Simulation of multi-annual time series of solar photovoltaic power
Ramirez Camargo, Luis; Schmidt, Johannes

Published in:
SUSTAINABLE ENERGY TECHNOLOGIES AND ASSESSMENTS

DOI:
10.1016/j.seta.2020.100829

Publication date:
2020

License:
CC BY

Document Version:
Final published version

Citation for published version (APA):
Ramirez Camargo, L., & Schmidt, J. (2020). Simulation of multi-annual time series of solar photovoltaic power: Is the ERA5-land reanalysis the next big step? SUSTAINABLE ENERGY TECHNOLOGIES AND ASSESSMENTS, 42, [100829]. https://doi.org/10.1016/j.seta.2020.100829
Simulation of multi-annual time series of solar photovoltaic power: Is the ERA5-land reanalysis the next big step?

Luis Ramirez Camargo, a,b,c,⁎, Johannes Schmidt a

⁎ Corresponding author at: Institute for Sustainable Economic Development, University of Natural Resources and Life Sciences, Vienna, Austria.
E-mail address: luis.ramirez-camargo@boku.ac.at (L. Ramirez Camargo).

ARTICLE INFO

Keywords: ERA5-land MERRA-2 Photovoltaics Renewable energy Open data

ABSTRACT

The simulation of multi-annual time series of photovoltaic electricity generation in high temporal resolution using reanalysis data has become a common approach. These time series are crucial to assess the viability of electricity systems with high shares of variable renewable generation. Our work combines the new ERA5-land reanalysis data set and PV_LIB to generate hourly time series of photovoltaic electricity generation for several years and validates the results using individual data of 23 large photovoltaic plants located in Chile. We use a clustering algorithm to differentiate between fixed and tracking systems, as meta-information on installation type was not available. Results are compared with photovoltaic output for these locations calculated using MERRA-2, a global reanalysis with five times lower spatial resolution, which is one established source for modelling photovoltaic generation time series. Accuracy and bias indicators are satisfactory for all plants, i.e. correlations are above 0.75 for all installations and above 0.9 for more than half of them, while the mean bias error is between -0.05 and 0.1 for all installations. However, the improvements in simulation quality over results obtained with MERRA-2 are minor. From our assessment of generation data quality, we conclude that efforts towards availability and standardization of data of individual installations are necessary to improve the basis for future validation studies.

Introduction

Solar photovoltaic systems (PV) play a major role in the renewable energy transition taking place around the globe. E.g. in Germany, an early deployer of PV, 20% (45 GW) of installed electricity generation capacity was PV in 2018 [1]. Newcomers to the renewable energy transition are catching up quickly. One example is Chile where more than 2.6 GW of PV have been installed since 2013, representing 11% of the total installed electricity generation capacity. More than 2 GW are currently under construction and environmental permissions for an additional 17.7 GW have been granted until the beginning of 2020 [2]. At the global scale the total installed capacity grew by around 115 GW in one year to reach a total of 627 GW at the end of 2019 [3].

There is considerable less work on the simulation of multi-annual time series of photovoltaic electricity generation in high temporal resolution using reanalysis data has become a common approach. These time series are crucial to assess the viability of electricity systems with high shares of variable renewable generation. Our work combines the new ERA5-land reanalysis data set and PV_LIB to generate hourly time series of photovoltaic electricity generation for several years and validates the results using individual data of 23 large photovoltaic plants located in Chile. We use a clustering algorithm to differentiate between fixed and tracking systems, as meta-information on installation type was not available. Results are compared with photovoltaic output for these locations calculated using MERRA-2, a global reanalysis with five times lower spatial resolution, which is one established source for modelling photovoltaic generation time series. Accuracy and bias indicators are satisfactory for all plants, i.e. correlations are above 0.75 for all installations and above 0.9 for more than half of them, while the mean bias error is between -0.05 and 0.1 for all installations. However, the improvements in simulation quality over results obtained with MERRA-2 are minor. From our assessment of generation data quality, we conclude that efforts towards availability and standardization of data of individual installations are necessary to improve the basis for future validation studies.

There is considerably less work on the simulation of multi-annual time series of PV generation available. These time series are fundamental for modelling the transition to renewable energies. Private companies such as solargis [7] and VAISALA [8] offer high resolution time series of solar radiation and PV output estimations for particular locations but at costs that would be difficult to cover by research projects interested in data for thousands of locations or large-scale systems analysis. Pfenninger and Staffell [9] addressed this issue with the renewables.ninja platform. This platform allows to freely generate multi-annual time series of hourly PV output for any location in the world.
The underlying solar radiation data, the Modern-Era Retrospective Analysis for Research and Applications (MERRA) [10], the Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2) [11] and the Surface Radiation Data Set – Heliosat (SARAH) [12], have been validated extensively in literature and are in use in numerous studies, see e.g. [13–15]. There is also a growing body of literature making use of the renewables.ninja data but the output of individual PV installations estimated there has only been validated by Pfenninger and Staffell [9] themselves. This restricts the geographical coverage of the validations to Europe. Similarly, the PVGIS platform of the Joint research centre of the European Commission [16] in its version 5 allows to generate hourly time series for up to 12 years (2005–2016) for locations in most parts of the world [17]. The underlying solar radiation data sets, that include data derived from the Climate Monitoring Satellite Application Facility (CM-SAF) [18], the regional reanalysis COSMO-REA6 [19], the US National Solar Radiation Database (NSRDB) [20] and the global reanalysis ERA5 [21], have been validated multiple times and a classification of which data set works better for which locations is also available [22,23]. In contrast to the solar radiation data used in PVGIS, to the best knowledge of the authors, validations of the derived time series of PV output of this platform have not been performed.

While satellite derived solar radiation data were shown to be usually more accurate than reanalysis data sets of previous generations, the accuracy of state of the art regional reanalysis data sets such as COSMO-REA6 is coming closer to the one of their satellite derived counterparts [23,24] and in particular the last global reanalysis of the European Centre for Medium-Range Weather Forecasts (ECMWF), ERA5, presents promising results compared to its predecessors. Urraca et al. [23] showed that ERA5 solar radiation data have an average bias on the global scale that is 50–75% lower compared to ERA-interim and MERRA-2. Huang et al. [25] found that ERA5 performs relatively robustly across Australia without notable deficiency and is more accurate than the Global Forecast System (GFS) and the Australian Community Climate and Earth-System Simulator (ACCESS). Trolliet et al. [26] compared five data sets including MERRA-2 and ERA5 for irradiance estimation in the tropical Atlantic ocean and stated that while the reanalysis data sets are not the best performing among all available data sets, ERA5 presented consistently higher correlations than MERRA-2, i.e. correlation coefficients greater than 0.85 for MERRA-2 and 0.89 for ERA5. Overestimation of global horizontal irradiation (GHI) by ERA5 has been, however, reported for 98 sites in China [27] and multiple locations in Norway [28]. Further low performance of ERA5 was reported for direct normal radiation in a location in Brazil, when comparing it to other 10 solar radiation data sets, where ERA5 scored a root mean square error (RMSE) of 266.9 W/m² (63.4%) but none of the other data sets in the comparison reached better values than 154.3 W/m² (37.5%) [29]. However, concerning GHI, ERA5 performed in the midfield of the evaluated data sets for the same location [29], did not have considerable lower performance than data from CAMS, CM-SAF and SARAH for a location in Methoni, in southwest Peloponessse, Greece [30], had similar bias compared to that of satellite data for inland regions with few clouds globally [23] and presented promising results as a complement for satellite-based databases in regions not covered by geostationary satellites when simulating PV systems [22].

Regarding PV output, Atencio Espejo et al. [31] present a comparison between ERA5 and PV_LIB derived PV power generation time series and measured data for one location in Milano, Italy for 2014–2016. They find correlations around 0.91 and a normalized RMSE around 11% when comparing calculated against measured PV output.

The satisfactory accuracy results have already been used as justification in multiple studies that employ ERA5 data as input to calculate PV output time series. These studies include an assessment of synergies of solar PV and wind power potential in West Africa at hourly resolution [32], an assessment of on-site steady electricity generation from hybrid (PV, wind power and battery) renewable energy systems for the entire territory of Chile [33], and the mapping at the global scale of degradation mechanisms as well as total degradation rates for a monocrystalline silicon PV module [34]. However, none of these studies presented a validation of the PV output data.

In July 2019, the Copernicus Climate Change Service (C3S) launched the ERA5-land [35] data set of the ECMWF, derived from ERA5, with a spatial resolution of 9 km × 9 km, which is more than three times and five times higher than the resolution of ERA5 and MERRA-2, respectively. To the best of our knowledge, no validations of the radiation variables in this data set or any validation of PV output derived from them have been performed so far. Considering the progress in the accuracy of radiation variables in the global reanalysis data sets, it could be expected that such high resolution data set allows the estimation of PV output time series that are considerably more accurate than calculations relying on previous global reanalysis generations. In order to test this, the present study proposes the calculation of PV output time series using radiation, temperature and wind speed data from ERA5-land together with the widely used PV_LIB library [36]. Moreover, it compares these estimates to PV generation data derived from MERRA-2 with PV_LIB as well as to measured data from the locations of large PV installations in Chile. The comparison is performed using hourly, daily and monthly capacity factors as well as typically used indicators such as the mean bias error (MBE), the Pearson correlation coefficient and RMSE. The selection of locations in Chile is related not only to the massive expansion of PV generation in this country but also to the fact that data of all large PV installations connected to the grid are open and freely available online through official sources. An additional particularity of this country is that most part of its territory is not reached by the Meteosat Second Generation (MSG) geostationary satellite, which is the main source of most of the solar radiation data sets that are the usual benchmark for solar radiation reanalysis data sets, also in use in renewables.ninja and PVGIS.

Data and methodology

Data sets and pre-processing

Four data sets are used to test the proposed hypothesis. These include measured data from PV installations in Chile, meta-information about the installations including locations, and the ERA5-land and MERRA-2 variables necessary for PV output estimation (Table 1). All of them are available openly and freely for academic uses. The time series have an hourly temporal resolution. Details of the data and the necessary pre-processing are presented in Section 2.1.1. and Section 2.1.2.

Table 1

| Data set                         | Provider                                      | Spatial Resolution of data set                        |
|----------------------------------|-----------------------------------------------|------------------------------------------------------|
| meta-information on PV installations | Ministry of Energy of the Chilean Government [37] | Point information (latitude and longitude)           |
| Electricity output of PV Chilean installations | Energía abierta Chile [38]                      | –                                                    |
| ERA5-land                        | ECMWF [35]                                    | Grid of 0.1° – 9 km                                   |
| MERRA-2                          | NASA [39]                                     | Grid of 0.5° latitude × 0.625° longitude – 50 km     |
Additionally, all values beyond 1.1 times the 99 percentile were classified as outliers and excluded from further calculations. Similarly, only values larger than 0 were preserved for the validation in order to avoid the inclusion of periods where plants were entirely offline as well as night periods, when prediction is trivial. A summary of the main characteristics of the measured data is provided in Table A1 in the appendix. Additionally the processed data set with the capacity factors is available under [43].

We then classified the installations in two steps: first, we identified installations with low data quality. Subsequently, we applied a clustering algorithm to the generation time series of the remaining installations to differentiate between systems with and without tracking. For the first step, i.e. for identifying erroneous data sets, we removed hours with no production from the data-set (this includes non-measured hours, but also night hours). We then calculated the longest series of consecutive hours, when generation was below 50% of average annual capacity factors and 70% below average monthly capacity factors. If these sequences were longer than 70 h, the installations were classified as “erroneous”. These thresholds proved to be a good predictor of simulation quality and visual inspection confirmed that these installations showed irregular behaviour (see Fig. 2 for an example). Further 10 installations were manually classified as “erroneous”, as they showed very low simulation quality and after visual inspection, serious anomalies were observed in the data. The time series of two installations, “SOLAR HORMIGA” and “SOLAR EL AGUILA I”, were partially erroneous but it was possible to preserve at least 13 Months of consecutive data by removing parts of the time series. In total, we classified 34 installations as “erroneous”.

The 23 residual installations were assumed to have acceptable data quality. However, several of the time series resemble patterns of the use of tracking devices, i.e. the show a very flat production profile during the day. However, only for a very small subset information on the installation type could be derived from meta-data found during an internet search. We therefore, in the second classification step, used a clustering approach to understand which PV plants use tracking devices. For that purpose, we aggregated the generation profiles to an average hourly profile (i.e. calculating the average for each hour of the day), and normalized the time series by the maximum, so that generation was scaled between 0 and 1. Subsequently, we applied k-clustering algorithm to the generation time series of the remaining installations with low data quality. Subsequently, we applied k-clustering with 4 clusters, as using 2 clusters only did not achieve a satisfying classification of profiles. From visual inspection of the clustered profiles, we could identify which profiles belong to tracking and which ones belong to fixed PV plants (see Fig. 3). However, we had to re-classify one installation afterwards from tracking to fixed (“LAS MOLACAS”), as it clearly shows a fixed PV plant profile. In total, 9 installations were classified as “fixed” and 14 as “tracking”. We verified our classification against available internet sources [44–46] and could verify that 5 out of the 14 “tracking” installations indeed use single-axis trackers. For the other installations, we could not verify the system type with publically available information. The detailed classification algorithms and manual reclassifications can be found in the R-scripts on github [47].
In the present study, we assume that the “fixed” systems aim at maximizing yield in a year by using a panel orientation towards the equator (in this case towards North) and an inclination equal to the latitude. These conditions represent optimally installed PV systems without tracking [9,48]. For the “tracking” systems, we assume horizontal single-axis tracking with use of back-tracking. This combination of characteristics is common in commercially available systems since more than one decade and has a higher output than horizontal single-axis tracking system without back-tracking [49]. We have also opted for horizontal and not vertical tracking systems since all systems for which metadata was available were configured this way.

ERA5-land and MERRA-2 data and PV power output using PV_LIB

ERA5-land is defined by the ECMWF as an enhanced version of ERA5 for land applications [50]. The main particularity compared to ERA5 is the spatial resolution of around 9 km, which is more than three times higher than the one of ERA5 and more than five times higher than the resolution of MERRA-2. The data will eventually cover the same time horizon as ERA5 (January 1950 to near real time) and the period 2014–2018 was retrieved from the Copernicus Climate Data Store [35]. MERRA-2 is the replacement of MERRA with the same spatial resolution but with multiple improvements in the underlying algorithms. In contrast to ERA5-land, MERRA-2 is available from 1980 on [11]. The data for the period 2014–2018 and the territory of Chile was obtained from the Goddard Earth Sciences Data and Information Services Center of the NASA and downloaded with the MERRAbin package [51].

The PV output is calculated using PV_LIB for python [36]. This is a widely used open source toolbox created by the PV Performance Modeling Collaborative (PVPMC) of the Sandia National Laboratories. It is in continuous development since 2014. The reported users include Espejo et al. [31], which employs PV_LIB in combination with ERA5 data to predict the output of one installation in Italy with promising results. For the simulation of the two types of systems, “fixed” and “tracking”, we calculated solar position using PyEphem, which is a wrapper of the C library XEphem, a widely used ephemeris library in development since 1990 [54]. For the “fixed” system it has been assumed that PV installations are oriented towards North and inclined by an angle equal to the latitude of the location. These characteristics are an approximation of the necessary conditions for maximum output during a year without any tracking system [9,48]. For the calculation of tilt and azimuth of the “tracking” systems, we used the PV_LIB implementation of the Tracking algorithm by Lorenzo et al. [49].

Fig. 3. Clustering of hourly averaged capacity factors. The figure shows the 4 clusters derived by k-clustering for the 23 considered installations. Cluster 1 and 3 were classified as “fixed”, clusters 2, 4 as “tracking”. “LAS MOLLACAS” (in green in cluster 2) was manually reclassified to “fixed”.

Table 2: Variables from ERA5-land and MERRA-2 used for the PV output model.

| Reanalysis Variable | Description | Units |
|---------------------|-------------|-------|
| ERA5-land 10 m u-component of wind (u10) | The horizontal speed of air moving towards East, at a height of ten meters above the earth surface | m s⁻¹ |
| ERA5-land 10 m v-component of wind (v10) | The horizontal speed of air moving towards North, at a height of ten meters above the earth surface | m s⁻¹ |
| ERA5-land 2 m temperature (t2m) | Air temperature at 2 m above the earth surface | K |
| ERA5-land Surface solar radiation downwards (SSRD) | Amount of solar radiation reaching the earth surface. This variable is accumulated from the beginning of the forecast time to the end of the forecast step | J m⁻² |
| MERRA-2 2-meter eastward wind (U2M) | The horizontal speed of air moving towards East, at a height of two meters above the earth surface | m s⁻¹ |
| MERRA-2 2-meter northward wind (V2M) | The horizontal speed of air moving towards North, at a height of two meters above the earth surface | m s⁻¹ |
| MERRA-2 2-meter air temperature (T2M) | Air temperature at 2 m above the earth surface | K |
| MERRA-2 Incident shortwave land (SWGDN) | Amount of instantaneous solar radiation reaching a horizontal unit earth surface as hourly average | W m⁻² |
conversion from horizontal irradiance to an inclined surface requires an estimation of the direct normal irradiance (DNI) and diffuse horizontal irradiance (DHI) from the derived instantaneous SSRD and SWGDN in Wm\(^{-2}\). The DNI is obtained using the Direct Insolation Simulation Code (DISC) as presented in [55] and implemented in PV_LIB. DHI is calculated following Equation (1):

\[
\text{DHI} = \text{GHI} - \text{DNI} \cdot \cos(\phi) \quad (1)
\]

where \( \phi \) is the zenith previously calculated with PyEphem.

\( \text{GHI}, \text{DNI}, \text{DHI}, \) air temperature and wind speed as well as technical details of technologies launched in 2014 are the input to the PV model. The selected PV panel, “Silevo_Triex_U300 Black”, is part of the module data base of the Sandia National Laboratories and the selected inverter “ABB_MICRO_0_3_I_OUTD US_240_240V” belongs to the list of approved systems of the US Clean Energy Council. From the data base available for PV_LIB these systems resemble the state of the art technology in 2014, the first year for which we have generation observations for PV installations in Chile. The resulting time series were adjusted to daylight summer time which is used in the observed generation timeseries.

**Accuracy indicators for PV output time series**

To allow comparability with the results obtained for locations in Europe in [9], the comparison is performed for capacity factors. Three commonly used indicators in solar and PV forecasting literature, MBE, RMSE and Pearson correlation coefficient [4], defined in equations 2–4 respectively, are calculated for the hourly values, as well as for the daily and monthly averages:

\[
\text{MBE} = \frac{1}{N} \sum_{i=1}^{N} (\hat{I}_i - I_i) \quad (2)
\]

\[
\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{I}_i - I_i)^2} \quad (3)
\]

\[
\rho = \frac{\text{Cov}(\hat{I}, I)}{\text{Var}(\hat{I})\text{Var}(I)} \quad (4)
\]

where \( \hat{I}_i \) and \( I_i \) are the measured and simulated capacity factors.

It can be expected that clear sky conditions can be well simulated by the models, as they are predictable by daily and annual variations. The advantage of using additional climatic data is that they allow to incorporate weather effects, in particular the presence of clouds, on PV generation output. A weather data based simulation therefore should be able to predict differences from clear sky conditions reasonably well.

We therefore also estimated how well the models simulate PV generation when the time series do not follow the expected seasonality under clear-sky conditions. In the following, we call these timeseries deseasonalized. For that purpose, we calculated clear sky GHI, DNI, DHI and global irradiance in plane profiles for each location using the Ineichen and Perez clear sky model [56] also available in PV_LIB. These where normalized to one by dividing the time series by the highest irradiance value. The normalized clear sky GHI time series were subtracted from the hourly capacity values of the measured and calculated PV output data of the “fixed” systems, while the normalized global irradiance in plane was subtracted from the hourly capacity values of measured and calculated PV output data of the tracking systems.

Moreover, the accuracy indicators are calculated for the spatially aggregated values in their raw and deseasonalized versions for the two sets “fixed” and “tracking”. This should present evidence if spatial aggregation contributes to reduce error.

**Results**

**Raw timeseries**

Pearson correlation coefficients and RMSEs for all sets of installations and the hourly, daily and monthly capacity factors for the ERA5-land and MERRA-2 derived PV output are presented in Fig. 4. For all median and average values of the indicators and all levels of temporal aggregation with the exception of the correlation of the monthly values for tracking systems, ERA5-land performs slightly better than MERRA-2. However, differences are small and the spread is very similar. The largest differences between indicators for ERA5-land and MERRA-2 based PV calculations are observed for the correlations of daily data. In this case, the correlation for ERA5-land data on average is higher by 0.107 and 0.042 for fixed and tracking installations respectively on average. Although the RMSE is higher for tracking systems, the difference in simulation quality between fixed and tracking systems in terms of correlations depends on the temporal resolution. While they are higher for fixed systems at hourly resolution, the performance for daily and monthly resolution tend to be better for tracking systems. With higher temporal aggregation, correlation decreases. This goes against expectations, but is mainly due to the impact of some outliers in the comparison of monthly correlations. These outliers can be explained by the incapability of the simulations to entirely reproduce the yearly seasonality of the measured data as well as the short time series.
available for the calculation of the indicator (around 13 values in each of the outlier cases).

MERRA-2 on average overestimates generation more than ERA5-land, i.e. MERRA-2 on average overestimates capacity factors by 0.030, while ERAS-land shows an overestimation of 0.018 for fixed systems. For tracking systems, the numbers are 0.057 (MERRA-2) and 0.034 (ERA5-land). The overestimation of PV output based on MERRA-2 is consistent with literature and the lower values for the ERA5-land can at least partially be attributed to the higher spatial resolution of the data set that contributes to resolve more accurately irradiance-relevant weather events [9]. This is confirmed by Fig. B1 in the appendix, which shows the MBE of hourly capacity factors.

2D density plots of observations versus simulations are shown in Fig. 5. In a perfect simulation, all points would be on a 45-degree line. This is not the case, but a clear pattern showing substantial correlation can be found, as confirmed by Fig. 4. In all cases, a high number of values can be obviously found at the lower range, as here values close to night-time with very low generation cumulate. It is also clearly visible, that tracking systems have a much higher share of capacity factors above 50%, as they are able to capture a higher share of solar radiation. This causes a clustering of capacity factors at the lower and higher end, while fixed systems have a more continuous spread of values over the whole capacity range.

The patterns for fixed systems, comparing ERA5-land and MERRA-2, look similar. For tracking systems, there is however an interesting difference: at very high capacity factors, MERRA-2 overestimates generation stronger than ERA5-land, as can be seen by one dark red box at a capacity factor of around 0.9 for observations, but at 1 for MERRA-2. This does not occur for ERA5-land. In general, the plots confirm the overestimation described above.

**Deseasonalized timeseries**

Fig. 6 shows the Pearson correlation coefficient for deseasonalized time series. Correlations are substantially lower in comparison to the indicators obtained with raw time series, nevertheless, there seems to be additional explanatory value in the reanalysis data: correlations are well above 0 in all instances. Again, MERRA-2 performs worse than ERA5-land in almost all comparisons and tracking systems perform substantially worse than fixed systems. This may indicate that there is room for improvement in the simulations.

Fig. 7 shows 2D density plots of observed capacity factors versus simulated ones for the deseasonalized time series. Capacity factors are in the range of slightly below 0 to 0.5. This means that at some point in time, reanalysis data shows higher radiation values than under theoretical clear-sky conditions. However, for most of the time clear-sky conditions show higher radiation values than ERA5-land and for fixed systems, most values are aligned around a 45 degree line. For tracking systems, results are worse, as also indicated by lower correlations in Fig. 6. Values are not strongly aligned along the 45 degree line, but are concentrated around 0. We explain the lower quality of tracking simulations by its dependence on DNI, which is not available in any global reanalysis data but simulated from GHI. Additionally, detailed technical information about system configurations is very limited for tracking systems. However, simulation quality depends strongly on them, while for fixed systems, details about the technical configuration have a smaller impact on results. As observed for the raw time series, MERRA-2 overestimates more than ERA5-land.

**Aggregated results**

The indicators for the spatially aggregated values for fixed and tracking systems have been calculated using the time series since the start of operation. Results are presented in Table 3 and show that spatial aggregation reduces errors. The correlations for the aggregated time series are larger than the averages for the individual installations in all cases. For the raw data, the correlations of the aggregated values are even higher than the highest individual value among all individual installations independent of the temporal resolution. Even the deseasonalized time series reach correlations for hourly capacity factors of 0.86 for the ERA5-land derived PV output and 0.80 for the MERRA-2 derived data for fixed systems. The values are relatively low for the hourly deseasonalized estimations for tracking systems, which in their aggregated form reach values below the highest individual ones. For the daily values, the deseasonalized aggregated data have better correlation than the installation with the maximum correlation. For the monthly values, this relation is inversed but the differences between maximum individual and aggregated correlations are very low. MBEs are below 0.05 in all cases with the exception of the raw data for tracking systems simulated with MERRA-2 data. Similarly, the RMSEs remain always below 0.10 for all data with the exception of the hourly raw data for tracking systems simulated with MERRA-2 data. The results for the aggregated values corroborate the results for individual installations. Time series of PV output calculated with ERA5-land data are consistently better than the ones calculated with MERRA-2 but the difference is minor for most of the indicators.
Understanding simulation quality

Finally, we aimed at understanding if simulation quality is influenced by how sunny locations are. For that purpose, we plotted mean capacity factors against the correlation between simulation and observation as quality indicator (Fig. 8). For fixed systems, the correlation is, however, negative at −0.41, if data points from ERA5-land and MERRA-2 are assessed together. This depends strongly on one outlier installation (“PFV LOS LOROS”, the two points on the lower right of the left panel in Fig. 8). If it is removed, correlation turns positive to 0.35, while the removal of other installations, one at a time, changes the correlation to up to around 0. Nevertheless, there is very weak evidence on a relationship between average capacity factor and simulation quality. For tracking systems, the correlation is higher at 0.74 (taking into account both data sources). This result is robust against removal of single installations, as correlation never falls below 0.7.

We have additionally assessed if clouds may explain simulation quality. For that purpose, we have calculated the correlation between the average difference of a location’s irradiance from clear-sky conditions as an indicator of the impacts of clouds and the simulation quality in terms of correlation. Here, a negative correlation should be expected, i.e. the larger the difference between clear sky conditions and actual conditions, the lower the simulation quality. For tracking systems this can be confirmed with a correlation of −0.42, which is robust when removing single installations, however fixed systems even show a positive correlation of 0.53 (also see Fig. C1 in Appendix). Single outliers do not change this result for fixed systems. We conclude that there is evidence that the simulation quality of tracking systems is reduced in more cloudy regions. For fixed systems such a claim is not substantiated by our results.

Discussion

The RMSE values for hourly and daily data of “fixed” systems using ERA5-land for the simulation are similar to the ones presented in [9] for European installations. While the installations evaluated in [9] present an RMSE with median around 0.1 for hourly data estimated using MERRA-2, comparable installations calculated using ERA5-land present an RMSE with median around 0.11. These values slightly deteriorate in our case to 0.12 for simulations made with MERRA-2. For daily data, [9] shows RMSE values around 0.05 while the equivalent here are around 0.07 when calculated with ERA5-land data and around 0.08 when using MERRA-2 as input. Moreover, the minimum RMSE values obtained using ERA5-land data reach the average accuracy of the PV
output calculated with SARAH irradiance data (dataset showing the best results in [9]) for installations in Europe. The correlation of hourly values between observations and ERA5-land based simulations is for seven out of nine installations with fixed configuration better than the ones reported by Atencio Espejo et al. [31]. They used a combination between ERA5 data and PV_LIB to simulate PV output at a location in Italy. Furthermore, no study was found that addresses validates tracking installations in a similar way. Considering that the accuracy indicators calculated here deteriorate in many cases for such configurations, future research should be dedicated to the simulation of such systems.

The limited improvement in the accuracy of the results obtained with ERA5-land in comparison to the ones obtained with MERRA-2 for PV output is unexpected as there is a large increase in the spatial resolution from MERRA-2 to ERA5-land. Also, for wind power generation, ERA5 (the basis for ERA5-land) improves strongly over MERRA-2. For instance, Olauson [57] showed that hourly wind power estimations made with ERA5 were considerably better than the ones obtained using MERRA-2. His results indicate that ERA5 derived wind power estimations have higher correlation and on average 20% lower mean absolute and root mean square errors for national aggregated data of 4 European countries and the Bonneville Power Administration in the North-Western United States as well as for 1051 individual wind turbines in Sweden. This indicates that improvements in the wind variables in the reanalysis data sets outperform improvements in the solar radiation variables.

A clear output of the validation exercise presented here is the need for more and better data for validation and forecasting model development purposes. Using standard assumptions about PV system configurations may lead to the generation of PV output profiles that are considerably different to the actual ones, since differences on e.g. tracking system type or orientation have a large impact on the estimated output of the systems. The lack of meta-information is a known issue of PV installation data bases in Europe, where the information is either not available or has errors [9], an experience that is now replicated in energy transition newcomer countries such as Chile and Brazil. These issues are relatively easy to correct, at least for new installations that will be integrated in the data bases, but awareness about the importance of data requirements is necessary. Moreover, selection of optimal input weather data and PV and inverter models can only be improved if respective metadata is available for a wide range of locations. Data driven analysis such as the clustering used here might contribute to identify technical characteristics of installations where metadata is not available. This would, however, require specialized knowledge to improve accuracy and leave questions open on the configuration details of particular installations, such as e.g. the exact type of tracking or the size and model of the inverter. Additionally, data quality checks by data providers would reduce the amount of work necessary to clean up generation data and would also improve the validity of such experiments.

It should be taken into account that there are uncertainties in the technical PV and inverter assumptions and models as well as in the models which estimate DNI from GHI or transform GHI into irradiance on an inclined surface, which are difficult to quantify in the present study. However, the main parameter which determines the hourly variability of PV output is still GHI. Therefore, following the results

| Table 3 |
| --- |
| Indicators for the aggregated hourly capacity values for raw and deseasonalized data as well as “fixed” and “tracking” configuration types. |
| Indicator | Data set | Type | Hourly | Daily | Monthly |
| --- | --- | --- | --- | --- | --- |
| | | | Raw | Deseasonalized | Raw | Deseasonalized | Raw | Deseasonalized |
| Pearson | ERA5-land | fixed | 0.980 | 0.972 | 0.989 | 0.962 | 0.957 | 0.986 | 0.972 |
| | MERRA-2 | fixed | 0.798 | 0.967 | 0.990 | 0.789 | 0.965 |
| | ERA5-land | tracking | 0.686 | 0.959 | 0.890 | 0.857 |
| | MERRA-2 | tracking | 0.450 | 0.054 | 0.052 |
| MBE | ERA5-land | fixed | 0.028 | 0.027 | 0.027 |
| | MERRA-2 | fixed | 0.034 | 0.033 | 0.033 |
| | ERA5-land | tracking | 0.054 | 0.053 | 0.052 |
| | MERRA-2 | tracking | 0.071 | 0.070 | 0.070 |
| RMSE | ERA5-land | fixed | 0.064 | 0.045 | 0.036 |
| | MERRA-2 | fixed | 0.075 | 0.054 | 0.044 |
| | ERA5-land | tracking | 0.088 | 0.066 | 0.060 |
| | MERRA-2 | tracking | 0.117 | 0.084 | 0.077 |

Fig. 8. Mean capacity factor vs. correlation between observation and simulation.
obtained, there is still considerable room for improvement in the calculation of the GHG equivalent variables in the reanalysis data sets. A generalization of results at the global scale can only be achieved if improvements, validation exercises and development of forecasting methods are also made for locations distant from typical hot spots of research. Beyond the necessity of open availability of data and standardization of data warehousing procedures from official institutions, the scientific community can contribute here by making not only their simulated datasets available but also reference data that have been gathered from data bases that are only available in regionally known repositories. A platform such as open power system data [58] is an excellent start but efforts have to be made to make data of locations outside the European borders also available in an open and usable way. This however relies on original data providers using open data licenses.

Conclusions

A workflow to estimate multi-annual time series of photovoltaic power output based on data derived from a recent global reanalysis data set, ERA5-land, and an open source software library for PV power estimation, PV_LIB, has been proposed. This has been assessed using open data from 23 large PV installations in Chile. The results are in general satisfactory since correlation values of hourly capacity factors for most of the installations are above 0.8 always and reach 0.9 or more in several cases, mean biased errors are +/− 0.1 and root mean square errors are around 0.2. When the time series are deseasonalized using a clear sky model the correlation improves for the daily and monthly values but deteriorates for the hourly ones. Still, there is value in the reanalysis products as deviations from clear-sky models can be explained to a significant extent by the irradiation data from the reanalysis products.

ERA5-land results do not represent a major improvement compared to the ones calculated using the MERRA-2 reanalysis dataset, which is several years older and has considerably lower spatial resolution. Nevertheless, simulation quality improves with ERA5-land for almost all combinations of indicators, temporal resolutions, and type of time series. Simulation quality of tracking systems can be partly explained by average capacity factors at locations, i.e. sunnier locations seem to show better quality. For fixed systems, this relationship could not be confirmed, potentially due to the low sample size.

Additional meta-information on system characteristics and regular quality checks by data providers could substantially decrease the amount of work necessary to clean up data and to define setups for validation experiments. While this will depend mainly on governmental institutions, researchers can contribute here by making available cleaned datasets with meta-information on open data repositories.

CRediT authorship contribution statement

**Luis Ramirez Camargo:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization. **Johannes Schmidt:** Conceptualization, Methodology, Software, Validation, Formal analysis, Data curation, Writing - original draft, Writing - review & editing, Visualization, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This project has received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 Research and Innovation Programme (reFUEL, grant agreement No. 758149). The authors would like to acknowledge the ECMWF, the Chilean National Comission of Energy, The Chilean Ministry of Energy, and NASA for making their data open and freely available for research purposes.

Finally, the authors would like to thank the three anonymous reviewers as well as Katharina Gruber for their valuable feedback on the previous versions of the paper.

Appendix A. Table with main characteristics of the reference PV installations and total generation from all available records

| Latitude            | Longitude          | Type*     | Capacity official AC [MW] | Capacity estimated AC [MW] | Capacity diff. [%] | Official operation start | First observation | Avg CF measured | Total generation measured since first observation [GWh] |
|---------------------|--------------------|-----------|---------------------------|---------------------------|-------------------|--------------------------|------------------|----------------|--------------------------------------------------------|
| ANDES SOLAR         | −24.00129          | t         | 21.8                      | 22.27                     | 2.2               | 28.05.2016               | 10.02.2016       | 0.31           | 174.19                                                  |
| BELLAVISTA          | −31.66627          | e         | 3                        | 3.05                      | 1.67              | 27.03.2017               | 04.04.2016       | 0.16           | 11.76                                                   |
| CARRERA PINTO ET-APA I | −27.00424         | e         | 20                       | 89.76                     | 348.8             | 03.05.2016               | 28.12.2015       | 0.23           | 535.21                                                  |
| CHANARES            | −26.37306          | e         | 36                       | 36.06                     | 0.17              | 31.03.2017               | 16.12.2014       | 0.24           | 311.01                                                  |
| CONOJO SOLAR        | −25.5041           | e         | 104                      | 104.9                     | 0.87              | 30.04.2014               | 04.05.2016       | 0.26           | 634.95                                                  |
| CORDILLERILLA       | −35.1627           | e         | 1.33                     | 1.27                      | −4.51             | 01.12.2016               | 21.11.2016       | 0.18           | 4.29                                                   |
| EL DIVISADERO       | −30.84765          | t         | 3                        | 3.16                      | 5.33              | 01.06.2017               | 10.08.2016       | 0.12           | 7.96                                                   |
| FV BOLERO           | −23.46973          | e         | 138.2                    | 130.79                    | −5.36             | 02.06.2015               | 08.03.2017       | 0.27           | 556.32                                                  |
| LA CHAPEANA         | −30.5165           | t         | 2.93                     | 2.76                      | −5.8              | 22.12.2017               | 19.01.2016       | 0.25           | 17.73                                                   |
| LA QUINTA SOLAR     | −33.02892          | t         | 3                        | 3.09                      | 3                 | 21.12.2017               | 06.09.2017       | 0.27           | 9.7                                                    |
| LALACKAMA           | −25.11923          | e         | 55                       | 56.8                      | 3.27              | 31.08.2015               | 03.01.2015       | 0.27           | 538.65                                                  |
| LALACKAMA 2         | −25.10889          | e         | 16.5                     | 17.18                     | 4.12              | 19.10.2016               | 05.01.2015       | 0.23           | 135.79                                                  |
| LAS ARAUCARIAS      | −33.34603          | e         | 0.14                     | 0.11                      | −21.43            | 26.10.2016               | 12.05.2016       | 0.16           | 0.41                                                   |
| LAS MOLLACAS        | −30.67633          | f         | 2.93                     | 2.71                      | −7.51             | 09.06.2016               | 19.01.2016       | 0.22           | 15.62                                                   |
| LAS TURCAS          | −33.80384          | t         | 3                        | 3.13                      | 4.33              | 08.11.2017               | 10.01.2017       | 0.25           | 8.48                                                   |
| LOS PUQUIOS         | −20.43731          | e         | 3                        | 1.97                      | −34.33            | 28.03.2014               | 01.01.2016       | 0.12           | 6.11                                                   |
| LUNA DEL NORTE      | −30.04019          | t         | 70.6801                  | 70.6801                   | 0.00              | 01.03.2016               | 30.09.2014       | 0.24           | 20.64                                                   |
| LUZ DEL NORTE       | −27.02625          | f         | 141                      | 136.67                    | 3.07              | 24.02.2016               | 01.01.2015       | 0.22           | 1047.74                                                 |
| MARIA ELENA FV      | −22.2204           | e         | 68                       | 65.67                     | −3.43             | 21.01.2015               | 01.01.2016       | 0.33           | 572.02                                                  |
| PAMPA SOLAR NOR-TE  | −25.53295          | t         | 69.3                     | 70.28                     | 1.41              | 08.01.2017               | 21.03.2016       | 0.28           | 479.67                                                  |
| PARQUE SOLAR CUIZ   | −31.66307          | e         | 3                        | 2.79                      | −7                | 13.09.2017               | 01.10.2017       | 0.24           | 7.23                                                   |
| Location                  | Latitude   | Longitude  | t  | f  | e  | Date       | Date       | MBE    |
|---------------------------|------------|------------|----|----|----|------------|------------|--------|
| PARQUE SOLAR FINIS TERRAE | −22.3446   | −69.5224   | 138| 140.74| 1.99 | 08.03.2017| 06.01.2016| 0.29   | 1079.16|
| PARQUE SOLAR Pampa Camarones | −18.8857   | −70.1125   | 6.24| 6.22| −0.32 | 04.11.2015| 07.05.2016| 0.29   | 42.62  |
| PVF LAGUNILLA             | −30.5075   | −71.1119   | 3  | 2.92| −2.67 | 05.02.2016| 04.11.2015| 0.18   | 14.25  |
| PVF LOS LOROS             | −27.8513   | −70.1684   | 46 | 47.5| 3.26  | 11.08.2015| 11.06.2016| 0.13   | 139.33 |
| PILOTO SOLAR CARRDONES    | −27.5934   | −70.4403   | 0.4| 0.26| −35   | 21.02.2017| 07.02.2017| 0.18   | 0.78   |
| PLANTA PV CERRO DOMINADOR | −22.7928   | −69.4573   | 99.05| 99.82| 0.78  | 01.09.2017| 21.07.2017| 0.32   | 403.21 |
| PMGD PICA PILOT CAMARO NES | −18.8857   | −70.1125   | 6.24| 6.22| −0.32 | 04.11.2015| 07.05.2016| 0.29   | 42.62  |
| PUERTO SECO SOLARR        | −22.4492   | −68.8701   | 3  | 3.13| 4.33  | 21.12.2017| 06.09.2017| 0.27   | 9.74   |
| PV SALVADOR               | −26.3106   | −70.6856   | 68 | 66.31| −2.49 | 08.09.2016| 24.07.2014| 0.26   | 671.63 |
| QUILAPILUN                | −31.1037   | −70.6879   | 103.02| 95.49| −7.31 | 10.06.2013| 25.07.2016| 0.22   | 448.8  |
| SAN FRANCISCO SOLAR       | −27.5934   | −70.4403   | 0.4| 0.26| −35   | 21.02.2017| 07.02.2017| 0.18   | 0.78   |
| SANTA JULIA               | −32.3010   | −71.1039   | 3  | 3.2 | 6.67  | 04.10.2016| 01.05.2016| 0.25   | 18.72  |
| SOLAR JAMA 1              | −27.5934   | −70.4403   | 0.4| 0.26| −35   | 21.02.2017| 07.02.2017| 0.18   | 0.78   |
| SOLAR JAMA 2              | −27.5934   | −70.4403   | 0.4| 0.26| −35   | 21.02.2017| 07.02.2017| 0.18   | 0.78   |
| SOLAR JAVIERA             | −30.5075   | −70.1125   | 3  | 3.2 | 6.67  | 04.10.2016| 01.05.2016| 0.25   | 18.72  |
| SOLAR LAS TERRAZAS        | −27.5934   | −70.4403   | 0.4| 0.26| −35   | 21.02.2017| 07.02.2017| 0.18   | 0.78   |
| SOLAR LLANO DE LAMPAS     | −27.5934   | −70.4403   | 0.4| 0.26| −35   | 21.02.2017| 07.02.2017| 0.18   | 0.78   |
| SOLAR PSF LOMAS COLORADAS | −31.1037   | −70.6879   | 103.02| 95.49| −7.31 | 10.06.2013| 25.07.2016| 0.22   | 448.8  |
| SOLAR TECHOS ALTAMIRA     | −33.4791   | −70.5380   | 0.15| 0.1 | −33.33| 08.08.2013| 24.04.2014| 0.14   | 0.57   |
| SPS LA HUAYCA             | −20.4545   | −69.5340   | 25.05| 25.8| 2.99  | 09.03.2017| 01.01.2016| 0.27   | 180.21 |
| TAMBO REAL                | −30.4853   | −70.7701   | 2.94| 2.86| −2.72 | 28.08.2014| 01.01.2014| 0.11   | 13.45  |
| TELTIL SOLAR              | −27.5934   | −70.4403   | 0.4| 0.26| −35   | 21.02.2017| 07.02.2017| 0.18   | 0.78   |
| URIBE SOLAR               | −27.5934   | −70.4403   | 0.4| 0.26| −35   | 21.02.2017| 07.02.2017| 0.18   | 0.78   |
| VALLE DE LA LUNA II       | −33.2717   | −70.8708   | 3  | 3.2 | 9.42  | 11.10.2017| 01.09.2017| 0.25   | 8.7    |

* t: tracking, f: fixed, e: erroneous.

Appendix B. Figures of the MBE for raw and deseasonalized data.
Appendix C. Relationship between simulation quality and difference between clear sky conditions and irradiation on ground.

Fig. C1. Mean absolute difference between clear-sky conditions and ERAS-land irradiation on surface vs. correlation between observation and simulation.

References

[1] Bundesnetzagentur. EEG in Zahlen 2018 2019.
[2] Comisión Nacional de Energía de Chile, Ministerio de Energía del Gobierno de Chile. Reporte mensual ERNC Enero 2020. 2020.
[3] IEA PVPs Tl, Becquerel Institute, Kizuka I, Jäger-Waldau A, Donoso J, Detolleaure A, et al. Snapshot of Global PV Markets – 2020. IEA Photovoltaic Power Systems Programme; 2020.
[4] Blaga R, Sabadus A, Stefu N, Dughir C, Paulescu M, Badescu V. A current perspective on the accuracy of incoming solar energy forecasting. Prog Energy Combust Sci 2019;70:119–44. https://doi.org/10.1016/j.pecs.2018.10.003.
[5] Sobti S, Kothari-Kamal S, Rahim NAbd. Solar photovoltaic generation forecasting methods: A review. Energy Convers Manag 2018;156:659–97. https://doi.org/10.1016/j.enconman.2017.11.019.
[6] Das UK, Tey KS, Seyyedmahmoudian M, Mekhilef S, Idris MY, Van Deverent W, et al. Forecasting of photovoltaic power generation and model optimization: A review. Renew Sustain Energy Rev 2018;81:912–28. https://doi.org/10.1016/j.rser.2017.08.017.
[7] solargis. Bankable solar data for better decisions 2020. https://solargis.com/ (accessed February 19, 2020).
[8] Vaiala. Wind and Solar Online Tools. Vaiala 2020. https://www.vaiala.com/en/wind-and-solar-online-tools (accessed February 19, 2020)
[9] Pfenninger S, Staffelli L. Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data. Energy 2016;114:1251–65. https://doi.org/10.1016/j.energy.2016.08.060.
[10] Rienecker MM, Suarez MJ, Gelaro R, Todling R, Bacmeister J, Liu E, et al. MERRA: NASA’s modern-era retrospective analysis for research and applications. J Clim 2011;24:3624–48. https://doi.org/10.1175/JCLI-D-10-00015.1.
[11] Gelaro R, McCarty W, Suárez MJ, Todling R, Molod A, Takacs L, et al. The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). J Clim 2017;30:5419–54. https://doi.org/10.1175/JCLI-D-16-0758.1.
[12] Müller R, Pfiffner U, Träger-Chatterjee C, Trentmann J, Greiner R. Digging the METEOSAT treasure—3 decades of solar surface radiation. Remote Sens 2015;7:8067–101. https://doi.org/10.3390/rs7068067.
[13] Ren G, Wan J, Liu J, Yu D. Spatial and temporal assessments of complementarity for renewable energy resources in China. Energy 2019;177:262–75. https://doi.org/10.1016/j.energy.2019.04.023.
[14] Guesmard CA, Habte A, Sengupta M. Reducing uncertainties in large-scale solar resource data: the impact of aerosols. IEEE J Photovolt 2018;8:1732–7. https://doi.org/10.1109/JPHOT.2018.2869954.
[15] Höltinger S, Mikovits C, Schmidt J, Baumgartner J, Arheimer B, Lindström G, et al. The impact of climatic extreme events on the feasibility of fully renewable power systems: A case study for Sweden. Energy 2019;178:695–713. https://doi.org/10.1016/j.energy.2019.04.128.
[16] Huld T, Müller R, Gambardella A. A new solar radiation database for estimating PV performance in Europe and Africa. Sol Energy 2012;86:1803–15. https://doi.org/10.1016/j.solener.2012.03.006.
[17] European Commission. Photovoltaic Geographical Information System (PVGIS), EU Sci Hub - Eur Com 2020. https://ec.europa.eu/jrc/en/pvgis (accessed June 20, 2020).
[18] Schulz J, Albert P, Behr H-D, Capron D, Denke H, Diewitte S, et al. Operational climate monitoring from space: the EUMETSAT Satellite Application Facility on Climate Monitoring (CM-SAF). Atmos Chem Phys 2009;9:1687–709. https://doi.org/10.5194/acp-9-1687-2009.
[19] Hans-ErOl-Zentrum für Wetterforschung. COSMO Regional Reanalysis - COSMO-REA6 2018. http://reanalysis.meteo.uni-bonn.de/COSMO-REA6 (accessed January 24, 2018).
[20] Sengupta M, Xie Y, Lopez A, Habte A, Maclurin G, Shelby J. The National Solar Radiation Data Base (NSRDB). Renew Sustain Energy Rev 2018;89:51–60. https://doi.org/10.1016/j.rser.2018.03.003.
[21] Hersbach H, Bell B, Berrisford P, Horiony A, Muñoz J, Nicolas J, et al. Global re-analysis: goodbye ERA-Interim, hello ERA n.d.:10.
[22] Urraca R, Huld T, Lindfors AV, Rihelå A, Martinez-de-Pison FJ, Sanz-Garcia A. Quantifying the amplified bias of PV system simulations due to uncertainties in solar radiation estimates. Sol Energy 2018;176:663–77. https://doi.org/10.1016/j.solener.2018.06.065.
[23] Urraca R, Huld T, Gracia-Amillo A, Martinez-de-Pison FJ, Kaspar F, Sanz-Garcia A. Evaluation of global horizontal irradiance estimates from ERRAS and cosmo-rea6 analyses using ground and satellite-based data. Sol Energy 2018;164:339–54. https://doi.org/10.1016/j.solener.2018.02.059.
[24] Ramirez Camargo L, Graber K, Nitsch F. Assessing variables of regional reanalysis data sets relevant for modelling small-scale renewable energy systems. Renew Energy under review.
[25] Huang J, Riku LJ, Qin Y, Katzyje J. Assessing model performance of daily solar irradiance forecasts over Australia. Sol Energy 2018;176:615–26. https://doi.org/10.1016/j.solener.2018.10.080.
[26] Trolliet M, Walawender JP, Bourlès B, Boilley A, Trentmann J, Blanc P, et al. Downwelling surface solar irradiance in the tropical Atlantic Ocean: a comparison of re-analyses and satellite-derived data sets to PIRATA measurements. Ocean Sci 2018;14:1021–56. https://doi.org/10.5194/os-14-1021-2018.
[27] Jiang H, Yang Y, Bai Y, Wang H. Evaluation of the total, direct, and diffuse solar radiation estimates from the ERA5 reanalysis data in China. IEEE Geosci Remote Sens Lett 2020;7:47–51. https://doi.org/10.1109/LGRS.2019.2916410.
[28] Babar B, Graversen R, Bostrom T. Solar radiation estimation at high latitudes: Assessment of the COSMOSAF databases, ASR and ERA-S. Sol Energy 2019;182:397–411. https://doi.org/10.1016/j.solener.2019.02.058.
[29] Salazar G, Guemyard C, Galdino JB, de Castro Villela O, Fraidenraich N. Solar irradiance time series derived from high-quality measurements, satellite-based models, and reanalyses at a near-equatorial site in Brazil. Renew Sustain Energy Rev 2020;11;109477 https://doi.org/10.1016/j.rser.2019.109478.
[30] Puloloue BG, Kambidzis HD, Kaskaoutis DG, Karagiannis D, Polo JM. Comparison between MRM simulations, CAMS and PVGIS databases with measured solar radiation components at the Methoni station. Greece. Renew Energy 2020;146:1372–91. https://doi.org/10.1016/j.renene.2019.07.064.
[31] Atencio Espejo FE, Grillo S, Luisi L. Photovoltaic Power Production Estimation Based on Numerical Weather Predictions. 2019 IEEE Milan PowerTech, Milan, Italy. IEEE: 2019, p. 1–6. https://doi.org/10.1109/PWTC.2019.8810897.
[32] Sterf S, Liersch S, Koch H, van Lipzig NPM, Thiery W. A new approach for assessing synergies of solar and wind power: implications for West Africa. Environ Res Lett 2018;13:094009 https://doi.org/10.1088/1748-9326/aad8f6.
[33] Ramirez Camargo L, Valdes J, Masip Macia Y, Dorner W. Assessment of on-site processing and mapping of degradation mechanisms and degradation rates of PV modules. Energies 2019;12:10479. https://doi.org/10.3390/en121210479.
[34] Copernicus Climate Change Service (C3S). C3S ERA-land reanalysis. Copernic C3SERA Land Change Serv 2019. https://cds.climate.copernicus.eu/cdsapp#!/home (accessed July 29, 2019).
[35] Andrews RW, Stein JS, Hansen C, Riley D. Introduction to the open source PV LIB modules. Energies 2019;12:4749. https://doi.org/10.3390/en12244749.
[36] Comisión Nacional de Energía – Ministerio de Energía - Gobierno de Chile. Energía Abierta 2020. http://energiaabierta.cl/ (accessed February 19, 2020).
[37] National Aeronautics and Space Administration. MERRA-2 Data Access 2020.
https://gmao.gsfc.nasa.gov/reanalysis/MERRA-2/data_access/ (accessed July 8, 2020).

[40] International Energy Agency. Energy Policies Beyond IEA Countries: Chile 2018. 2018.

[41] Valdés J, Poque González AB, Ramirez Camargo L, Valín Fernández M, Masip Macía Y, Dorner W. Industry, flexibility, and demand response: Applying German energy transition lessons in Chile. Energy Res Soc Sci 2019;54:12–25. https://doi.org/10.1016/j.erss.2019.03.003.

[42] seatgeek. fuzzywuzzy. SeatGeek; 2020.

[43] Ramirez Camargo L, Schmidt J. Time series of electricity output for large grid connected photovoltaic installations in Chile 2020;1. https://doi.org/10.5281/zenodo.3939047.

[44] CVE Chile. PGMD plants and direct solar power sales in Chile. Cap Vert Énerg 2020. https://www.cvegroup.com/en/our-sites/cve-chile/ (accessed June 30, 2020).

[45] Tilseco. La Chapeana y Mollacas. Tilseco 2020. http://tilseco.com/portfolio/la-chapeana-y-mollacas/ (accessed June 30, 2020).

[46] Coordinador Eléctrico Nacional - SING. Información Técnica - Centrales Solares 2020. http://cdec2.cdec-sing.cl/pls/portal/cdec_web_cdec_sing SP_pagina?p_id=5193 (accessed June 20, 2020).

[47] Ramirez Camargo L, Schmidt J. https://github.com/inwe-boku/PV_from_era5. Institute for Sustainable Economic Development; 2020.

[48] Ramirez Camargo L, Stoeglehner G. Spatiotemporal modelling for integrated spatial and energy planning. Energy Sustain Soc 2018:8. https://doi.org/10.1186/s13705-018-0174-x.

[49] Lorenzo E, Narvarte L, Muñoz J. Tracking and back-tracking. Prog Photovolt Res Appl 2011;19:747–53. https://doi.org/10.1002/pip.1085.

[50] European Centre for Medium-Range Weather Forecasts. ERA5-Land: data documentation. Copernic Knowl Base - ECMWF Conflu Wiki 2020. https://confluence.ecmwf.int/display/CKB/ERA5-Land%3A+data+documentation# ERA5Land:datadocumentation-HowtociteERA5-Land (accessed February 20, 2020).

[51] Schmidt J. MERRAbin. 2019.

[52] Hoyer S, Hamman JJ. xarray: N-D labeled Arrays and Datasets in Python. J Open Res Softw 2017:5. https://doi.org/10.5334/jors.148.

[53] Silva J, Ribeiro C, Guedes R. Roughness length classification of Corine Land Cover classes. Proc. Eur. Wind Energy Conf. Milan Italy 2007:7–10.

[54] Downey E. XEphem 2015. http://www.clearskyinstitute.com/xephem/xephem.html (accessed June 30, 2020).

[55] Maxwell EL. A quasi-physical model for converting hourly global horizontal to direct normal insolation. Department of Energy: U.S; 1987

[56] Ineichen P, Perez R. A new airmass independent formulation for the Linke turbidity coefficient. Sol Energy 2002;73:151–7. https://doi.org/10.1016/S0038-092X(02)00045-2.

[57] Olauson J. ERA5: The new champion of wind power modelling? Renew Energy 2018;126:322–31. https://doi.org/10.1016/j.renene.2018.03.056.

[58] Neon Neue Energieökonomik, Europa-universität Flensburg, DIW Berlin, Technical University of Berlin, ETH zürich. Open Power System Data 2017. https://open-power-system-data.org/ (accessed February 28, 2020).