May 2023

OASES project: Work Package 3

WP3.1 Literature and data study report

Version 1

Authors:
Jan-Hendrik Grobler (CSIR)
Doctor Malatji (CSIR)
Stefan Karamanski (CSIR)
Gert Wessels (CSIR)

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

Disclaimer

The content of this report reflects only the authors' view. The European Commission is not responsible for any use that may be made of the information it contains.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

Document information

Grant Agreement	963530
Project Title	Long-Term Joint EU-AU Research and Innovation Partnership on Renewable Energy
Project Acronym	LEAP-RE
Project Coordinator	Vincent Chauvet (Vincent.chauvet@lgi-consulting.com) – LGI
Project Duration	1st October 2020 – 31st December 2025 (63 Months)
Related Work Package	WP3
Related Task(s)	WP3.1
Lead Organisation	CSIR
Contributing Partner(s)	CSIR, Fraunhofer IEE
Due Date	Month 6
Submission Date	15 June 2023
Dissemination level	

History

Date	Version	Submitted by	Reviewed by	Comments
3 May 2023	1	CSIR		
6 May 2023	2	Fraunhofer IEE	Malte Lindenmeyer	Minor changes

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

Table of contents

List of figures

This project has received funding from the European Union's Horizon 2020
Research and Innovation Program under Grant Agreement 963530.

List of tables

This project has received funding from the European Union's Horizon 2020
Research and Innovation Program under Grant Agreement 963530.

Abbreviations and Acronyms

Acronym	Description
ANN	Artificial Neural Networks
AR	Autoregressive
ARCH	Autoregressive conditional heteroscedasticity
ARFIMA	Autoregressive fractionally integrated moving average
ARIMA	Autoregressive integrated moving average
ARIMAX	Autoregressive integrated moving average with exogenous variables
ARMA	Autoregressive moving average
ARMAX	Autoregressive moving average model with exogenous variables
CBAM	ClimaCell Bespoke Atmospheric Model
CCA	Canonical Correlation Analysis
CDF	Cumulative Distribution Function
CMIP6	Coupled Model Intercomparison Project - Phase 6
DNI	Direct Normal Irradiance
ECMWF	European Center for Medium-Range Weather Forecasts
ERA5	ECMWF Reanalysis version 5
ESM	Earth Systems Model
ETSAP	Energy Technology Systems Analysis Program
GARD	Generalized Analog Regression Downscaling
GCM	Global Climate Model

This project has received funding from the European Union’s Horizon 2020
Research and Innovation Program under Grant Agreement 963530.

GCM	Global Circulation Model
GFS	Global Forecast System
GP	Genetic Programming
GRAF	Global High-Resolution Atmospheric Forecasting
GWA	Global Wind Atlas
IEA	International Energy Agency
ICON	Icosahedral Nonhydrostatic (modelling framework)
KIM	Korean Integrated Model
LARS-WG	Long Ashton Research Station Weather Generator
MA	Moving average
MACA	Multivariate Adaptive Constructed Analogs
MAE	Mean absolute error
MAPE	Mean absolute percentage error
MASE	Mean absolute scaled error
MBE	Mean bias error
MRE	Mean relative error
MSE	Mean square error
MSPE	Mean square percentage error
NCEP	National Centers for Environmental Prediction
NCUM	National Centre for Medium Range Weather Forecasting Unified Model
NMAE	Normalized mean absolute error

This project has received funding from the European Union’s Horizon 2020
Research and Innovation Program under Grant Agreement 963530.

NMSE	Normalized mean squared error
NRMSE	Normalized root mean squared error
NWP	Numerical weather prediction
NWS	National Weather Service
PV	Photovoltaic
RCM	Regional Climate Model
RMSE	Root mean squared error
SAST	South African Standard Time
SOM	Self-Organizing Map
SRCNN	Super Resolution Convolutional Neural Network
TIMES	The Integrated MARKAL-EFOM System
VAR	Vector autoregressive
WP	Work Package
WRF	Weather Research and Forecasting

Summary

The overall objective of the Open Access AU-EU Ecosystem for Energy System Modelling (OASES) project is the development and demonstration of a sustainable AU-EU ecosystem for energy system modelling, based on open-source software and open access data. The project consists of five interlinked work packages (WPs), each divided into several subtasks.

This report is the first deliverable within Work Package 3 (WP3), namely a report on currently available datasets for wind and solar resources, as well as on software, methodologies and techniques for wind power and PV time series generation.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

1. Introduction

The overall objective of the **O**pen **A**cces**s** AU-EU Ecosystem for **E**nergy **S**ystem Modelling (OASES) project is the development and demonstration of a sustainable AU-EU ecosystem for energy system modelling, based on open-source software and open access data. The project will build easy-to-use modelling workflows for different spatial scales, utilizing RES data developed in the project as well as data and tools from other similar high-quality efforts. The workflows will be used by six example case studies, each with different scope, that can be replicated using code, data, tutorials, and documentation from the proposed project. By so doing, the project enables local actors to learn and perform energy system scenario analysis relevant for their needs.

The work plan of the project is based on five work packages (WPs), each divided into several subtasks. Figure 1 gives an overview of the structure and lead partners of the work packages.

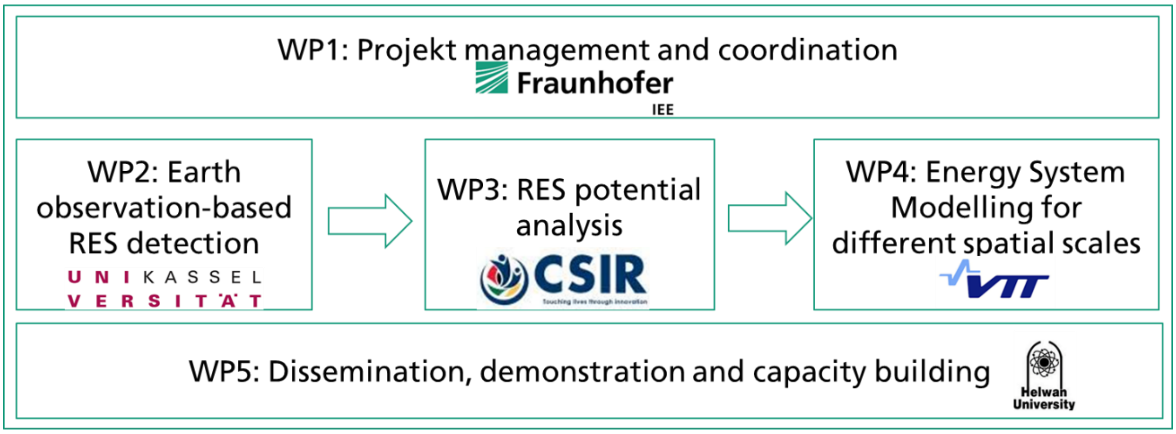

Figure 1 Overview of the structure and lead partners of the work packages

This report is the first deliverable within Work Package 3 (WP3). WP 3 covers a number of tasks, starting with a resource assessment (available data, software and methods) and ending with the generation of high-resolution RES time series. These RES time series are then used as input for energy system models in WP4. Table 1 provides a summary of the tasks related to the deliverables in WP3.

D3.1	Report on current available datasets on wind and solar as well as software, methodologies and techniques for wind power and solar PV time series generation.
D3.2	Description of the general model for wind power and solar PV time series generation for any future scenario.
D3.3	Description of the time series data sets for wind and solar generation for the selected national case studies of Egypt, Algeria and South Africa.

Table 1 Deliverables for WP3

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

This report provides a starting point for the other tasks in WP3. It identifies available datasets, methods and software that can be used for developing the required RES datasets. From these, a selection will be made and further used during the rest of the WP3 tasks.

This report is structured in a simple way. The section following this introduction, Section 2, provides a discussion of the potential data sources for WP3 available globally and locally. This is followed by a section describing possible methods that have been identified through a search of available literature. A final section deals with software currently available and considered by the team for usage. The report ends with a brief conclusion.

Since this document is the starting point of WP3, it is foreseen that the content, findings and recommendations in this document may still change throughout WP3 – while working with the data other issues may arise, or an identified method or software may require additions or changes. Therefore, this document is presented as a first version, and further versions may be developed as learning is obtained in WP3.

This project has received funding from the European Union's Horizon 2020
Research and Innovation Program under Grant Agreement 963530.

2. Datasets

Time-series datasets are critical in developing energy models and forecasts. The available datasets are explored in this section.

2.1 Global datasets

2.1.1 The Prediction of Worldwide Energy Resources (POWER)

The Prediction of Worldwide Energy Resources (POWER) dataset, developed by NASA Langley Research Centre, intends to provide solar and meteorological data. NASA claims that this freely and publicly available data supports renewable energy, building energy efficiency and agricultural needs (Zhang, et al., 2008). The interface of the online platform where the data can be accessed is displayed in Figure 2.

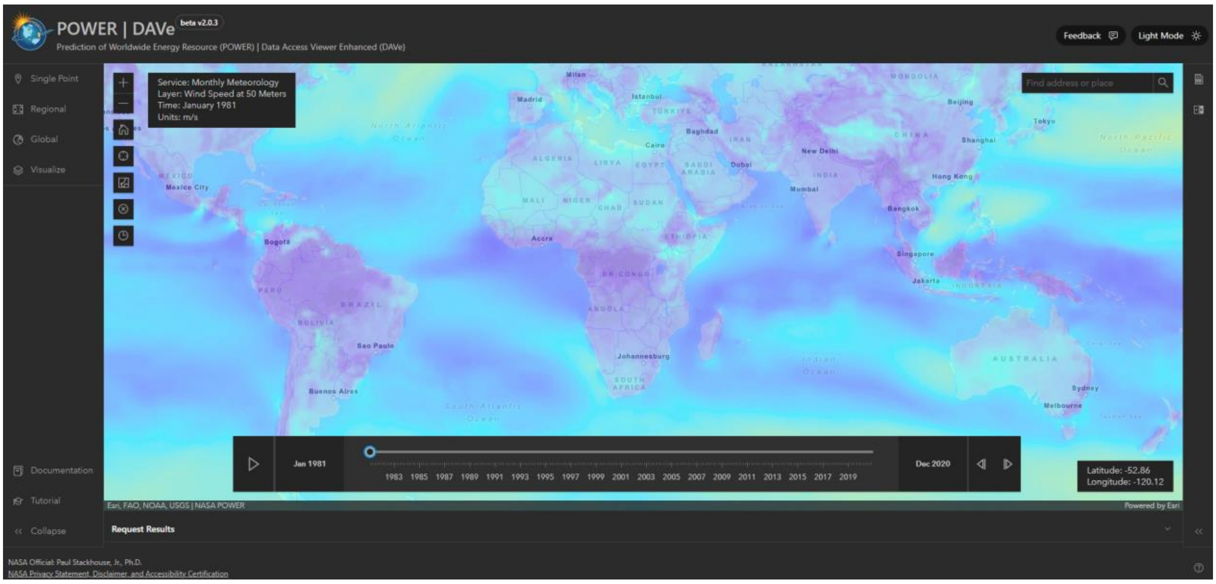

Figure 2: POWER online interface (POWER, n.d.)

Figure 3: POWER logo (The NASA POWER Team, 2022)

The POWER dataset provides hourly, daily, monthly and yearly radiation and meteorological data (more information on POWER's temporal overview can be found at https://power.larc.nasa.gov/docs/services/api/temporal/). The radiation dataset starts from 1 January 1984 to near-real time. The meteorological dataset starts from 1 January 1981 to near-real time (The NASA Power Team, n.d.). The downloadable time-series radiation data is shown in Table 2.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

All Sky Surface Longwave Downward Irradiance	Clear Sky Surface Longwave Downward Irradiance
All Sky Surface Shortwave Downward Irradiance	Clear Sky Surface Shortwave Downward Irradiance
Top-Of-Atmosphere Shortwave Downward Irradiance	

Table 2: POWER radiation data

The downloadable time-series meteorological data is shown in Table 3, below:

Precipitation Corrected Sum	Temperature at 2 m
Surface Pressure	Temperature at 2 m Minimum Average
Specific Humidity at 2 m	Temperature at 2 m Maximum Average
Relative Humidity at 2 m	Wet Bulb Temperature at 2 m
Dew/Frost Point at 2 m	Earth Skin Temperature
Eastward Wind at 10, 50 m	Northward Wind at 10, 50 m
Wind Direction at 10, 50 m	Wind Speed at 10, 50 m

Table 3: POWER meteorological data

The time-series data can be downloaded from a single point (given by a lat/long co-ordinate), a specific region (given by four lat/long co-ordinates) or globally. Data requests can be done through the POWER portal, found at https://power.larc.nasa.gov/beta/data-access-viewer/. The downloaded data can be in JSON, CSV, ASCII or NetCDF file formats. An example of the monthly temperature at 2 m, from 1996-2008 downloaded from the POWER dataset is displayed in Figure 4. This example, in CSV format, is for a single point, and the lat/long co-ordinate is shown. Annual means are also included.

This project has received funding from the European Union's Horizon 2020
Research and Innovation Program under Grant Agreement 963530.

-BEGIN HEADER-
NASA/POWER CERES/MERRA2 Native Resolution Monthly and Annual
Dates (month/day/year): 01/01/1996 through 12/31/2008
Location: Latitude -33.89 Longitude 19.21
Elevation from MERRA-2: Average for 0.5 x 0.625 degree lat/lon region = 449.62 meters
The value for missing source data that cannot be computed or is outside of the sources availability range: -999
Parameter(s):
T2M MERRA-2 Temperature at 2 Meters (C)
-END HEADER-

PARAMET	YEAR	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	ANN
T2M	1996	21.16	20.56	18.27	17.34	14.68	10.62	8.19	9.19	10.9	13.37	13.65	17.95	14.64
T2M	1997	20.1	19.54	18.81	15.3	14.07	9.55	10.26	11.05	14.39	16.73	16.44	19.65	15.48
T2M	1998	20.21	22.61	19.32	17.3	13.17	10.76	9.8	10.72	11.83	15.4	16.86	19.74	15.6
T2M	1999	21.64	21.69	21.47	17.73	14.01	12.28	10.77	11.83	11.73	16.81	18.09	23.45	16.77
T2M	2000	21.18	21.2	18.27	16.44	13.98	12.29	10.76	12.71	12.03	15.61	18.65	19.05	16
T2M	2001	20.47	22.02	20.19	16.79	14.61	11.88	10.33	10.61	12.3	15.97	18.61	20.14	16.12
T2M	2002	19.92	21.96	21.31	17.23	12.89	9.62	9.56	10.71	13.8	14.79	16.69	21.49	15.8
T2M	2003	20.65	22.02	19.58	17.17	13.8	10.84	10.29	8.99	12.15	15.9	18.36	18.66	15.66
T2M	2004	22.16	21.82	17.88	16.94	15.25	11.48	10.43	11.53	13.33	15.52	19.62	20.56	16.36
T2M	2005	20.55	21.38	20.4	16.17	13.08	10.4	12.15	9.61	13.29	15.48	17.58	18.55	15.69
T2M	2006	21.15	22.07	18.42	16.12	11.98	11.29	9.86	9.7	13.12	14.4	16.86	18.53	15.25
T2M	2007	22.3	20.64	19.38	16.78	14.02	10.5	9.61	9.98	11.96	15.11	15.39	19.12	15.37
T2M	2008	20.61	21.3	19.92	16.94	14.83	11.11	9.8	10.66	10.29	14.46	16.11	19.57	15.45

Figure 4: Monthly temperature at 2 m, from 1996-2008, downloaded from the POWER dataset

The resolution of a dataset is an important factor in determining which applications the dataset can be used for and for taking into consideration any accuracy sacrifices that are made. The horizontal resolution of the radiation data is a global 1° x 1° lat/long grid. The horizontal resolution of the meteorological data is a 0.5° x 0.625° lat/long grid (The NASA Power Team, n.d.).

2.1.2 Global Solar Atlas

The Global Solar Atlas (GSA) is an online platform where solar datasets can be obtained to determine areas with a high solar potential. This free publicly available dataset can be used in the preliminary stages of determining the viability of solar farms and for local governments to understand the solar potential of their areas. The GSA allows users to visualise areas of interest using interactive maps, as displayed in Figure 5.

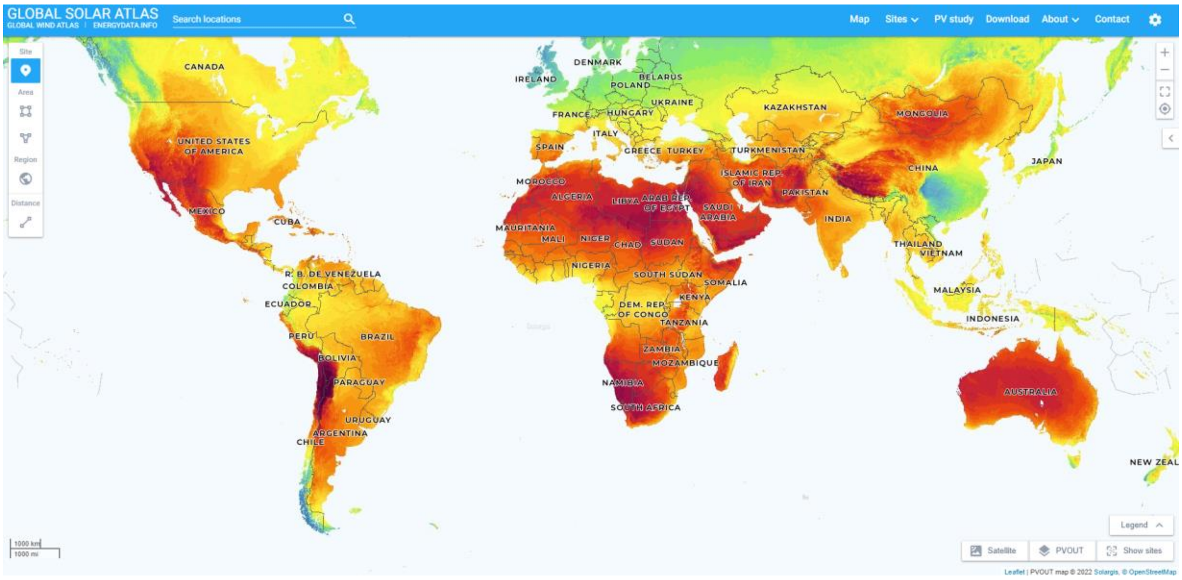

Figure 5: Global Solar Atlas online interactive map (Global Solar Atlas, n.d.)

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

The GSA also features a PV energy yield calculator, where user defined inputs can be given, and long-term energy yield can be calculated. This allows stakeholders to understand seasonal and intra-day variability of PV production better (Global Solar Atlas, n.d.).

Data can be viewed on the online platform itself or downloaded as a PDF (visual plots) or an Excel document (raw data). The data that can be downloaded, as well as its respective temporal and spatial resolution, is specified in Table 4.

Data parameter	Acronym	Unit	Temporal aggregation	Spatial resolution	Source(s)
PV Electricity Output	PVOUT	kWh/kWp or kWh	12 x 24 (month x hour) profiles	30 arcsec (~1 km)	Solargis
Global Horizontal Irradiation	GHI	kWh/m2	Annual average	9 arcsec (~250 m)	Solargis
Diffuse Horizontal Irradiation	DIF	kWh/m2	Annual average	9 arcsec (~250 m)	Solargis
Direct Normal Irradiation	DNI	kWh/m2	12 x 24 (month x hour) profiles	9 arcsec (~250 m)	Solargis
Optimum inclination [º] for inclined and fixed equator facing PV modules	OPTA	º	Annual average	2 arcmin (~4 km)	Solargis
Air Temperature at height of 2m	TEMP	°C	Annual average	30 arcsec (~1 km)	ERA5, post-processed by Solargis
Elevation	ELE	m	-	3 arcsec (~90 m)	SRTM v4.1 and other multiple sources, post-processed by Solargis

Table 4: Available downloadable data from the GSA (Global Solar Atlas, n.d.)

The time over which the GSA has recorded data varies per region. A summary of the temporal and geographical coverage is illustrated in Figure 6. Data is not available above 60°N, as shown with gray shading on the figure. Air temperature values are calculated from 1994-2021 globally.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

Figure 6: Temporal and geographical coverage of the GSA (Global Solar Atlas, n.d.)

An area in the Karoo Hoogland Local Municipality in South Africa (-31.052934°, 20.731201°) was selected as an example location. Below is an example of a downloaded file, in Excel format, including direct normal irradiation (DNI) for this location.

Direct normal irradiation [Wh/m²]

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
0 - 1	0	0	0	0	0	0	0	0	0	0	0	0
1 - 2	0	0	0	0	0	0	0	0	0	0	0	0
2 - 3	0	0	0	0	0	0	0	0	0	0	0	0
3 - 4	0	0	0	0	0	0	0	0	0	0	0	0
4 - 5	0	0	0	0	0	0	0	0	0	0	0	2
5 - 6	286	29	0	0	0	0	0	0	0	186	466	559
6 - 7	695	644	389	218	50	5	1	65	420	633	719	766
7 - 8	816	797	710	627	533	409	433	553	667	752	810	866
8 - 9	901	883	810	732	666	629	660	678	770	828	884	931
9 - 10	952	938	867	784	730	703	730	750	835	880	929	970
10 - 11	969	955	895	816	760	738	764	788	870	900	941	989
11 - 12	948	935	883	815	777	756	784	808	878	910	938	975
12 - 13	919	889	846	791	774	750	787	810	874	887	918	940
13 - 14	877	828	799	752	740	718	763	780	841	842	876	895
14 - 15	822	780	735	690	674	659	701	725	780	774	828	842
15 - 16	763	725	665	608	556	525	591	632	682	700	768	785
16 - 17	697	655	550	253	25	22	90	255	397	585	675	710
17 - 18	575	377	82	0	0	0	0	0	0	25	231	506
18 - 19	26	0	0	0	0	0	0	0	0	0	0	1
19 - 20	0	0	0	0	0	0	0	0	0	0	0	0
20 - 21	0	0	0	0	0	0	0	0	0	0	0	0
21 - 22	0	0	0	0	0	0	0	0	0	0	0	0
22 - 23	0	0	0	0	0	0	0	0	0	0	0	0
23 - 24	0	0	0	0	0	0	0	0	0	0	0	0
Sum	10246	9435	8231	7086	6285	5914	6304	6844	8014	8902	9983	10737

Figure 7: Example of DNI downloaded data from the GSA in the Karoo Hoogland Local Municipality

Not only can the data be downloaded in Excel format, but it can also be downloaded in PDF format, where it can be visualised easily. Examples of visualisations are shown in Figure 8, Figure 9 and Figure 10 for the Karoo Hoogland Local Municipality. Firstly, in Figure 8, site information can be seen. This includes:

- The map data as per Table 4;
- An image of the geographic location selected;
- Horizon and sunpath, in South African Standard Time (SAST) in this case; and
- PVOUT map, showing the long-term PV power potential.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

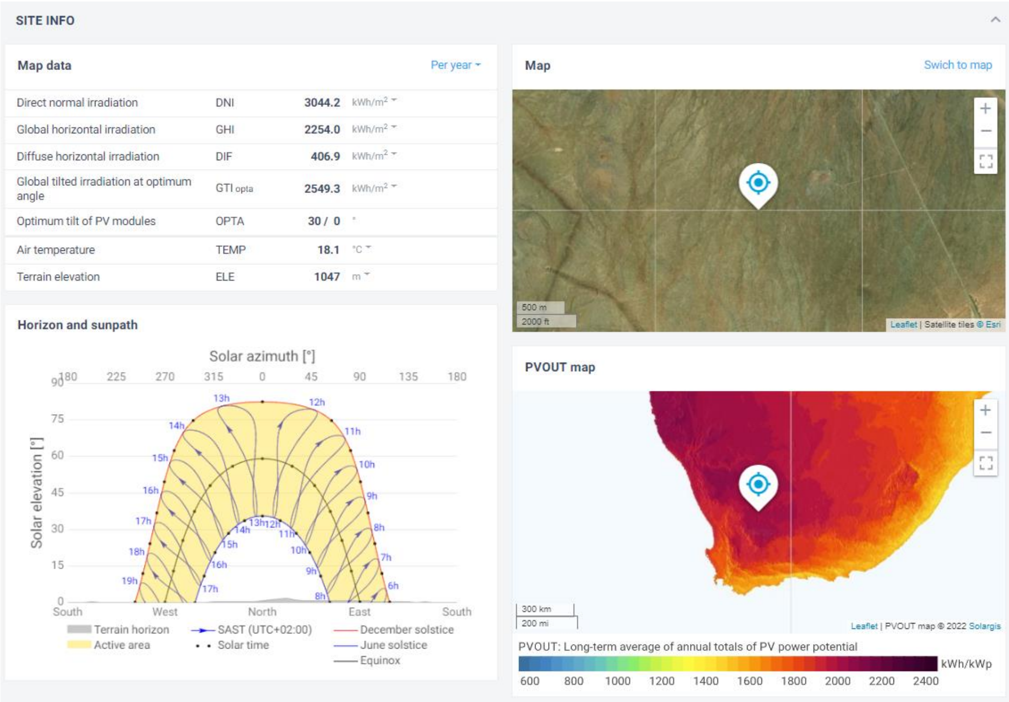

Figure 8: Site information downloaded example

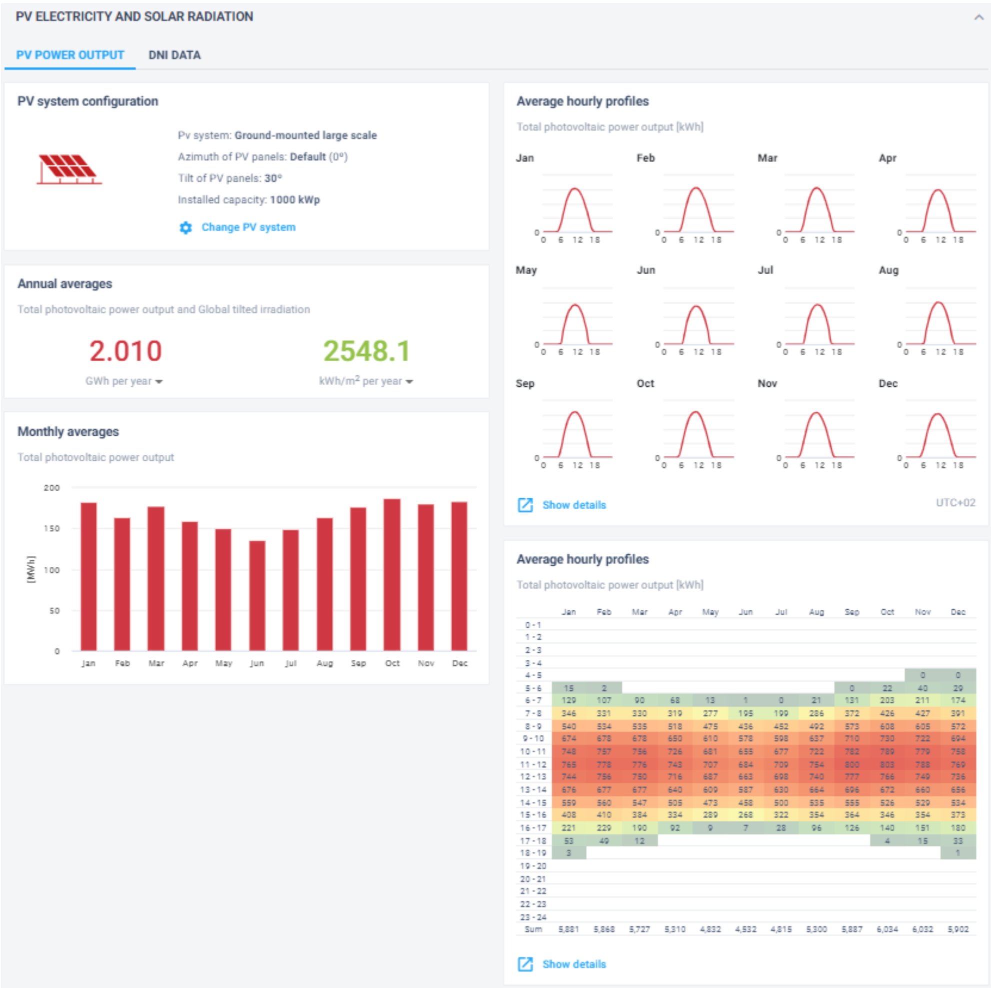

Figure 9: PV power output information downloaded example

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

Secondly, in Figure 9, PV power output information can be seen. This includes:

- PV system configuration (which is user customisable)
- Annual energy averages
- Monthly energy averages
- Hourly energy averages, per month of year
    - In curve format
    - In heatmap format

Figure 10: DNI information downloaded example

Lastly, in Figure 10, DNI information can be seen. This includes:

- Annual average DNI
- Monthly average DNI

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

- Hourly DNI averages, per month of year
    - In curve format
    - In heatmap format

The GSA is validated through solar measurement stations which capture high quality irradiation data that are used to improve the solar model. This is funded by The Energy Sector Management Assistance Program (ESMAP), which allows for a minimum of 2 years of data to be collected, resulting in a significantly lower amount of uncertainty for developers and planners (Global Solar Atlas, n.d.). This means solar projects can be of a higher quality and more effective. The validation countries are Malawi, Maldives, Pakistan, Tanzania, Zambia (Global Solar Atlas, n.d.).

In terms of accuracy, the GSA expects uncertainty for annual GHI and DNI values to be ±4% and ±9% respectively (Global Solar Atlas, n.d.) for:

- Most of Europe and North America (approx. below 50°N) and Japan.
- Mediterranean regions, Arabian Peninsula (except the Gulf region) and Morocco.
- South Africa, Chile, Brazil, Australia (Global Solar Atlas, n.d.).

However, the GSA expects a higher uncertainty for annual GHI and DNI values at ±8% and ±14% respectively in specific situations (Global Solar Atlas, n.d.) for these:

- High latitudes (approx. above 50°).
- Countries in humid tropical climate (e.g., equatorial regions of Africa, America and Pacific, Philippines, Indonesia, and Malaysia) and coastal zones (approximately up to 15 km from a body of water).
- Regions with high and dynamically changing concentrations of atmospheric aerosols (Northern India, West Africa, Gulf region, some regions in China).
- High mountainous regions with regular snow and ice coverage, and high-reflectance deserts.
- Regions with limited or no availability of high-quality ground measurements (Global Solar Atlas, n.d.).

2.1.3 Global Wind Atlas

The Global Wind Atlas (GWA) is another online platform, where wind datasets can be obtained to determine areas with a high wind potential. This data is intended for use by policymakers, planners, and investors (Global Wind Atlas, n.d.). The web-based application portal, illustrated in Figure 11, allows for freely downloadable datasets and high-resolution GIS-compatible maps. The GWA is a collaborative effort between the Technical University of Denmark's Department of Wind Energy and the World Bank Group (Global Wind Atlas, n.d.).

This project has received funding from the European Union's Horizon 2020
Research and Innovation Program under Grant Agreement 963530.

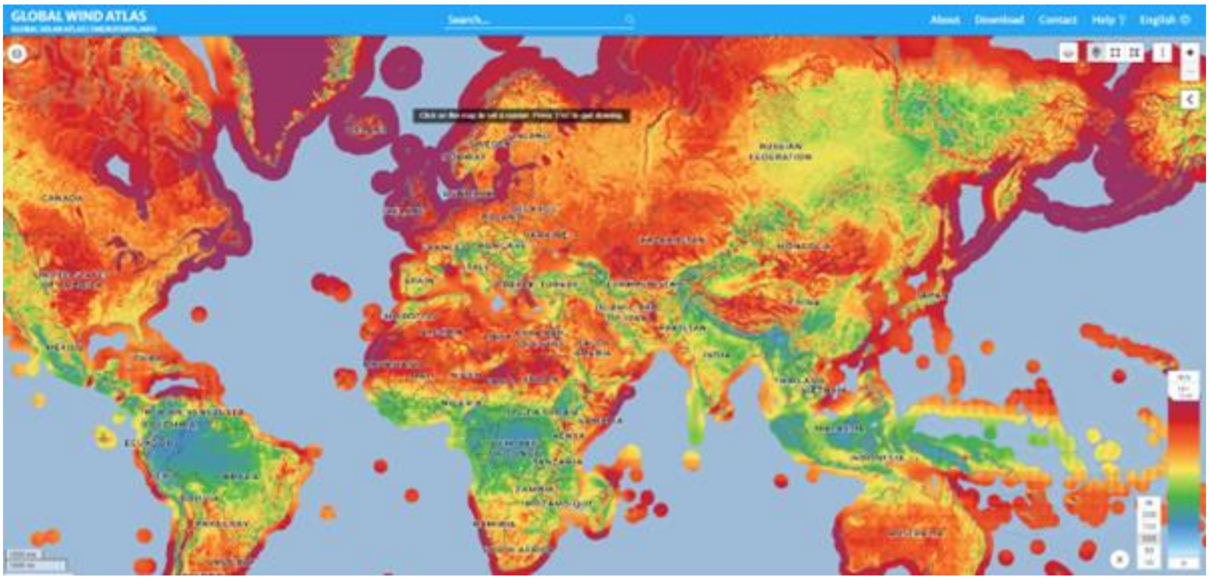

Figure 11: Global Wind Atlas web-based portal

The GWA offers global onshore and offshore (up to 200km from shore) wind data at 10, 50, 100, 150 and 200 m elevation. Wind resource mapping from the GWA is at a 250 m horizontal grid spacing (Global Wind Atlas, n.d.). Users of the GWA can access the data at yearly, monthly and hourly temporal resolutions.

A single point or custom area can be selected as the location of interest. The GWA's map generation tool allows users to see (and download in PNG/PDF format) maps of capacity factor, mean wind speed, wind power density, orography and roughness length of the area. In addition, the GWA has an energy yield calculator that allows users to select a generic or custom wind turbine, and use this to acquire annual energy production, capacity factor and full load hours – this can be downloaded in GIS compatible format. Furthermore, WAsP LIB files can be downloaded.

Mean Power Density	Wind Speed Variability - Hourly
Wind Frequency Rose	Wind Speed Variability - Monthly
Wind Speed Rose	Wind Speed Variability - Annually
Wind Power Rose	Hourly vs. monthly (radar plot)
Mean Wind Speed	Hourly vs. monthly (cross table)

Table 5: Summary of downloadable data in JSON or CSV format

This project has received funding from the European Union's Horizon 2020
Research and Innovation Program under Grant Agreement 963530.

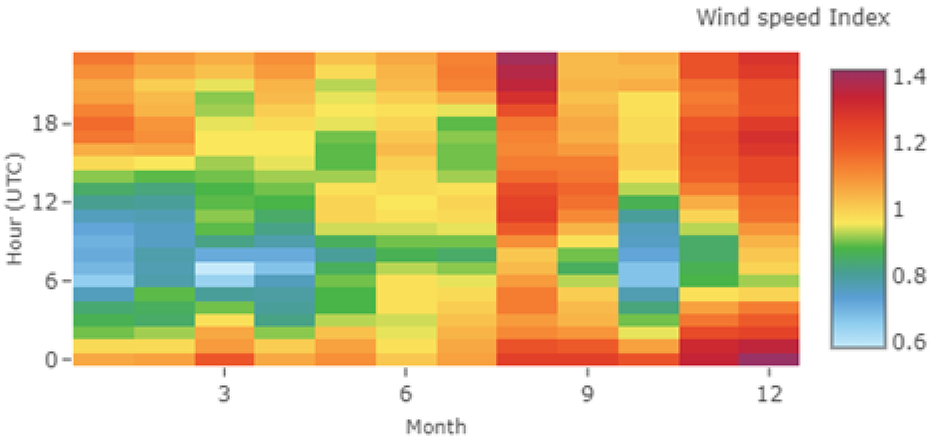

Figure 12: Example of hourly wind speed variability downloadable plot from the GWA

The GWA dataset is validated using measurement data, funded by The Energy Sector Management Assistance Program (ESMAP), as well as publicly available high quality wind data (Global Wind Atlas, n.d.). This is an ongoing process, performed by the Wind Energy Department of the Technical University of Denmark (DTU). The validation countries, as illustrated in Figure 13, are: Bangladesh, Maldives, Pakistan, Papua New Guinea, Vietnam, Zambia.

Figure 13: GWA validation countries (outlined in bold)

There are 35 sites where validation has been performed, where the mean absolute bias of the mean wind speeds is 14%, with a standard deviation of 10% (Global Wind Atlas, n.d.).

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

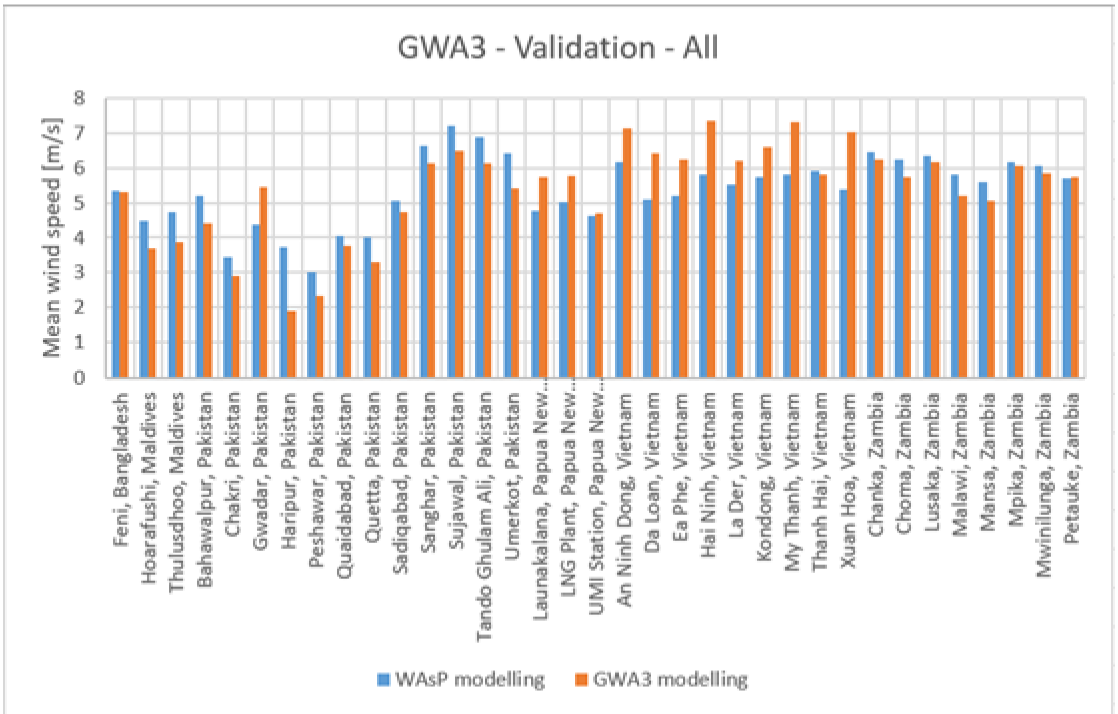

Figure 14: Mean wind speed modelling comparison between WAsP and GWA

As with many other models, there are limitations of the GWA. The mesoscale modelling process introduces uncertainties (Global Wind Atlas, n.d.) in:

- representativeness of the large scale forcing and sampling
- model grid size
- description of the surface characteristics
- model spin-up
- simulation time
- modelling domain size

The microscale modelling process introduces uncertainties (Global Wind Atlas, n.d.) in:

- the orographic flow model within WAsP
- the surface description and departures from the reference wind profile

The GWA allows for the addition of a RIX layer on the map, which is an index of ruggedness of the terrain. As the terrain's RIX increases, the data becomes more uncertain. This is due to the BZ-model being used past its operational envelope. Errors can consequently occur in the flow modelling, resulting in the model producing higher than expected mean wind speeds.

2.2 Local datasets

Although global datasets provide a lot of information, a country might need more detail to address a specific need. As a result, an organization might generate an energy atlas focusing specifically on a country. These datasets often prove to be invaluable for validation purposes.

To simplify the dissemination of data there are platforms that will have a searchable database. Energydata.info is one such platform that provides data for free. Other examples

This project has received funding from the European Union's Horizon 2020
Research and Innovation Program under Grant Agreement 963530.

include datahub.Geocradle.eu and African-energy.com although the latter makes use of a subscription service to access the data.

The Solar Atlas of Egypt (https://solea.gr/solar-atlas-of-egypt-2/) is a website developed under the framework of Geocradle.eu. It is based on 15 years of climatological radiation data from EUMETSAT (European Organization for the Exploitation of Meteorological Satellites). It provides mean monthly data at about 5-kilometer resolution.

The Renewable Energy Atlas of Morocco is a decision support tool developed by NOVELTIS for ADEREE, a Moroccan national agency with the aim to develop local renewable energy. Unfortunately, this data appears to be only available for ADEREE.

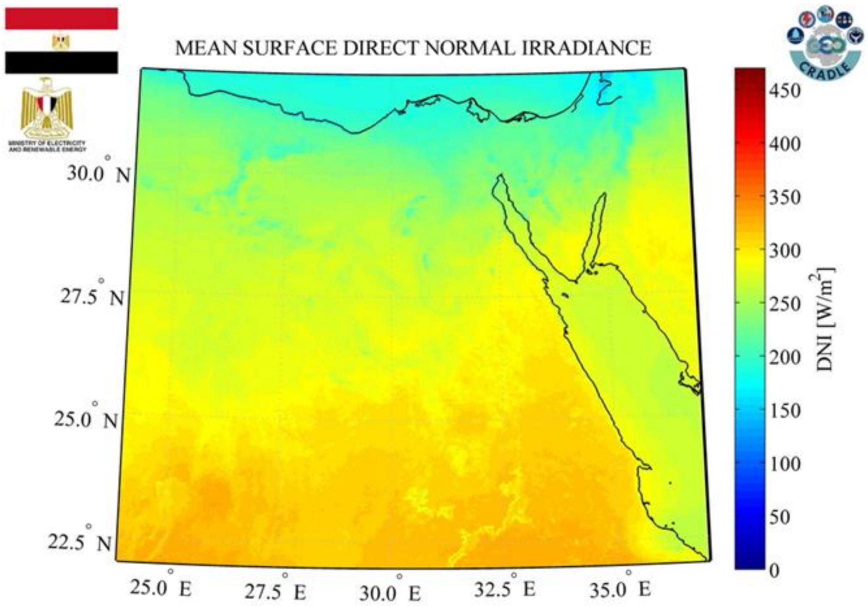

Figure 15: Mean surface direct normal irradiance for February in Egypt from the Solar Atlas of Egypt.

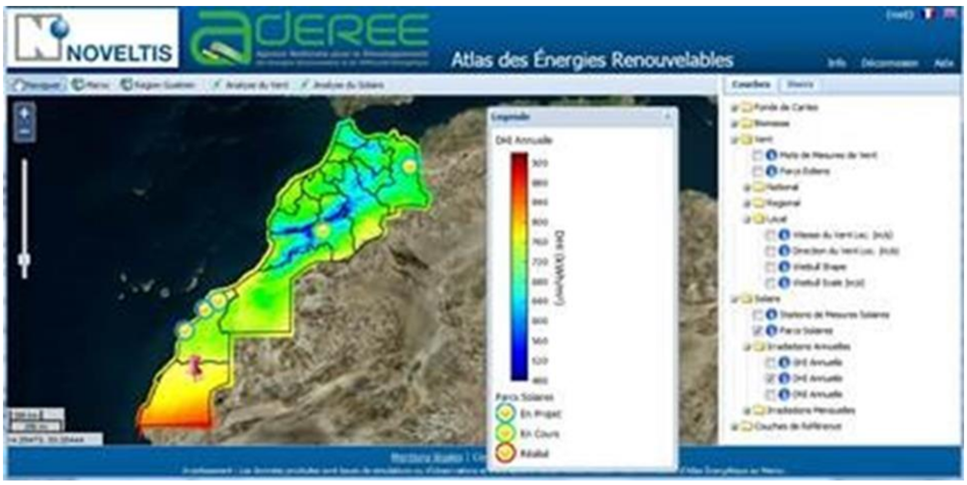

Figure 16: A renewable energy decision support tool developed by NOVELTIS for the Moroccan Agency ADEREE.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

The Wind atlas of South Africa (WASA, https://www.wasaproject.info/) is a project of the South African National Energy Development Institute with the objective to enable large scale wind exploitation in South Africa. The data is freely available.

This project has received funding from the European Union's Horizon 2020
Research and Innovation Program under Grant Agreement 963530.

3. Methodologies and Techniques

3.1 Time series methods in forecasting

An excellent overview is provided by Ghofrani and Alolayan of the most often used time series methods for solar and wind forecasting. Their overview forms a chapter in the book published by IntechOpen that is available online for free (Mohamudally, 2018). A summary of the methods is given in Table 6.

Name	Description	Reference
Autoregressive (AR)	The current (forecasted) value is determined as a linear combination of the past values of the same variable and a term for signal noise.	(University of New Mexico, 2006)
Moving average (MA)	Combines *n* past noise values to develop a time-series.	(Inman, Pedro, & Coimbra, 2013)
Autoregressive moving average (ARMA)	The model consists of a combination of AR and MA terms.	(Inman, Pedro, & Coimbra, 2013)
Autoregressive moving average model with exogenous variables (ARMAX)	Exogenous variables such as cloud cover, humidity, wind speed and direction are added to the ARMA model.	(Inman, Pedro, & Coimbra, 2013)
Autoregressive integrated moving average (ARIMA)	This model is used for non-stationary time series.	(Wang & Niu, 2009)
Autoregressive fractionally integrated moving average (ARFIMA)	Generalized version of ARIMA, used for 'long-memory' forecasting.	(Contrearas-Reyes & Palma, 2013)
Autoregressive integrated moving average with exogenous variables (ARIMAX)	Includes previous values of an exogenous time-series to enhance its performance and accuracy. Able to deal with sudden changes in trends.	(Inman, Pedro, & Coimbra, 2013)
Vector autoregressive (VAR)	Generalizes the AR model by characterizing linear dependences between two or more time-series.	(Hatemi, 2004)
Autoregressive conditional heteroscedasticity (ARCH)	Used for time series with specific variances for the error terms.	(Engle, 1982)

Table 6: Summary of time series methods (Mohamudally, 2018)

This project has received funding from the European Union's Horizon 2020
Research and Innovation Program under Grant Agreement 963530.

The metrics used to determine the errors / residuals for forecast method validation are also listed as:

- Mean squared error (MSE)
- Normalized mean squared error (NMSE)
- Root mean squared error (RMSE)
- Normalized root mean square error (NRMSE)
- Mean absolute error (MAE)
- Normalized mean absolute error (NMAE)
- Mean relative error (MRE)
- Mean bias error (MBE)
- Mean absolute percentage error (MAPE)
- Mean absolute scaled error (MASE)
- Mean square percentage error (MSPE)

Ghofrani and Alolayan concludes with a review of articles that use time series methods for solar radiation and wind speed forecasting, either individually or in hybrid form. The information is provided in tables. The table outlining the articles on solar radiation forecasting includes references to 19 articles spanning a forecast horizon from as short as 5 minutes to as long as a month. The table outlining the articles on wind speed forecasting includes 20 articles with forecast horizons ranging from one hour to one month.

3.2 Time series methods in energy system modelling

The inclusion of renewable energy in energy system models presents two problems. Firstly, finding wind and photovoltaic (PV) power data with sufficient resolution, and, secondly, finding datasets of sufficient size. Typically, many years of data (ideally spanning multiple decades) are required to ensure reliable model outputs. In recent years, long term data has become more readily available through to the use of global reanalysis (Pfenninger, 2017).

Pfenniger compared resampling, clustering, and heuristics as methods to reduce the time resolution of wind and solar PV power time series. The paper attempted to answer two questions: firstly, how accurate are different methods to reduce time resolution and secondly, how can time resolution be reduced in the most efficient way while maintaining scientific accuracy? It concluded that heuristic approaches appear more stable than statistical clustering.

Due to high inter-year variability, long term data sets were required to ensure the reliability of the model results. The analysis was performed with a model of the UK power system which provided 25 years' worth of wind and solar PV data and based on the open-source Calliope high-resolution modeling framework.

There are two main approaches that can be used to include full time series data in energy models: time slices and representative years (or days).

This project has received funding from the European Union's Horizon 2020
Research and Innovation Program under Grant Agreement 963530.

3.2.1 Time slices

The 'time slices' approach selects sets of time steps from the full time series to represent key characteristics such as weekdays and weekends, different times of the day, and different seasons. Energy system modelling software such as TIMES makes use of this technique. The Integrated MARKAL-EFOM System (TIMES) is energy modelling software developed under the Energy Technology Systems Analysis Program (ETSAP) of the International Energy Agency (IEA) that uses linear-programming to produce a least-cost energy solution (optimized within several user constraints) over medium to long-term time horizons. TIMES is available as open source under GNU's Not Unix (GNU) general public license.

3.2.2 Representative years

In the 'representative years' approach, typical years (or days) are artificially constructed from the full time series either to a) cover as much variability as possible, or b) by clustering the data. Both methods aim to preserve the statistical properties of the time series.

3.2.3 Disadvantages

The disadvantages of these methods are the potential loss of concurrency and continuity. Stormy, cold days may increase heating demand and reduce solar PV yields, but simultaneously increase wind power production. If subsets of typical heat demand days are mixed with subsets of typical wind production days, such correlations (and hence concurrency) are typically lost. When picking days randomly to generate a representative year, continuity is lost if the model contains states that need to be carried over from one day to the next, such as the charge of storage facilities, for example.

3.3 Statistical clustering

Statistical clustering is a way to group observations such that the similarity of the observations within a cluster is higher than between clusters. The approach groups the observations on the values of the variables associated with each observation, and there are different methods based on how "similarity" is measured. It is often used when studying demand profiles, but the technique has also been applied to process other inputs. A popular clustering method is the so-called 'k-means' method which aims to partition observations into any number of clusters, such that each cluster contains only observations closest to the cluster center, as shown in Figure 17 (Weston.pace, 2007).

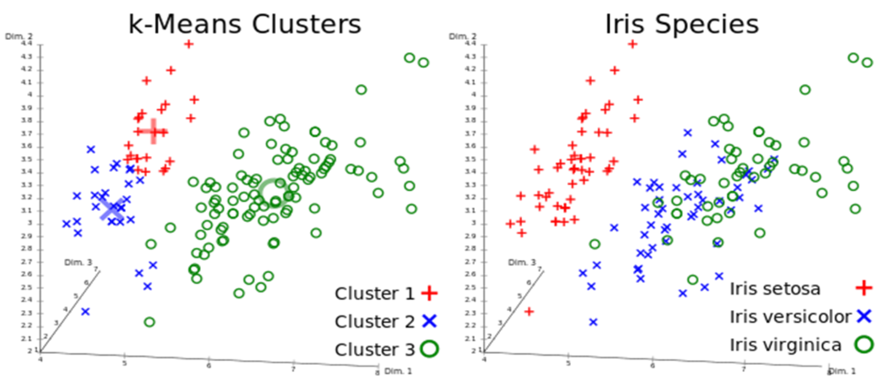

Figure 17: Randomly selected cluster means (left); and final cluster division (right)

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

3.4 Downscaling methods

Climate impact and risk assessments make use of global climate models (GCMs), but these present two key barriers.

Firstly, the spatial resolution of a GCM is usually too coarse for direct use in many applications. A single pixel in a GCM may have a size of 100 x 100 km, with one (average) value assigned for a given characteristic such as elevation, for example. That single pixel might however cover a mountain range or coastline.

Secondly, GCMs are designed to capture global patterns and are valuable for predicting the planet's response as a whole, for example to determine elevated greenhouse gas concentrations. They are not as accurate in predicting climate patterns in any particular location, where they usually display a certain degree of bias. Since these biases differ between models, some models will perform better than others at any selected location.

The information from GCMs often requires further processing to produce a bias-corrected higher resolution dataset. A process called "downscaling" produces such datasets that are spatially and temporally tuned for detailed projection of climate impacts and risks.

Downscaling is the process where coarser datasets are processed to produce higher spatial-temporal resolution output (illustrated in Figure 18). This is typically required to produce datasets at scales that are useful at a regional and local scale. There are two fundamental approaches for the downscaling of coarse-grid output to a finer resolution: dynamical and statistical.

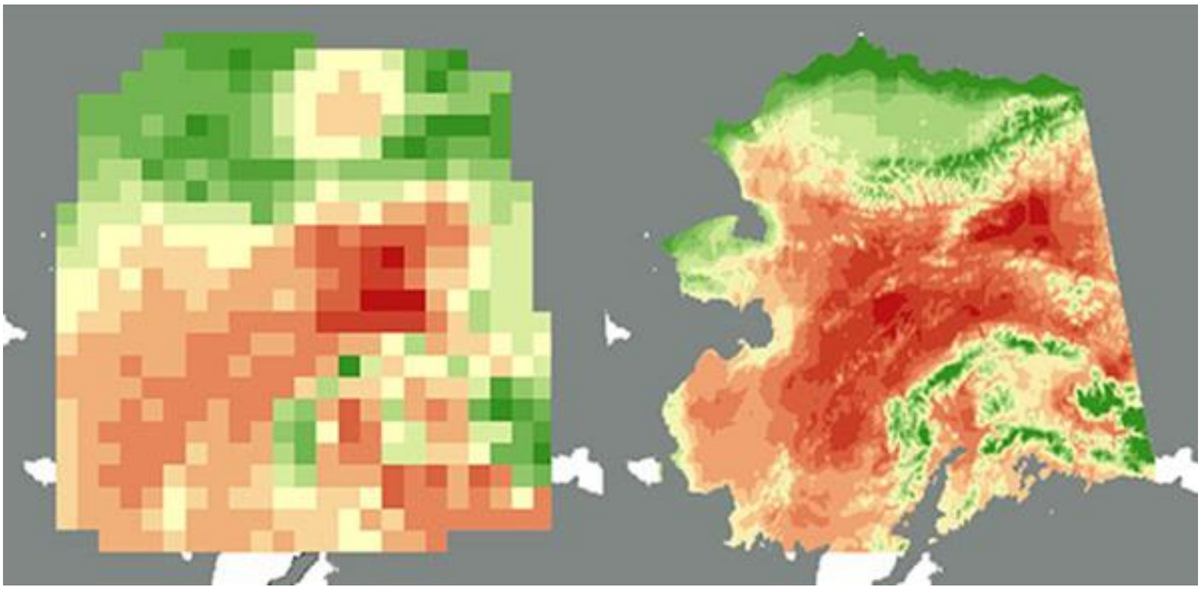

Figure 18: Original data at 0.5° resolution (left) and downscaled to 2 km (right)

3.4.1 Dynamical downscaling

The dynamic approach is taken when a higher resolution climate model is embedded within a lower resolution one, such as a Global Circulation Model (GCM). GCMs are available from several climate centers, but variables such as wind speed are often required at much smaller scales. Computers and programs that run the GCMs are limited to gross representations of the geographic, geologic and atmospheric details and therefore many small-scale features cannot be represented, even though they may significantly impact the

This project has received funding from the European Union's Horizon 2020
Research and Innovation Program under Grant Agreement 963530.

local and regional climate. For example, near-surface wind speeds have particular importance for wind energy resource estimation, but surface wind speeds exhibit variability at much smaller spatial scales than would typically be available from coupled atmosphere-ocean GCMs.

Regional Climate Models (RCMs) are examples of higher resolution climate models. They are sub models, nested in GCMs to produce higher resolution and cover smaller areas. RCMs get their boundary conditions from GCMs and typically simulate climate variables at the 25-50km scale. Unfortunately, RCMs also inherit several of the CGMs' limitations, including the distortion of climate variable distributions. Biases at the station level can be improved by using neural networks and multiple linear regression to further post-process RCM outputs.

It should be noted that RCMs require significant computational capacity. They may cover a smaller area than a GCM, but the grid cells are smaller; more surface information is included, and additional processes are modelled. An RCM might therefore involve a larger number of computations than a GCM that covers the entire globe (Wilby, et al., 2009). The time required to run RCMs depends heavily on its complexity and the computer resources available but is typically in the order of 24 hours for every one year modelled. A model covering 50 years can therefore require several months to compute (ARD, 2014).

3.4.2 Statistical downscaling

Statistical downscaling assumes that the regional climate is conditioned by the large-scale climatic state and local features such as topography, land-sea distribution and land use. It uses statistical methods to establish empirical relationships between GCM output (called predictors) and local small-scale climate variables, often at station level. The latter is referred to as predictands. The GCM outputs must be available where the predictands are located, therefore interpolation of the GCM output is usually needed. In essence, statistical downscaling is a process of training a model to establish a statistical relationship between the GCM dataset and an observational one. This is first done over a historical period before it is employed to generate future predictions.

Statistical downscaling is an attractive alternative to RCMs for the following reasons: (ARD, 2014)

- The methods used are computationally inexpensive in comparison.
- They do not require such a high level of technical expertise to implement.
- Station-scale climate information can be provided instead of at a 20-50 km scale.

The approach is based on the following inherent assumptions:

- The statistical relationship between the predictor and predictand does not change over time (the so-called 'stationarity assumption').
- The predictor carries any relevant climate change signal.
- There is a strong relationship between the predictor and predictand.
- GCMs accurately simulate the predictor.

Downscaling algorithms differ in the following ways:

This project has received funding from the European Union's Horizon 2020
Research and Innovation Program under Grant Agreement 963530.

- Methods can use a single-variate approach, which means the variable is predicted using only one predictor. A multi-variate approach allows any number of time-varying climatological or bio-geophysical variables.
- Different downscaling methods incorporate different levels of spatial information. Some only use local GCM information, while others also refer to other datasets.
- Methods range in maturity from well-established to experimental.

The three major types of statistical downscaling are a) weather classification; b) weather generators; and c) regression models.

3.4.2.1 Weather classification

Weather classification is used for both spatial and temporal downscaling and predicts the value of the local variables based on large-scale atmospheric states. A future atmospheric state is obtained from a GCM and matched with its most similar historical atmospheric state. The value of the local variable is then replicated under the future atmospheric state. Since all possible weather conditions are needed to find a suitable match, a large amount of data is required, typically in the order of 30 years. Analyzing such a large amount of data is computationally demanding in comparison to regression models. Some of the most popular methods associated with weather classification are described below.

The analog method

The large-scale atmospheric circulation from the GCM is compared to historical data and the most similar is taken as its analog. The observed local weather is then associated with the projected large-scale pattern. This method is most appropriate for non-normal distributions, such as daily rainfall, but is incapable of predicting values that are outside the range of the historical record. New extremes associated with climate change will therefore not be predicted.

Cluster analysis

This method, as described in subsection 3.2.3, is also used for weather classification.

Artificial Neural Networks (ANNs)

The name and structure of artificial neural networks are inspired by the human brain, and they function like biological neurons. They are arranged in three types of layers: the input layer, one or more hidden layers, and an output layer, as show in Figure 19. Each neuron connects to another and has a weight and threshold (bias) associated with it. If the output of any individual node is above the specified threshold value, that node is activated, sending data to the next layer of the network. Otherwise, no data is passed along. The weights and biases are adjusted during a training period so that the inputs can be matched to the appropriate outputs.

This project has received funding from the European Union's Horizon 2020
Research and Innovation Program under Grant Agreement 963530.

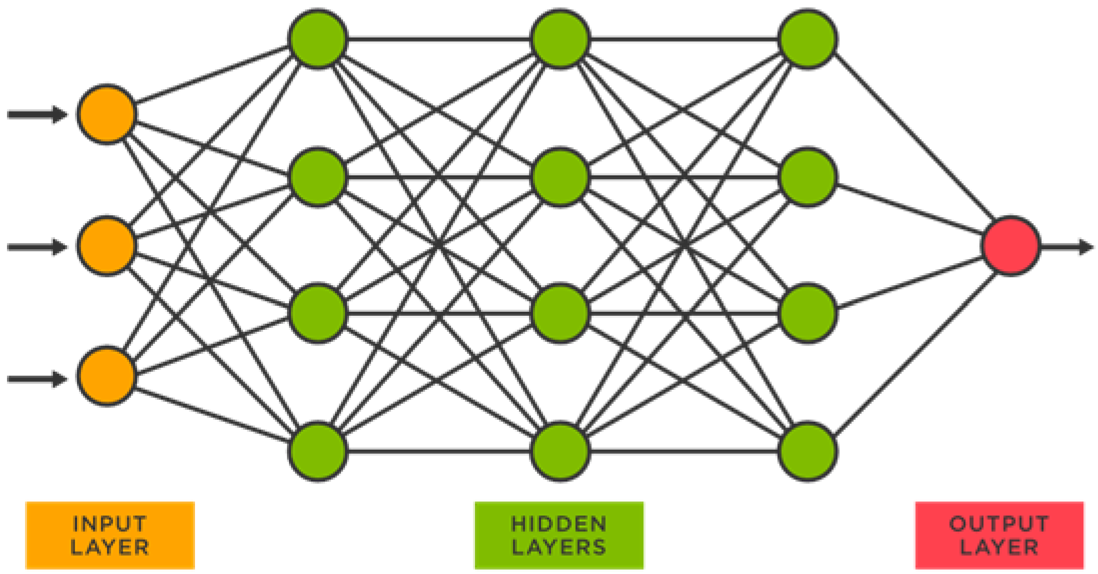

Figure 19: The structure of an artificial neural network (TIBCO, 2022)

Once the network is appropriately trained, high-resolution local climate data can be derived from large-scale circulation information.

Self-Organizing Map (SOM)

This method is an unsupervised machine learning technique used to produce a lower-dimensional representation of a higher dimensional dataset where geometric relationships between points indicate their similarity. Since the lower-dimensional representation is usually two-dimensional, it is often used as a visualization technique. The technique was developed by Finnish professor Teuvo Kohonen in the 1980's. A SOM belongs to the ANN group of methods but uses competitive learning rather than error-correction learning.

Figure 20 shows a simple example where a SOM is used to map three-dimensional data (colours consisting of combinations of blue, red and green) down to a two-dimensional grid. For this data set, a good mapping would group the red, green, and blue colors far away from one another and place the intermediate colors between their base colors (yellow close to red and green, teal close to green and blue, etc.). Figure 20 shows that this result was indeed obtained.

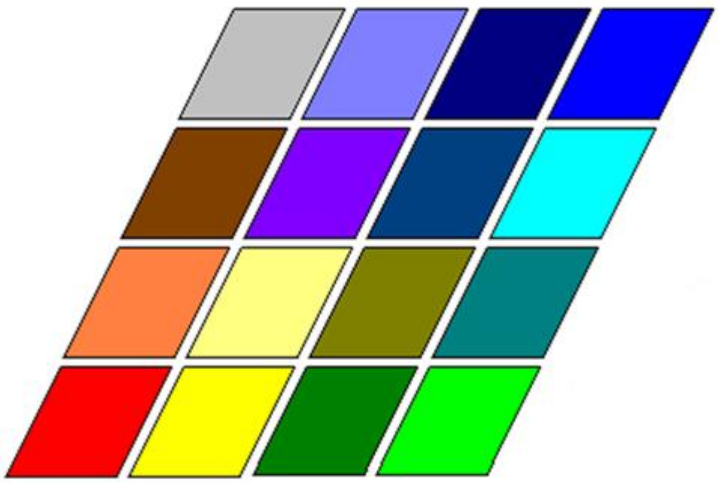

Figure 20: Example of a self-organizing map (Pang, 2003)

This project has received funding from the European Union's Horizon 2020
Research and Innovation Program under Grant Agreement 963530.

This technique can be used to obtain future local-scale conditions from GCM-simulated large-scale future projections.

SOM offers several advantages: (Yin, 2011)

- Distributions can be normal or non-normal (no assumptions are made).
- All data is represented in the model.
- The relationship between nodes is easy to visualize.

3.4.2.2 Weather generators

A weather generator produces sequences of variables at a daily timescale from monthly input. Multiple daily time series can be generated from a single set of monthly input since any number of daily sequences with the same statistical characteristics can be associated with a single larger timescale input set. The method is data intensive, sensitive to missing data and requires many years of input data. In addition, concurrency (see paragraph 3.2.3) must be preserved in the data. Weather generators are typically used for both spatial and temporal downscaling.

The Long Ashton Research Station Weather Generator (LARS-WG)

LARS-WG is a statistical downscaling technique that can be used to generate daily time series at a single site using as little as one year of historical data. The daily time series includes precipitation, temperature, and solar radiation. The method is well validated and generally performs well even in diverse climates (Chisanga, Phiri, & Chinene, 2017).

MarkSim General Circulation Model (GCM)

The MarkSim GCM allows the stochastic generation of daily weather data in three steps:

- downscaling GCM output spatially using the delta method.
- stochastically generating daily time series data using a procedure calibrated on clustered observations.
- selecting an analogue among the clusters that best matches values generated by the GCMs.

The method is described in detail in a paper by Jones and Thornton (Jones & Thornton, 2013) and can be accesses online (ILRI, 2014).

3.4.2.3 Other methods

Regression models

Regression models relate the predictors and predictands using either a linear or non-linear relationship. Linear methods are widely used because they are straightforward to implement, but the simplest methods require that both the predictor and the predictand follow a normal distribution, which may not always be the case with all observed variables. An example of a variable with a non-normal distribution is daily rainfall. The typical pattern of frequent small amounts of rainfall interspersed with a few heavy events makes the distribution non-symmetrical. There are a wide range of methods within the general class of regression models, and some have been developed for specific types of non-normal data such as count data, binary data, and so on.

This project has received funding from the European Union's Horizon 2020
Research and Innovation Program under Grant Agreement 963530.

An example where regression models were used with success is the analysis by (H. Shirkhani, 2013). They used a linear model and two polynomial models to link the large-scale datasets from the National Centers for Environmental Prediction (NCEP) to station-level wind velocity at the Agadez station in Niger. A model was developed for each of the 365 days of the year and calibrated using the 1950-1985 period. The model was then verified using the 1986-1988 period, and good agreement was found between the simulated and observed wind velocity.

MACA, DeepSD and GARD

An organization called CarbonPlan recently released global downscaled climate data spanning multiple downscaling methods (CarbonPlan, 2022). Their current (2022) release includes downscaled datasets produced from multiple statistical downscaling methods. Simulations from six global climate models from the Coupled Model Intercomparison Project - Phase 6 (CMIP6) and three different future emissions scenarios were downscaled using four methods: Multivariate Adaptive Constructed Analogs (MACA), DeepSD, and two implementations of the Generalized Analog Regression Downscaling (GARD) code.

- MACA is a statistical method for downscaling Global Climate Models (GCMs) from their native coarse resolution. It captures observed patterns of daily near-surface meteorology and simulated changes in GCM experiments. It is slightly preferable to direct daily interpolated bias correction in regions of complex terrain due to its use of a historical library of observations and multivariate approach (Climatology Lab, 2022).

- DeepSD is a generalized stacked super resolution convolutional neural network (SRCNN) framework for statistical downscaling of climate variables. DeepSD augments SRCNN with multi-scale input channels to maximize predictability in statistical downscaling (Cornell University, 2017).

- The GARD code was designed to provide a simple statistical downscaling method relying on regressions and statistical transformations from various inputs, such as precipitation; humidity; wind; etc. to various outputs, which include precipitation and temperature. Components of the core algorithms in GARD are documented in (Clark & Hay, 2004) and (Clark & Slater, 2006).

The resulting datasets include projections of daily maximum and minimum temperature and precipitation through to the end of the 21st century for the entire globe at a 0.25° resolution (approximately 25 km x 25 km). All the models were trained with the help of the ERA5 reanalysis product.

Their web browser features an interactive mapping tool to assist with exploring and comparing results directly since the datasets comprise several terabytes of data. The map tool shows downscaled data alongside those from the original CMIP6. The authors state that uncertainty associated with the choice of downscaling method has received less attention than other physical modeling steps. They therefore implemented multiple algorithms to enable comparisons between different methods.

Delta method

The delta method provides a simple way to downscale GCM projections statistically. Two methods of implementation are found in the literature, namely the additive and the multiplicative implementation.

The additive implementation is shown in Figure 21. The difference between present and future simulations is added to the present observation. In the figure the mean difference

This project has received funding from the European Union's Horizon 2020
Research and Innovation Program under Grant Agreement 963530.

between present simulation (green) and future simulation (red) is calculated and added to the present observation (blue) to make a downscaled future prediction.

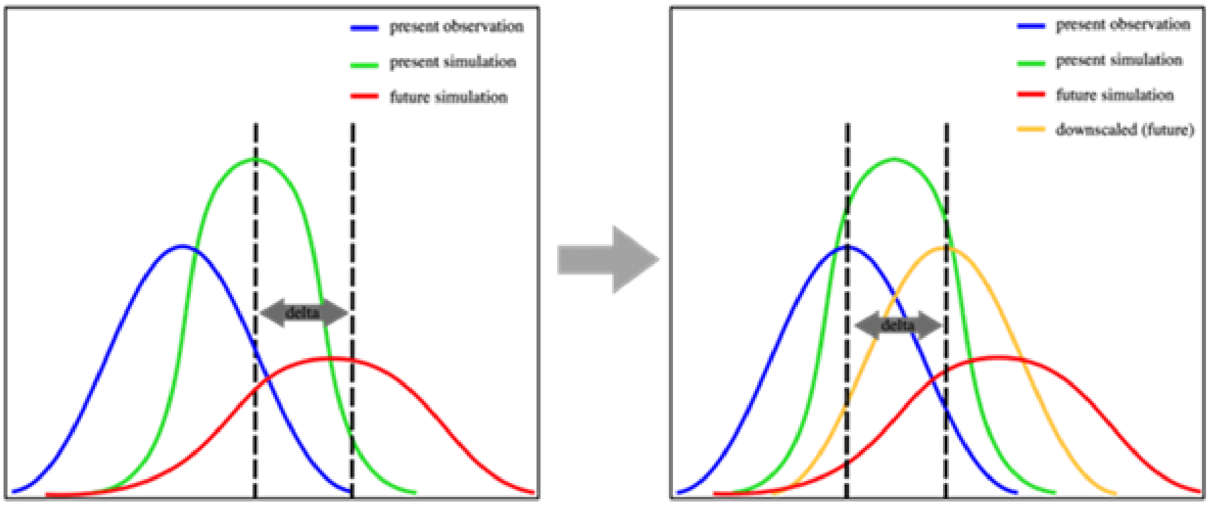

Figure 21: The Delta method (additive) (Regional Climate Model Evaluation System, 2022)

The multiplicative implementation takes the ratio between GCM simulations of the future and current climate and then multiplies it with the currently observed regional-scale value to obtain future regional values (see formula below).

$$T_f = \frac{T^F}{T^C} T_c$$

Where T_f is the future regional value; T_c is the current regional value; T^F is the future GCM value; and T^C is the current GCM value.

Hamlet et al (Hamlet, Salathé, & Carrasco, 2022) used both methods in their research. The multiplicative implementation was used for precipitation to avoid potential sign problems and the additive implementation was used for temperature to avoid problems with division by zero on the centigrade scale.

3.5 Other useful methods and measures

Genetic Programming (GP)

Genetic Programming is a relatively new method that is used to generate a set of interactive computer programs. The concept was proposed by Alan Turing in 1950 but was not successfully implemented until the 1980s. John Koza patented the first algorithm in 1988 and remains a leader in this field of research (Virtusa, 2022). The concept is modelled on biological evolution where an initial set of 'unfit' programs are continuously improved through stochastic variation (Banzhaf, 2002). The process is iterative, where only programs that satisfy prespecified criteria are allowed to survive. Those that produce solutions with acceptable quality are crossed to produce the next generation. Genetic programming will generally terminate once it reaches a predefined fitness measure. The latest research in this field can be found in the proceedings of dedicated conferences and journals.

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

Canonical correlation analysis (CCA)

For two sets of variables with correlations among the variables, a canonical-correlation analysis will find linear combinations of variables from each set that have the maximum degree of correlation with each other. The main assumption is that the relationship between predictand and predictor groups is linear and that they are normally distributed. It should also be noted that if co-variance is detected, it does not imply causality. A co-variance between two variables might be caused by the fact they are both correlated with a third variable. Therefore, results should be interpreted with knowledge of the underlying physical relationships.

Cumulative distribution functions (CDFs)

A cumulative distribution function is obtained by summing or integrating the probability density function, and it provides the probability that a random variable takes on a value less than or equal to any given value. While the probability density function can be used to obtain the probability that a discrete, continuous, or mixed variable takes on a specific value (i.e., by solving the probability density function for that value), when the variable is continuous it is difficult to obtain one exact probability value (must the probability be obtained for a value of 3.0, or for a value of above 3.0 but less than 3.5?). In such a case it is better to use the cumulative distribution function to determine the probability of the value falling above or below a certain value, or within a particular interval.

This project has received funding from the European Union’s Horizon 2020
Research and Innovation Program under Grant Agreement 963530.

4. Software

Soliciting forecasts from numerical weather prediction (NWP) models has hitherto been crucial to energy forecasting practices, such as solar, wind, or electricity load forecasting. Owing to their capacity for capturing the dynamics of the weather system, NWP models can cover forecast horizons up to a few days with a reasonable accuracy, which is unparalleled by other forecasting methods.

GFS

- The Global Forecast System (GFS) is a global numerical weather prediction system containing a global computer model and variational analysis run by the United States' National Weather Service (NWS)

- The mathematical model is run four times a day and produces forecasts for up to 16 days in advance, but with decreased spatial resolution after 10 days. The forecast skill generally decreases with time (as with any numerical weather prediction model) and for longer term forecasts, only the larger scales retain significant accuracy. It is one of the predominant synoptic scale medium-range models in general use.

- As with most works of the U.S. government, GFS data is not copyrighted and is available for free in the public domain under provisions of U.S. law. Because of this, the model serves as the basis for the forecasts of numerous private, commercial, and foreign weather companies.

ECMWF

The European Centre for Medium-Range Weather Forecast (ECMWF) is one of the (if not the) best NWP models to date. Therefore, energy forecasters can leverage this state-of-the-art dataset in their investigations on various research problems. The geographical extent covers most of Europe and North America. A total of 14 weather variables that are relevant to energy forecasting are included in the dataset, such as surface solar radiation downwards, U and V wind components, and surface temperature. To ensure the uptake of this dataset, example R scripts are provided, giving demonstrations on how the data can be read, processed, and used in energy forecasting applications. Note that:

- The European Centre for Medium-Range Weather Forecasts (ECMWF) is an independent intergovernmental organization supported by most of the nations of Europe.
- They produce numerical weather forecasts and monitor planetary systems that influence weather.
- They carry out scientific and technical research to improve forecasting skills.
- They also maintain an archive of meteorological data.

To deliver their core mission, the Centre provides:

- Twice-daily global numerical weather forecasts
- Air quality analysis
- Atmospheric composition monitoring
- Climate monitoring
- Ocean circulation analysis
- Hydrological prediction

This project has received funding from the European Union's Horizon 2020
Research and Innovation Program under Grant Agreement 963530.

IBM Global High-Resolution Atmospheric Forecasting

- IBM GRAF is a high-precision, rapidly updating, global weather model that updates hourly and at a 3km resolution to provide a picture of weather activity around the globe.
- It was the first operational global weather model to run on a GPU-accelerated supercomputer.
- It assimilates weather data rapidly into forecasts.
- The approach is claimed to provide improved global mapping of the atmosphere through a partnership with the National Center for Atmospheric Research (NCAR) to improve its latest generation global weather model.
- It uses very good data and observations. IBM GRAF has the capability to incorporate previously untapped data sources to help overcome the lack of specialized weather equipment in many parts of the world.
- GRAF achieves what has been something of a "holy grail" for global numerical weather forecasting, namely the ability to run at a resolution so fine that no approximations are needed to simulate how individual thunderstorms behave.

ClimaCell's Bespoke Atmospheric Model (CBAM)

ClimaCell Inc. recently introduced the ClimaCell Bespoke Atmospheric Model, an automated, on-demand, scalable, cloud computing service for operational NWP modeling. The spatial and temporal gaps in the publicly accessible weather observations are filled by CBAM using private observations, which is important in areas with short data cycles and quick forecast cycles.

- It uses one of the best available public models, NOAA's High-Resolution Rapid Refresh (HRRR), but works in resolutions of 3 km. CBAM gets to the highest resolution that exists in the market today: tens of meters. This resolution allows them to account for things like terrain and buildings and see otherwise invisible features, such as wind turbulence around buildings.
- CBAM can provide updates every few minutes.
- HRRR and similar models are only available in a few developed countries, while most of the world remains without access to such forecasts. However, CBAM can operate anytime, anywhere, from a single wind farm to the entire Indian subcontinent.

Weather Research and Forecasting (WRF)

Simulations of the atmosphere are produced by WRF. The process is broken down into several stages: setting up the model domains, ingesting the input data, constructing the initial circumstances, and using the forecast model. The forecast model component operates within the WRF software framework, which controls I/O and communications amongst parallel processors. WRF is mostly written in Fortran and may be compiled using a variety of compilers. It runs primarily on platforms with UNIX-like operating systems, including laptops and supercomputers (Jordan G. Powers, 2017)

WRF simulations begin with the WRF Preprocessing System (WPS). To set up the user's model domains, the WPS, a group of utilities, first collects geographic data (such as topography and land use). The required first-guess atmospheric data is then taken in,

This project has received funding from the European Union's Horizon 2020
Research and Innovation Program under Grant Agreement 963530.

reorganized, and extrapolated to the user's domains (such as a worldwide study or model forecast). Finally, input fields are added to the model's vertical levels, and lateral boundary conditions are generated. WRF is then ready to begin its analysis. This is achieved by the forecast component, which consists of a dynamic solver and physics packages for atmospheric processes (e.g., microphysics, radiation, planetary boundary layer). WRF may also be run as a global model on a latitude–longitude grid. While Global WRF was originally built to study planetary atmospheres, it has come to be used for terrestrial forecasting, chemistry, and climate, to initialize ensembles of WRF with convection-permitting resolutions (e.g., 3-km grids).

Regional Climate Model

The Regional Climate Model (RCM), also known as the RegCM, was created over a long time. It has been used in many projects comparing regional models, and a broad community has used it for a variety of regional climate studies, from process studies to paleoclimate and fully developed future regional climate projections (Ozturk, 2018).

The RegCM modeling system has four components: Terrain, ICBC, RegCM, and Postprocessor. Terrain and ICBC are the two components of RegCM pre-processor. Terrestrial variables (including elevation, land use and sea surface temperature) and three-dimensional meteorological data are horizontally interpolated from a latitude and longitude mesh to a high-resolution domain on either a Normal or Rotated Mercator, Lambert Conformal, or Polar Stereographic Projection. Vertical interpolation from GCM levels to the vertical coordinate system of RegCM is also performed. Since the vertical and horizontal resolution and domain size can vary, the modeling package’s programs employ parameterized dimensions requiring a variable amount of core memory, and the requisite hard-disk storage amount is varied accordingly (Ozturk, Turpb, Türkeşd, & Kurnaz, 2018).

The RegCM model solves a set of primitive dynamic equations describing the atmospheric motion, with parametrizations for physical processes as per:

- Radiation (Short Wave and Long Wave)
- Convection
- Turbulent Diffusion
- Moist (Clouds and Precipitation)
- Fluxes exchanged with surface (Soil model and Ocean fluxes)
- Tracer transport and chemistry (Aerosols and full chemistry)

The dynamic equations are discretized using a finite differences technique on a three-dimensional computation grid with fixed horizontal resolution and terrain following vertical coordinate. The model has three dynamic cores:

- Hydrostatic equation solver
- Non-hydrostatic equation solver with pressure coordinates
- Non-hydrostatic equation solver with height coordinates

The primitive equations for the three solvers are different and have different prognostic variables used to identify the atmospheric state.

ICON modelling framework

This project has received funding from the European Union’s Horizon 2020
Research and Innovation Program under Grant Agreement 963530.

The Icosahedral Nonhydrostatic (ICON) Modelling Framework is being developed in a collaboration between the German Weather Service (DWD) and Max-Planck Institute for Meteorology (MPI-M) to establish a unified modelling system for numerical weather prediction and climate modelling (Zängl, 2013).The planning started with compiling a list of mandatory and/or desirable properties of the common modelling system, with the most important of these summarized here:

- Development of a nonhydrostatic dynamic core to allow using the model over the full range of spatial resolutions relevant for NWP.

- Exact numerical conservation of atmospheric variables that are conserved in nature: this goal was formulated as mandatory for mass and as desirable for energy. In addition, mass-consistent transport of tracer variables was requested, which is particularly crucial for chemistry modelling because chemical reactions can depend very sensitively on the relative mixing ratios of the participating constituents, but also important for NWP because spurious sources or sinks in the moisture budget potentially deteriorate the forecast quality.

- Conservative atmosphere-ocean coupling: achieving this goal without highly sophisticated remapping algorithms requires computing the atmosphere and ocean components of the modelling system on conforming grids. Therefore, an ocean circulation model using (basically) the same grid structure as the atmospheric component is being developed as well.

Korean Integrated Model (KIM)

The Korean Integrated Model a recently developed nonhydrostatic global atmospheric model based on a cubed-sphere grid. It was deployed in April 2020 as an operational weather forecasting model at the Korea Meteorological Administration (KMA). The model has 91 vertical levels, with the highest level at 0.01 hPa. The KIM physics package includes a gravity wave drag (GWD) parameterization developed by Kim and Arakawa. KIM uses the hybrid-4DEnVar approach as the core data assimilation method.

Indian Institute of Tropical Meteorology Earth System Model

Earth Systems Models (ESMs) are useful for enhancing a fundamental understanding of the climate system, its multi-scale variability, global and regional climatic phenomena and making projections of future climate change (Krishnan, et al., 2019). The Indian Institute of Tropical Meteorology Earth System Model includes the following:

- An atmospheric general circulation model, which is a global spectral model with triangular truncation of 62 waves (T62, grid size ~200 km), and consists of 64 vertical levels with top model layer extending up to 0.2 hPa.

- A land surface model (Noah LSM) with four layers.

National Centre for Medium Range Weather Forecasting

The NCMRWF Unified model (NCUM) analysis-forecast system has used for Numerical Weather Prediction (NWP) for nearly a decade. The NCUM system is made up of components for observation pre-processing, observation processing and quality control, data assimilation, forecasting model, and tools for post processing.

- A hybrid 4D-Var data assimilation method is used. Flow dependent error information is provided by a 6-hour forecast from the high resolution global NEPS (N1024, ~12 km resolution) for all the four data assimilation cycles.

This project has received funding from the European Union's Horizon 2020
Research and Innovation Program under Grant Agreement 963530.

- Data assimilation is done at N320L70 resolution (~40 km horizontal resolution with 70 vertical levels reaching the height of 80 km) with N144L70 Hessian based preconditioning.
- A six hourly data assimilation cycle (time window of the assimilation) is used, centered at 00, 06, 12 and 18 Coordinated Universal Time (UTC).
- A variational bias correction is also used for satellite radiance observations.

Atlite

Atlite is an open-source Python library based on xarray that converts weather data into energy systems data. It can accept weather data from a server or file and can use a predefined cut-out to load the required variables. Time series data for renewable energy technologies like wind turbines or solar photovoltaic panels can be generated based on detailed mathematical models. It is designed to work with big data sets while maintaining low computational requirements.

5. Conclusion

In this work package we aim to generate high resolution energy time series data from weather data. Based on this literature study there are currently very limited weather or climate data available at an appropriate resolution. Even though some countries do have high resolution data, it might be outdated or incomplete. It is evident that whichever weather or climate data will be used for this study will have to be downscaled to a higher resolution to generate the energy potential time series at the target resolution.

Based on the scale of this project and its timelines, it was decided to make use of statistical rather than dynamic downscaling due to the limited computational resources available. This will also make the approach more accessible to a larger audience.

Atlite has been identified as the tool to generate time series data since it is open source and can accept different weather datasets. Currently it accepts ERA5 and Sarah-2 and will need to be adapted to accept downscaled datasets supplied by the user.

This project has received funding from the European Union's Horizon 2020
Research and Innovation Program under Grant Agreement 963530.

Bibliography

ARD, T. T. (2014). *A Review of Downscaling Methods for Climate Change Projections.* Burlington: United States Agency for International Development.

Banzhaf, W. (2002, 11 2). Artificial Intelligence: Genetic Programming. Dortmund, Germany.

CarbonPlan. (2022, June 30). *Open data and tools for multiple methods of global climate downscaling*. Retrieved from Data and science for climate action: https://carbonplan.org/research/cmip6-downscaling-explainer

Chisanga, C. B., Phiri, E., & Chinene, V. R. (2017). Statistical Downscaling of Precipitation and Temperature Using Long Ashton Research Station Weather Generator in Zambia: A Case of Mount Makulu Agriculture Research Station. *American Journal of Climate Change*, 6, 487-512.

Clark, M. P., & Hay, L. E. (2004). Use of medium-range numerical weather prediction model output to produce forecasts of streamflow. *Journal of Hydrometeorology*, 5(1), 15–32.

Clark, M. P., & Slater, A. G. (2006). Probabilistic quantitative precipitation estimation in complex terrain. *Journal of Hydrometeorology*, 7(1), 3-22.

Climatology Lab. (2022, November 1). *MACA*. Retrieved from Climatology Lab: https://www.climatologylab.org/maca.html

Contrearas-Reyes, J., & Palma, W. (2013). Statistical analysis of autoregressive fractionally integrated. *Computational Statistics*, 28(5):2309-2331.

Cornell University. (2017, March 9). *DeepSD: Generating High Resolution Climate Change Projections through Single Image Super-Resolution*. Retrieved from Computer Science > Computer Vision and Pattern Recognition: https://arxiv.org/abs/1703.03126

Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 50(4):987-1007.

Global Solar Atlas. (n.d.). *Data Outputs*. Retrieved from https://globalsolaratlas.info/support/data-outputs?c=11.523088,8.4375,3

Global Solar Atlas. (n.d.). *Getting Started*. Retrieved from https://globalsolaratlas.info/support/getting-started

Global Solar Atlas. (n.d.). *Global Solar Atlas*. Retrieved from https://globalsolaratlas.info/map?c=11.609193,8.4375,3

Global Wind Atlas. (n.d.). *Introduction*. Retrieved from Global Wind Atlas: https://globalwindatlas.info/en/about/introduction

Global Wind Atlas. (n.d.). *Key Features*. Retrieved from Global Wind Atlas: https://globalwindatlas.info/en/about/KeyFeatures

Global Wind Atlas. (n.d.). *Purpose*. Retrieved from https://globalwindatlas.info/en/about/purpose

This project has received funding from the European Union's Horizon 2020 Research and Innovation Program under Grant Agreement 963530.

Global Wind Atlas. (n.d.). *Validation*. Retrieved from https://globalwindatlas.info/en/about/validation

H. Shirkhani, O. S. (2013). Downscaling Method for Wind Data: Case Study of Agadez in Niger. *3rd Specialty Conference on Disaster Prevention and Mitigation.* Montréal, Québec.

Hamlet, A. F., Salathé, E. P., & Carrasco, P. (2022, 12 08). *Statistical Downscaling Techniques for Global Climate Model Simulations of Temperature and Precipitation with Application to Water Resources Planning Studies.* Retrieved from digital.lib: https://digital.lib.washington.edu/researchworks/bitstream/handle/1773/38428/2010-5.pdf?sequence=1

Hatemi, A. (2004). Multivariate tests for autocorrelation in the stable and unstable VAR models. *Economic Modelling*, 21(4):661-683.

ILRI. (2014, 1 1). *MarkSim DSSAT weather file generator*. Retrieved from MarkSim DSSAT weather file generator: https://gisweb.ciat.cgiar.org/MarkSimGCM/

Inman, R., Pedro, H., & Coimbra, C. (2013). Solar forecasting methods for renewable energy integration. *Progress in Energy and Combustion Science*, 39:535-576.

Jones, P. G., & Thornton, P. K. (2013). Generating downscaled weather data from a suite of climate models for agricultural modelling applications. *Agricultural Systems*, 1-5.

Jordan G. Powers, J. B. (2017). The Weather Research and Forecasting Model: Overview, System Efforts, and Future Directions. *BAMS*.

Krishnan, R., Swapna, P., Vellore, R., Narayansetti, S., Prajeesh, A., Choudhury, A., . . . Sanjay, J. (2019). The IITM Earth System Model (ESM): Development and Future Roadmap. In D. Randall, J. Srinivasan, R. Nanjundiah, & R. Mukhopadhyay, *Current Trends in the Representation of Physical Processes in Weather and Climate Models.* Singapore: Springer Nature Singapore Pte Ltd.

Mohamudally, N. (2018). *Time Series and Renewable Energy Forecasting.* online publication: IntechOpen. Retrieved from https://www.intechopen.com/books/6138

Ozturk, T. ,. (2018). Future projections of temperature and precipitation climatologyfor CORDEX-MENA domain using RegCM4.4. *Atmospheric Research*, 87-107.

Pang, K. (2003, September 1). *Self-organizing Maps*. Retrieved from Self-organizing Maps: https://www.cs.hmc.edu/~kpang/nn/som.html

Pfenninger, S. (2017). Dealing with multiple decades of hourly wind and PV time series in energy models: A comparison of methods to reduce time resolution and the planning implications of inter-annual variability. *Applied Energy*, 197:1-13.

POWER. (n.d.). *Data Access Viewer Enhanced (DAVe)*. Retrieved from https://power.larc.nasa.gov/beta/data-access-viewer/

Regional Climate Model Evaluation System. (2022, 12 8). Retrieved from JPL(NASA): https://rcmes.jpl.nasa.gov/content/statistical-downscaling

The NASA POWER Team. (2022, 3 4). *About the Prediction Of Worldwide Energy Resources (POWER) Project*. Retrieved from ArcGIS StoryMaps: https://storymaps.arcgis.com/stories/52116d331ff64e468fe9351fc1c76423

The NASA Power Team. (n.d.). *Data Sources*. Retrieved from NASA POWER: https://power.larc.nasa.gov/docs/methodology/data/sources/

TIBCO. (2022, November 27). *What is a Neural Network?* Retrieved from TIBCO: https://www.tibco.com/reference-center/what-is-a-neural-network

University of New Mexico. (2006, 9 17). *AR, MA and ARMA models*. Retrieved from Department of Mathematics & Statistics: https://www.math.unm.edu/~ghuerta/tseries/week4_1.pdf

Virtusa. (2022, 12 8). *Digital Themes*. Retrieved from Genetic Programming: https://www.virtusa.com/digital-themes/genetic-programming

Wang, W., & Niu, Z. (2009). Time series analysis of NASDAQ composite based on seasonal ARIMA. *Proceeding of the International Conference on Management and Service Science* (pp. 1-4). Wuhan: Curran Associates.

Weston.pace. (2007, July 26). *k-means clustering*. Retrieved from Wikipedia: https://en.wikipedia.org/wiki/K-means_clustering

Wilby, R. L., Troni, J., Biot, Y., Tedd, L., Hewitson, B. C., Smith, D. M., & Sutton, R. T. (2009). A review of climate risk information for adaptation and development planning. *International Journal of Climatology*, 29(9), 1193-1215.

Yin, C. (2011, December 1). *Applications of Self-Organizing Maps to Statistical Downscaling of Major Regional Climate Variables.* Retrieved from Research Commons: https://researchcommons.waikato.ac.nz/handle/10289/5733

Zängl, G. (2013). ICON: The icosahedral nonhydrostatic modelling framework of DWD and MPI-M. *Proc. ECMWF Seminar on Numerical Methods for Atmosphere and Ocean Modelling*.

Zhang, T., Chandler, W. S., Hoell, J. M., Westberg, D., Whitlock, C. H., & Stackhouse, P. W. (2008). A Global Perspective on Renewable Energy Resources: Nasa’s Prediction of Worldwide Energy Resources (Power) Project. *Proceedings of ISES World Congress 2007 (Vol. I – Vol. V), I-V*, 2636-2640.

This project has received funding from the European Union’s Horizon 2020 Research and Innovation Program under Grant Agreement 963530.