$.

**for** $t = 1, \ldots, N$ **do**
    calculate sun elevation index $i$ according to sun elevation for $I_t$;
    calculate sun hour angle index $j$ according to sun hour angle for $I_t$;
    calculate the lower $\alpha/2$ percentile $B_{i,j}^{\alpha/2}$ from bin $B_{i,j}$;
    calculate the upper $100 - \alpha/2$ percentile $B_{i,j}^{100-\alpha/2}$ from bin $B_{i,j}$;
    generate lower prediction interval $L_t^{100-\alpha} = \hat{I}_t + B_{i,j}^{\alpha/2}$;
    generate upper prediction interval $U_t^{100-\alpha} = \hat{I}_t + B_{i,j}^{100-\alpha/2}$;
**end**

**Result:** Out-of-sample hourly daytime (100-$\alpha$) upper and lower prediction interval, $L_t^{100-\alpha}$ and $U_t^{100-\alpha}$ respectively .

# p1i1

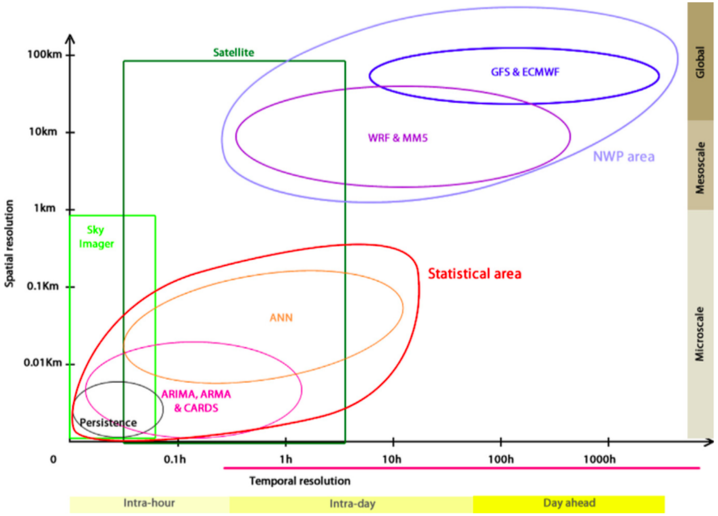
A. Grantham, Y. R. Gel, and J. Boland, “Nonparametric short-term probabilistic forecasting for solar radiation,” Sol. Energy, vol. 133, pp. 465–475, 2016.

# New work: p2i2

We make improvements to the:

- point forecast $\rightarrow$ i2
- probabilistic forecast $\rightarrow$ p2

# Point forecast i2



# Point forecast i2

The new point forecasting method combines perfect knowledge of the day-ahead daily solar radiation with our previous point forecasting method. Obviously perfect knowledge of the day-ahead daily solar radiation is not feasible.

Ideally we would prefer to use a day-ahead daily forecast from a numerical weather prediction (NWP) model because a daily NWP forecast is known to be very accurate. However, a NWP is unavailable at this time. Instead we use perfect knowledge of the day-ahead daily value as a proxy.

The idea here is to demonstrate the potential performance improvements of combining a daily NWP with an hourly solar radiation forecast, generated from our statistical model. Statistical methods perform better at hourly time scales and NWP methods perform better at daily time scales.

# Point forecast i2

In order to combine the hourly point forecasting with the known day-ahead (or a NWP day-ahead), we take the hourly Fourier $F_t$ component for the day-ahead and scale it so that the daily sum of the scaled $F_t$ matches the known day-ahead forecast.

# Probabilistic forecast i2

i1: global systematic variation in variance of solar radiation.

# Probabilistic forecast i2

i1: global systematic variation in variance of solar radiation.

i2: we look at conditional heteroscedasticity (change in variance).

The final errors are uncorrelated but dependent - the squared error terms, a proxy for variance, are correlated. This is the so-called ARCH effect- autoregressive conditional heteroscedastic. Usually when this happens one uses an ARCH or GARCH model for forecasting the variance.

# Probabilistic forecast i2

However we found that instead an exponential smoothing form was more useful.

$$S_{t+1} = \alpha Z_t^2 + (1-\alpha)S_t, \quad 0 < \alpha < 1, t \geq 2.$$

with $S_2 = Z_1^2$.

Since we are forecasting the variance, and then constructing a prediction interval using this, we have to perform this assuming the noise is normally distributed, which is not true.

So, we first had to use a normalising transformation, then forecast the variance, construct the PIs, and then transform back.

# Probabilistic forecast i2

Conditional heteroscedastic probabilistic forecast method:

1. Find $F(Z_t, i, j)$, the cumulative probability of $Z_t$ for bin $i, j$.
2. Transform $Z_t$ according to $\gamma_t = F^{-1}(Z_t, i, j)$, with $\gamma_t \sim N(0, 1)$. Note that this is done each time according to the bin currently referenced.
3. Find the Exponential Smoothing forecast model $\psi_t^2 = \alpha\gamma_t^2 + (1-\alpha)\psi_t^2$, with $0 < \alpha < 1$.
4. Apply the forecast model to get prediction intervals for $\sigma_t$. For instance, for a 95% PI, use $\hat{\sigma}_t \pm 1.96\psi_t$
5. Apply the inverse transform to take these limits of the prediction intervals back to the equivalent values for $\hat{R}_t$. Note once again that one has to do this with reference to the particular bins according to time of day and solar elevation.
6. Add the Fourier Series Representation to all to get forecast plus bounds for the original data.

# Preformance metrics: point forecast

Table 1: Point forecast: normalised root mean square error (NRMSE) (%)

| Point forecast | Original (p1) | Known day ahead (p2) |
|---|---|---|
| Adelaide | 19.14 | 15.56 |
| Darwin | 22.75 | 19.16 |
| Mildura | 15.29 | 12.34 |

# Preformance metrics: point forecast

Table 2: Point forecast: mean bias error (MBE) (%)

| Point forecast | Original (p1) | Known day ahead (p2) |
|---|---|---|
| Adelaide | 0.72 | 0.39 |
| Darwin | 0.85 | 0.52 |
| Mildura | 1.32 | 0.57 |

# Preformance metrics: point forecast

Table 3: Point forecast: mean absolute error (MAE) (%)

| Point forecast | Original (p1) | Known day ahead (p2) |
|---|---|---|
| Adelaide | 13.25 | 10.57 |
| Darwin | 15.86 | 13.22 |
| Mildura | 10.83 | 8.43 |

# Performance metrics: probabilistic forecast

Prediction interval coverage probability (PICP) for a given confidence level $\alpha$:

$$PICP = \frac{1}{L}\sum_{t=1}^{L} C_t,$$

where $L$ is the total number of forecasts and

$$C_t = \begin{cases} 1, & L_t^{100-\alpha} \leq I_{t+1} \leq U_t^{100-\alpha}, \\ 0, & \text{otherwise.} \end{cases}$$

Because coverage is easily obtained by having wider PI widths, we use the normalised averaged width (PINAW):

$$PINAW = \frac{1}{L I_{max}} \sum_{t=1}^{L} (U_t^{100-\alpha} - L_t^{100-\alpha}),$$

where $I_{max} = 1000\ Wm^{-2}$.

# Performance metrics: probabilistic forecast

The *coverage width-based criterion* (CWC) metric quantifies the trade-off between coverage and prediction interval width

$$CWC = PINAW(1 + \gamma(PICP)e^{-\mu((PICP)-\alpha)}),$$

where $\mu = 10$ and

$$\gamma = \begin{cases} 0, & PICP \geq /alpha, \\ 1, & PICP < alpha. \end{cases}$$

# Performance metrics: probabilistic forecast

The *coverage width-based criterion* (CWC) metric quantifies the trade-off between coverage and prediction interval width

$$CWC = PINAW(1 + \gamma(PICP)e^{-\mu((PICP)-\alpha)}),$$

where $\mu = 10$ and

$$\gamma = \begin{cases} 0, & PICP \geq /alpha, \\ 1, & PICP < alpha. \end{cases}$$

BUT this doesn't penalise over-coverage the same as under-coverage. So we use

$$CWC = PINAW(1 + (PICP)e^{-\mu|(PICP)-\alpha|}),$$

and treat over-coverage and under-coverage equally.

# Performance metrics: probabilistic forecast

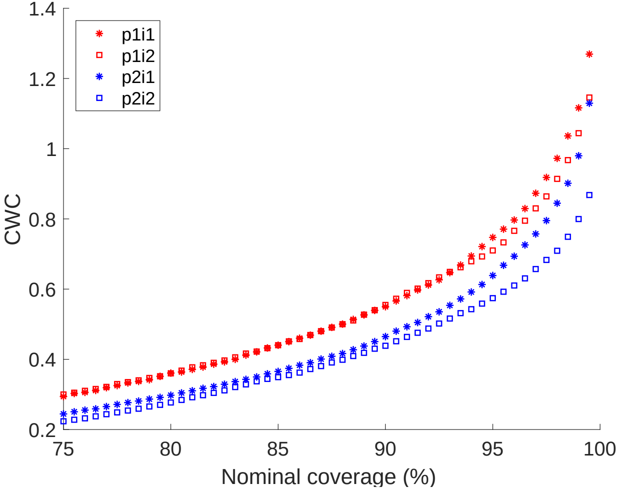



Figure 8: *Adelaide: coverage width-based criterion* (CWC)

# Performance metrics: probabilistic forecast

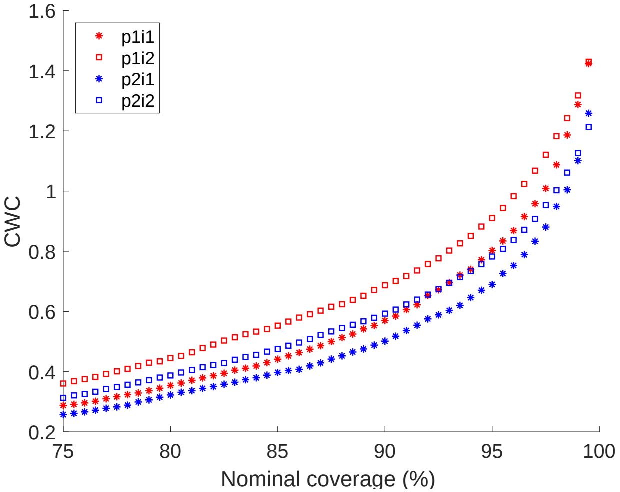



Figure 9: Darwin: coverage width-based criterion (CWC)

# Probabilistic forecasts

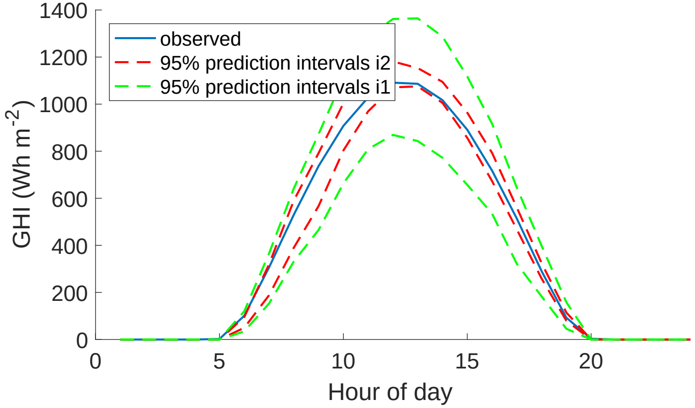



Figure 10: Adelaide: 95% probabilistic forecast for a clear sky

# Probabilistic forecasts

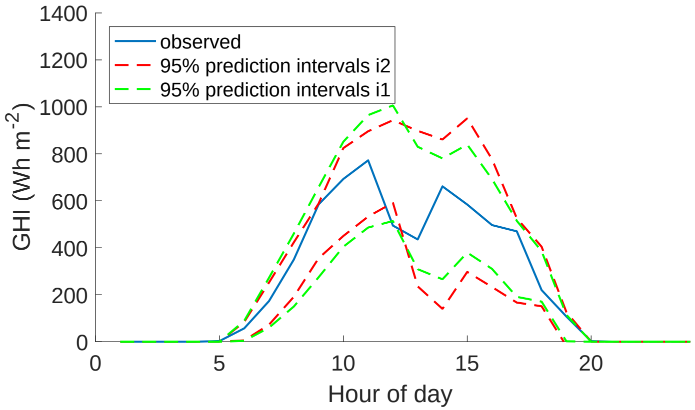



Figure 11: Adelaide: 95% probabilistic forecast for a cloudy sky

# Probabilistic forecasts

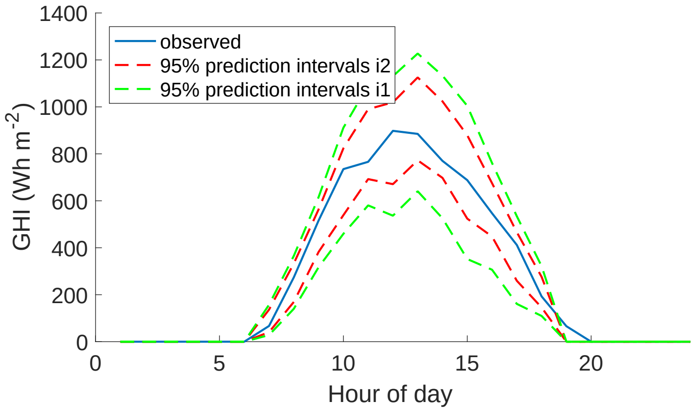



Figure 12: Darwin: 95% probabilistic forecast for a clear sky

# Probabilistic forecasts

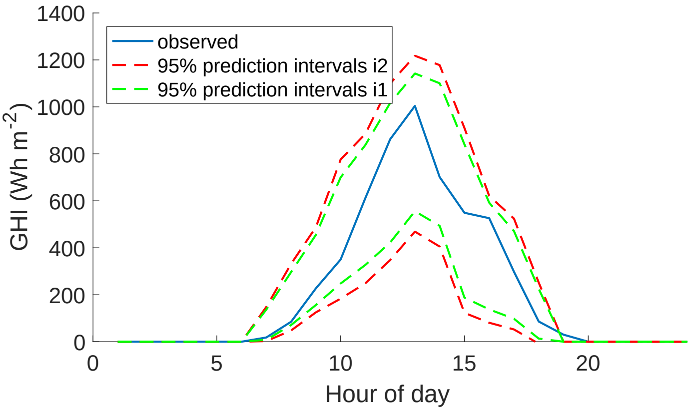



Figure 13: Darwin: 95% probabilistic forecast for a cloudy sky

# Probabilistic forecasting conclusions

Results are mixed:

- Adelaide and Mildura: i2 is better than i1
- Darwin: i1 is better than i2

These results suggest that the ideal probabilistic forecasting method might be climate specific.

Coming soon: incorporate a *n*-step ahead hourly numerical weather prediction (NWP) forecast into our one-step-ahead hourly point forecast.

# Synthetic sequences of GHI

The hourly global horizontal irradiation (GHI) $I_t$ for Mildura is given as

$$I_t = F_t + A_t + Z_t,$$

# Synthetic sequences of GHI

The hourly global horizontal irradiation (GHI) $I_t$ for Mildura is given as

$$I_t = F_t + A_t + Z_t,$$

The procedure for generating synthetic hourly GHI is

$$\widetilde{I}_t = F_t + \phi_1(F_{t-1} - \widetilde{I}_{t-1}) + \phi_2(F_{t-2} - \widetilde{I}_{t-2}) + \phi_3(F_{t-3} - \widetilde{I}_{t-3}) + Z_t,$$

where $\widetilde{I}_t$ is the synthetic hourly GHI for time point $t$ (in hours), $F_t$ is the Fourier component and $\phi_1$, $\phi_2$ and $\phi_3$ are the coefficients for the autoregressive order three (AR(3)) component. The procedure follows the hourly model but includes a bootstrapped white noise term $Z_t$. The white noise term $Z_t$ is bootstrapped (sampled with replacement) from $B_{i,j}$ (two-dimensional matrix of errors), according to the corresponding sun elevation and hour angle.

# Synthetic sequences of GHI

The synthetic hourly GHI will include patterns of GHI that have not occurred in the recorded data but are nonetheless equally as likely to occur. That is, the synthetic sequences have the same statistical properties as the observed:

- the same underlying hourly distributions
- the same serial hourly correlation structure
- the same underlying daily sum distribution
- the same daily serial correlation structure.

# Synthetic sequences of GHI

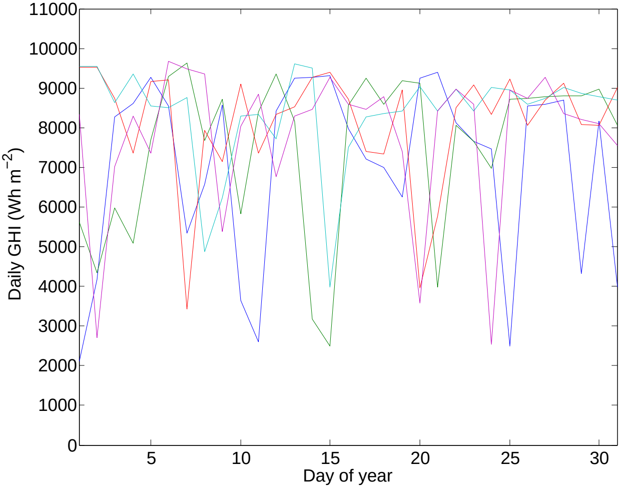



Figure 14: Example of five synthetic daily global horizontal irradiation (GHI) estimates for the month of January.

# Synthetic sequences of GHI

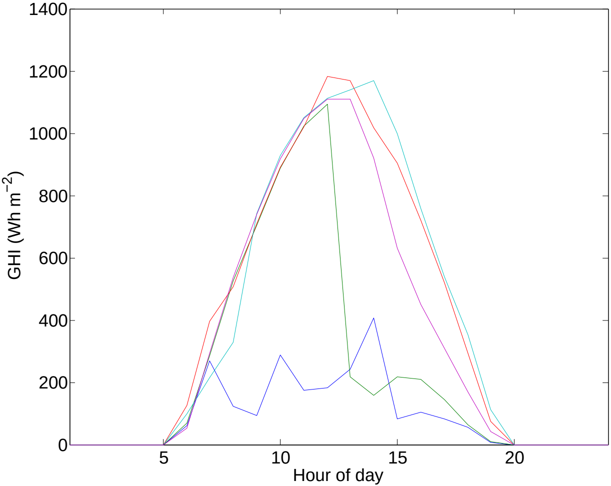



Figure 15: Example of five synthetic hourly global horizontal irradiation (GHI) estimates for January 1.

# Synthetic sequences of GHI

Table 4: Frequency distribution of consecutive days in 100 years with daily global horizontal irradiation (GHI) below 2,000 Wh $m^{-2}$ using the eight years of observed daily GHI, 1995–2002, and the 1,000 synthetic daily GHI $\widehat{I_t^*}$ years (instances), for Mildura.

| Consecutive days | Observed | Synthetic |
|---|---|---|
| 1 | 1512.5 | 1596.5 |
| 2 | 362.5 | 367.0 |
| 3 | 112.5 | 105.4 |
| 4 | 50.0 | 32.7 |
| 5 | 12.5 | 10.5 |
| 6 | 0.0 | 3.6 |
| 7 | 0.0 | 1.1 |
| 8 | 0.0 | 1.0 |
| 9 | 0.0 | 0.1 |

# Synthetic sequences of GHI conclusions

The synthetic data can be used, for example, as input for testing the performance and operation of a solar energy system for all supply scenarios resulting in the design of a more reliable system.

This work is part of a bigger endeavour, where we start with synthetic yearly, then synthetic seasons from that and then daily, then hourly and then 5 minute or minute, all in a chain.