Solar Energy 315 (2026) 114785

Contents lists available at ScienceDirect

Solar Energy

journal homepage: www.elsevier.com/locate/solener

Photovoltaic power modelling in high latitudes with empirical and machine learning models using transfer learning

Väinö Anttalainen [a,*], Lauri Karttunen [a], Sami Jouttijärvi [a], Anders V. Lindfors [b], Juha A. Karhu [b], Hugo Huerta [c], Samuli Ranta [c], Tarmo Lipping [d], Kati Miettunen [a]

[a] Department of Mechanical and Materials Engineering, University of Turku, Vesilinnantie 5, Turku, 20500, Finland
[b] Finnish Meteorological Institute, Erik Palménin aukio 1, Helsinki, 00560, Finland
[c] Department of Energy Systems and Vehicles, Turku University of Applied Sciences, Joukahaisenkatu 7, Turku, 20520, Finland
[d] Faculty of Information Technology and Communication, Tampere University, Pohjoisranta 11 A, Pori, 28100, Finland

HIGHLIGHTS

- Models are compared as general and site-specific fine-tuned models.
- Comparison is conducted using five multi-year datasets across Finland.
- Training the general models with data from other systems reduces error trends.
- Fine-tuning results in similar accuracy for the models.
- Machine learning models demonstrate better low-irradiance accuracy than empirical models.

ARTICLE INFO

Keywords:
Power modelling
Empirical models
Machine learning
High latitudes
General models
Fine-tuned models

ABSTRACT

Accurate power models for photovoltaics (PV) are essential to ensure expected system performance, thus avoiding economic losses and reductions in resource efficiency. Currently, insights are lacking on how reliably models perform in situations for which historical data are not available and how accurately they cover locations with high seasonality, such as the Nordics, where PV capacity is growing rapidly. Importantly, besides long summer days, in winter months, the Nordic areas experience low irradiances, for which power modelling is poorly understood but should be expanded to enable continuous monitoring. To fill current gaps in the literature, this work compares various empirical models (Huld, PVWatts, PVUSA) and investigates how significant improvements can be made with machine learning (ML) models (multilayer perceptron, gradient boosting). The models are analysed as general models, which can be applied directly to new systems, and as site-specific fine-tuned models, for which previous data from that system are required. The data include multi-year power output and on-site weather measurements from five systems across Finland with varying installations. Novel insights include that utilising transfer learning with high-quality data resulted in R^2 values of 0.944–0.994, and superior accuracy compared to the default models with minimal filtering that had R^2 values even as low as 0.78. Fine-tuning with site-specific data increased the R^2 values to 0.966–0.997. Additionally, fine-tuned ML models had lower nRMSE and MAE values than empirical models at low irradiance levels, highlighting their potential for low-light monitoring.

1. Introduction

To mitigate climate change, an increase in sustainable energy is needed while reducing the use of fossil-based energy. Photovoltaic (PV) technology is one option for this, as costs have decreased rapidly [1]. PV is also gaining popularity in Nordic countries, including Finland. For example, in August 2025, there were 313 PV sites with capacities exceeding 1 MW, under construction or in the planning stage, with a total capacity of 26.18 GW [2]. If realised, these sites will increase the total PV capacity in Finland around 16-fold. This increased interest in PV has resulted in a booming demand for research on various aspects of PV in high-latitude conditions.

* Corresponding author.
Email address: vaino.anttalainen@utu.fi (V. Anttalainen).

https://doi.org/10.1016/j.solener.2026.114785
Received 4 February 2026; Received in revised form 29 April 2026; Accepted 7 June 2026
Available online 15 June 2026

Nomenclature

Abbreviations

PV Photovoltaic
ML Machine learning
MLP Multi-layer perceptron
GBM Gradient boosting model
TL Transfer learning
LSTM Long short-term memory
PVUSA Power model by the Photovoltaics for Utility Scale Applications project
LOSO-CV Leave-One-Site-Out Cross-Validation
POA Plane-of-array
FMI Finnish Meteorological Institute
TUAS Turku University of Applied Sciences
SI Supplementary Information
UTU University of Turku
HEL System in Helsinki
KUO System in Kuopio
SOT-20 System in Sodankylä with 20-degree panel tilt
SOT-90 System in Sodankylä with 90-degree panel tilt
TKU System in Turku
Wp Watt-peak
AC Alternating current
DC Direct current
GHI Global horizontal irradiance, W/m^2
DHI Diffuse horizontal irradiance, W/m^2
DNI Direct normal irradiance, W/m^2
AOI Angle of incidence of the direct solar beam
QC Quality control
IEC International Electrotechnical Commission
SD Standard deviation
PLR Performance loss rate, %/year
NREL National Renewable Energy Laboratory
nRMSE Normalised root mean squared error
MAE Mean absolute error
def Default model
gen General model
fit Fine-tuned model
IQR Interquartile range

Symbols

V Voltage, V
I Current, A
G_{POA} Plane-of-array irradiance, W/m^2
WS Wind speed, m/s
T_{air} Air temperature, °C
T_m Panel temperature, °C
T_{cell} Cell temperature, °C
B_{POA} Beam component of plane-of-array irradiance, W/m^2
D_{POA} Sky-diffuse component of plane-of-array irradiance, W/m^2
R_{POA} Ground-reflected component of plane-of-array irradiance, W/m^2
α_{albedo} Albedo of the ground
β_{tilt} Tilt of the panel, °
$\mathrm{DNI}_{\mathrm{QClimit}}$ Quality control limit for maximum direct normal irradiance, W/m^2
α Solar elevation, °
$\mathrm{DNI}_{\mathrm{QC}}$ Direct normal irradiance to which quality control method is applied, W/m^2
G_{eff} Reflection corrected plane-of-array irradiance value, W/m^2
ΔT Temperature difference between panel and cell, °C
$a_{\mathrm{cell}}, b_{\mathrm{cell}}$ Parameters for cell temperature calculation
P_0 Nominal power of the system, W$_{\mathrm{p}}$
P_{AC} Alternating current power, W
P Direct current power, W
$P_{\mathrm{without\_plr}}$ Power without performance losses, W
$P_{\mathrm{with\_plr}}$ Power with performance losses, W
Δt Years from initial date
G'_{POA} Plane-of-array irradiance divided by the reference plane-of-array irradiance
T' Panel temperature with reference panel temperature subtracted, °C
$k_1, \dots, k_6$ Coefficients for the Huld model
$k'_1, \dots, k'_6$ Coefficients for the Huld model divided by the nominal power of the system
a, b, c, d Coefficients for the PVUSA model
a', b', c', d' Coefficients for the PVUSA model divided by the nominal power of the system
C_T Temperature correction term
γ Temperature coefficient of the panel, 1/°C
R^2 Coefficient of determination
n Total number of observations
i ith observation
y_i Actual value of the ith observation
$\hat{y}_i$ Predicted value of the ith observation
$\bar{y}$ Mean of the actual observations

Subscripts and superscripts

meas Measured value
modelled Modelled value
rc Reflection-corrected irradiance component

High-latitude conditions create an opportunity for a variety of high-performing solar installations with various production profiles [3–5]. However, the conditions consist of factors that make PV system performance assessment challenging. Strong seasonality, with long summer days and cold, snowy winter months, directly motivates studies on the effects of device temperature [6] and snow [7]. Snow can either decrease power production when it covers the panels or increase power production due to its high albedo. In addition to the weather effects, the sun elevation angle in high latitudes is often low. Low sun elevation angles result in reflection losses and shading from surrounding buildings, trees, and other panels. These effects pose challenges for reliable PV output modelling [8] and forecasting [9,10], as well as calculating performance loss rates [11]. One understudied aspect of high-latitude PV systems is their reliable monitoring, especially in different irradiance conditions. High accuracy during low light conditions must be addressed to achieve continuous system monitoring and to detect performance flaws even during winter.

PV system monitoring is important, as ensuring peak system performance can secure economic gains. Low-performing systems lead to decreased resource and land use efficiency, and consequently reduce environmental and financial benefits. The importance of real-time and automatic monitoring increases with PV system size as manual inspection becomes increasingly time-consuming and economically prohibitive. One method for automatic system monitoring and fault detection is to compare actual power output with expected power output [12–16]. Øgaard et al. showed that such a comparison was a more suitable monitoring approach than specific yield or performance ratio-based approaches at high latitudes when an ML model was used [14]. Other models that were not data-driven produced less stable metrics, highlighting the benefits of site-specific models in challenging weather conditions. However, when historical data are not available, site-specific models become unfeasible, and other modelling approaches must be used, creating a need to find the most suitable alternatives.

The challenge is to reliably detect faults in highly seasonal high-latitude conditions, especially if site-specific models are not available. However, knowledge of which power models are the most reliable in these conditions is unavailable due to the lack of a comprehensive comparison of multiple models. Modelling can be carried out using various models, including IV-curve-based models [17], empirical models [17], statistical models, and machine learning (ML) models [18,19]. Our work focuses on empirical models due to the frequency of their use in the relevant high-latitude literature [8,9,11,14,20]. In addition, our work analyses multi-layer perceptron (MLP) and gradient boosting (GBM) ML models due to their ability to map non-linear dependencies, such as ambient conditions to power output, making them well-suited for PV power modelling [14,21–23]. The data-driven nature of empirical and ML models makes their usage and comparison similar.

The distinct characteristics of high-latitude conditions (such as periods of cold temperatures, low irradiance, and snow coverage) create a need to validate power models in these conditions, which is currently limited to only a few models. Alcañiz et al. showed that the accuracy of predictive power models decreases if they are used on data from different climates than those on which they were trained [24]. Although their study focuses on ML models, it highlights the need for model validation in local climates. Similarly, Hajjaj et al. highlight the need to compare the accuracy of several models in the same climate to find the most suitable model for that climate [25].

Previous literature related to high-latitude conditions is limited to studies of one or two models [8,9,12], using data from one system [12], or using a dataset that is less than a year long [9,12]. These shortcomings can easily impact reliability and generalisability and indeed previous literature has not reached a consensus. Böök et al. showed that the empirical Huld model works well for monitoring purposes in Finnish conditions using data from two systems [8]. Nevertheless, Brester et al. showed that the MLP model outperformed the Huld model in day-ahead output forecasting using some of the same data [9]. In 2015, Platon et al. compared the PVSAT-2 empirical model to an MLP model in Canada using data from a 120 kW system covering March to July 2014 [12]. Even though the MLP showed better accuracy, the difference from the empirical model was so small that the authors preferred the simplicity of the one-equation empirical model over the more complex MLP model. Øgaard et al. showed that the random forest ML model was more accurate than the empirical PVWatts and single diode models using six rooftop systems from Norway [14], highlighting the benefits of data-driven models in high-latitude conditions. This work aims to bridge the gap and identify generalisable power modelling features in high-latitude conditions by comparing five models using multi-year datasets from five systems.

In addition, this work compares the accuracy of general and site-specific models, which is often missing. Fine-tuning is limited to models with modifiable coefficients and systems with available historical data. For a new system without historical measurement data, monitoring needs to be done using general models that are not optimised for the system. Huld et al. provided default coefficients for the Huld model, which can then be used as a general model [26], and Böök et al. showed that fine-tuning the Huld model increased the accuracy for two PV systems in Finland compared to the default coefficients [8]. If default values are not provided, general models can be trained using transfer learning (TL) [27]. TL is a method in which the model is constructed using data from a source domain and then applied to a new target domain with similar characteristics to the source domain [27]. TL methods have been actively researched in recent years, but the majority of the works cover advanced ML models which are often dependent on previous observations, such as long short-term memory (LSTM) models that are used for output forecasting [27,28]. This work applies the TL approach to MLP and GBM models, as well as empirical Huld and PVUSA models, to examine whether simpler empirical models benefit from the approach. Since the models are trained using data from four systems and tested using data from the system that is left out, the approach can be seen as Leave-One-Site-Out Cross-Validation (LOSO-CV). Anamiati et al. trained a general MLP model using data from other PV systems and successfully applied the trained model to a new system [29]. However, they focused on the accuracy of only the general model, omitting the comparison to the site-specific model. Comparing the accuracy of the general model to the fine-tuned model offers another validation method for the general model, since the accuracy of the fine-tuned model acts as a reference point. Further analysis of different general models and their comparisons to fine-tuned models is needed due to the rapid increase in new PV sites without historical data in the future, and will be carried out in this work.

Different empirical models are shown to give increasingly erroneous results when the irradiance drops below 800 W/m^2 or when the module temperature drops below zero [30]. However, when using ML models, Livera et al. showed that filtering out plane-of-array (POA) values under 100 W/m^2 decreased the accuracy of the models due to the smaller training dataset [31]. These conflicting results call for a comparison of model performances at different irradiance levels. This study aims to address this gap by splitting the irradiance into range bins and assessing the model accuracy across the bins.

This work carries out a comprehensive comparison of conventional empirical Huld, PVUSA, and PVWatts models with machine learning MLP and GBM models in high-latitude climates by utilising two novel approaches. First, general and fine-tuned models are compared to determine suitable models to use with and without fine-tuning. Such a comparison is often omitted, but it offers valuable insights into the benefits of fine-tuning. In addition, the comparison shows how close general models get to the accuracy of site-specific models when data are not available for fine-tuning. Second, the models are analysed under different irradiance levels to assess their accuracy under varying conditions. The specific aim is to determine the conditions in which general and fine-tuned models perform well, and whether the increased complexity of ML models offers significant improvements to modelling accuracy. Our work here builds on the master's thesis by Anttalainen [32], which focused mainly on one system. By covering several systems at different latitudes (60–67° N), regions with varying conditions (snowy North, coastal, and inland), tilt (15–90°) and azimuth angles (123–217° from North clockwise) and adding detailed analyses across different irradiance conditions, the current contribution aims to determine results and features of system modelling that generalise reliably to other high-latitude locations, for instance in Sweden, which has very similar climatic conditions to Finland.

2. Methodology

2.1. Experimental setups and data collection

The data used in this study are from five PV systems in four locations across Finland covering different latitudes (60–67° N), as shown in Fig. 1 and described in Table 1. All of the systems consist of monofacial polycrystalline silicon panels. The systems in Helsinki, Kuopio, and Sodankylä are maintained by the Finnish Meteorological Institute (FMI), and their data are openly available [33]. Karhu et al. provided detailed descriptions of these systems [7]. The Turku system is maintained by Turku University of Applied Sciences (TUAS), and its data are available [34]. The variables used and their resolutions are described in Supplementary Information (SI) Section S1. A summary of the processing steps is shown in Fig. 2.

The system in Helsinki (HEL) is installed on a flat rooftop. During low sun elevation angles, the system suffers from self-shading, which means that each panel row casts its shadow on the next panel row. During winter, snow often accumulates on the panels due to their low tilt and low clearance (small distance between the leading edge of the panels and the roof), resulting in the snow piling up in front of the panels.

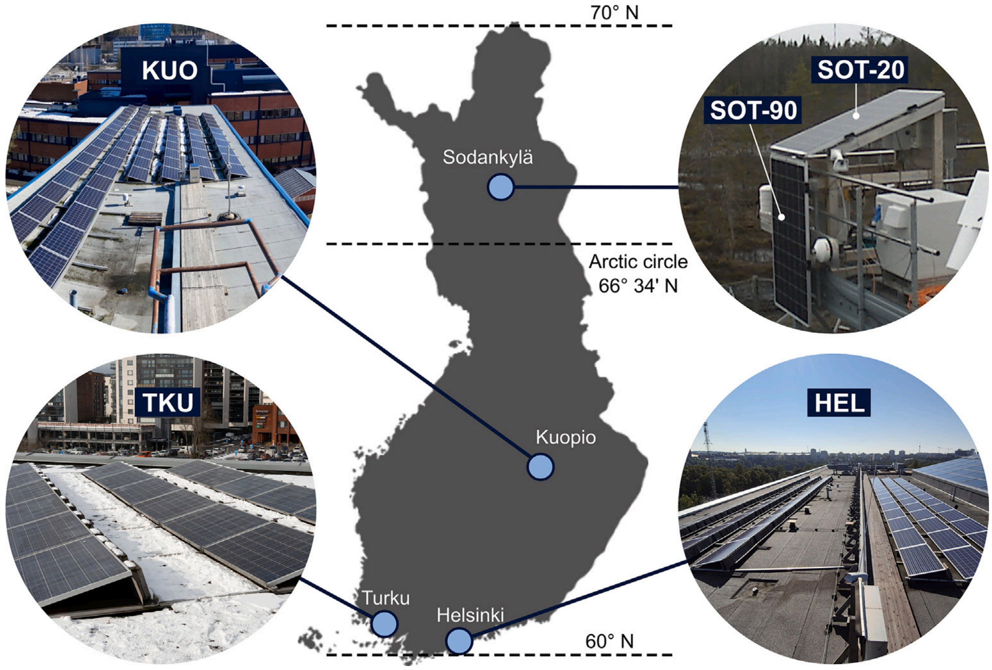

Fig. 1. Locations and photographs of the systems studied. Photographs of the Helsinki, Kuopio, and Sodankylä systems were taken by Karhu/FMI. The photograph of the Turku system was taken by Anttalainen/UTU. Reproduced from [35]. Copyright Karttunen et al. under CC-BY 4.0, 2026.

Table 1
Parameters of systems used. Panel azimuth is measured from north clockwise.

	HEL	KUO	SOT-20 and SOT-90	TKU
Location	Helsinki	Kuopio	Sodankylä	Turku
Latitude (° N)	60.20	62.89	67.37	60.45
Longitude (° E)	24.96	27.63	26.65	22.30
Altitude (m)	30	85	180	28
Altitude from the ground (m)	17	10	4	9
Maintainer	FMI	FMI	FMI	TUAS
Commission date	26 Aug 2015	22 Sep 2016	2 Feb 2017	Jul 2017
Monitoring period	Dec 2015–Dec 2021	Apr 2017–Nov 2020	Feb 2018–Nov 2021	Jan 2018–Sep 2020
Number of panels	84	78	1 and 1	18
System nominal power (Wp)	21,000	20,280	260 and 260	4500
Panel manufacturer	SolarWorld	SOLARWATT	SolarWorld	Kingdom Solar
Panel model	SW 250 poly	BLUE 60P	SW 260 poly	KD-P250W
γ (%/°C)	–0.41	–0.38	–0.41	–0.408
Panel azimuth (°)	135	217	123	180
Panel tilt (°)	15	15	20 and 90	15
Inverter	ABB TRIO-20.0-TL-OUTD	ABB TRIO-20.0-TL-OUTD	INV-350-60 micro-inverter	Fronius Symo Hybrid 4.0

The system in Kuopio (KUO) resembles the Helsinki system and likewise suffers from self-shading at low sun elevation angles and snow cover during winter.

There are two PV systems in Sodankylä, each consisting of a single 260 Wp nominal power panel. One panel is installed at a 20° tilt (SOT-20) and the other at a 90° tilt (SOT-90). Both panels are installed on a platform four meters above the ground with a high clearance preventing any piling of snow. The SOT-90 system benefits from high albedo caused by snow during winter due to its high tilt angle and location in northern Finland, which experiences significant snowfall.

The system in Turku (TKU) is installed on a flat rooftop. It is affected by two trees located 15 metres southwest of the panels and by an I-V room located 2.5 metres east of the panels. The system also suffers from snow cover during winter due to low clearance.

2.2. Irradiance and temperature modelling

In this work, empirical power models that require POA irradiance (G_{POA}), wind speed (WS), air temperature (T_{air}), panel temperature (T_{m}), and cell temperature (T_{cell}) are used. SI Table S1 shows that the TKU and SOT-20 systems are missing measurements for G_{POA}, which are modelled using beam (B_{POA}), sky-diffuse (D_{POA}), and ground-reflected (R_{POA}) components of G_{POA} as follows:

$$G_{\mathrm{POA}}^{\mathrm{modelled}} = B_{\mathrm{POA}} + D_{\mathrm{POA}} + R_{\mathrm{POA}}. \tag{1}$$

B_{POA} and R_{POA} can be calculated from GHI and DNI components as follows:

$$B_{\mathrm{POA}} = |\mathrm{DNI} \cdot \cos(\mathrm{AOI})| \tag{2}$$

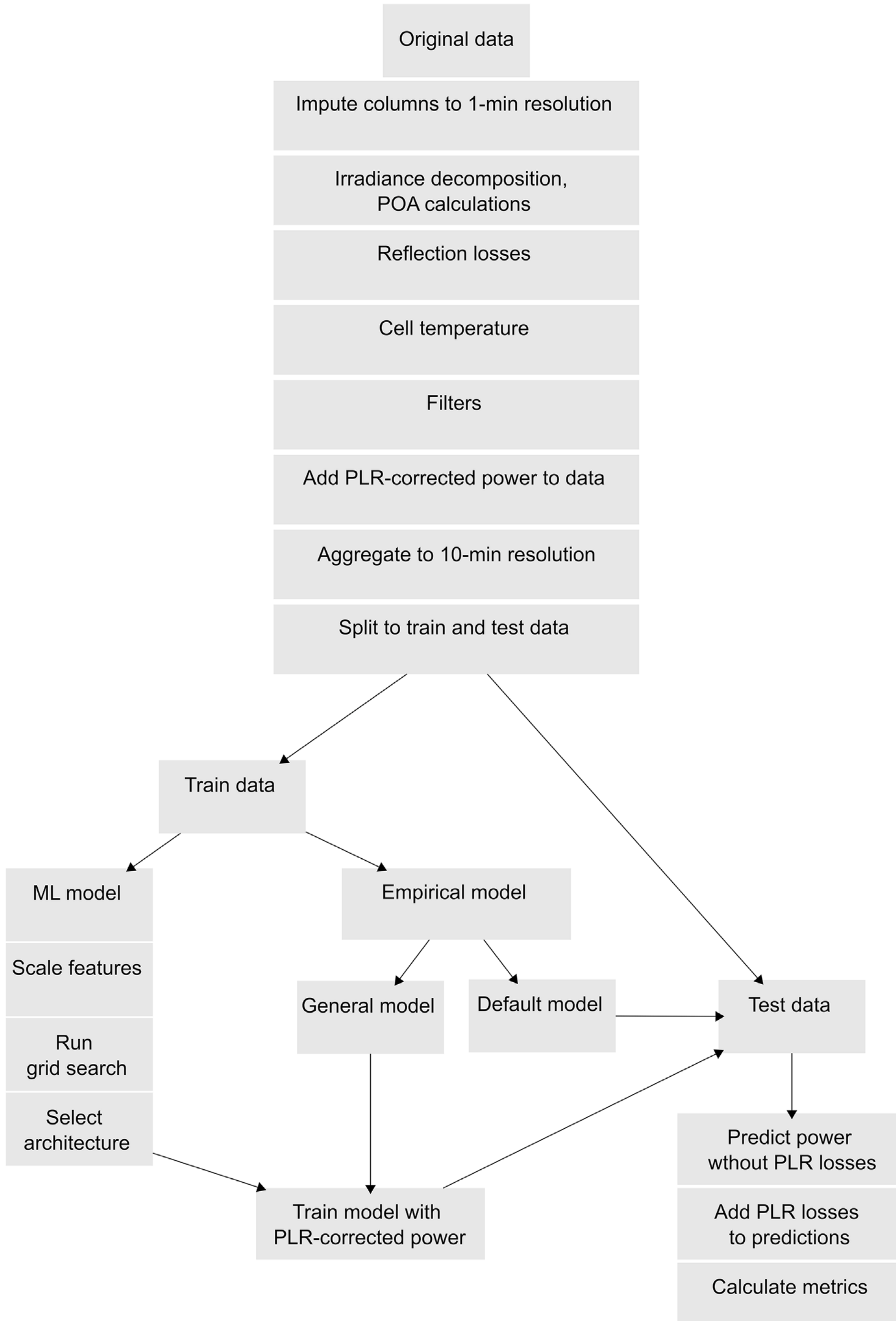

Fig. 2. Summary of the processing steps.

$$R_{\mathrm{POA}} = \mathrm{GHI} \cdot \alpha_{\mathrm{albedo}} \cdot \left(1 - \cos(\beta_{\mathrm{tilt}})\right) /2, \tag{3}$$

where AOI refers to the angle of incidence of the direct solar beam, α_{albedo} refers to the albedo, and β_{tilt} refers to the tilt of the panel.

Albedo value $\alpha_{\mathrm{albedo}} = 0.151$ was used for days without snow, since it is expected to be representative of these systems and is, for example, the default value in the FMI Open PV Forecast application [36]. Albedo value for snow varies depending on the freshness of the snow. Typical albedo values range from 0.6 (old dirty snow) up to 0.9 (fresh clean snow) [37]. In this work, albedo value $\alpha_{\mathrm{albedo}} = 0.7$ was used when snow was present. In HEL, KUO, SOT-20, and SOT-90, snow depth on the ground is reported once a day (at 00:00 UTC). Snow albedo is set if two consecutive measurements show snow (snow depth $>$ 0 cm). Snow depth data for TKU is downloaded from FMI's data service [38] using the weather station in Turku (Artukainen).

The D_{POA} component is calculated with the Perez sky-diffuse model [39]. Because the TKU system is only measuring GHI, the values for DHI and DNI were decomposed using the Erbs model [40]. The quality control (QC) method for DNI introduced by Böök et al. [8] is used for measured and decomposed DNI values. The QC method filters out unrealistically high DNI values as follows:

$$\mathrm{DNI}_{\mathrm{QClimit}} = A \cdot \exp(B \cdot \alpha) + C \tag{4}$$

$$A = -838 \tag{5}$$

$$B = -0.112 \tag{6}$$

$$C = 951 \tag{7}$$

$$\mathrm{DNI}_{\mathrm{QC}} = \min(\mathrm{DNI}, \mathrm{DNI}_{\mathrm{QClimit}}), \tag{8}$$

where α is the solar elevation. The authors also suggest setting DNI values to 0 if $\alpha \leq 0.5$, which was used in this work.

Because part of the irradiance is reflected from the panel surface, especially when the irradiance is coming at a high AOI, not all of the incoming irradiance is absorbed by the cells. The reflection losses are calculated using the models proposed by Martin and Ruiz [41]. The equations used for reflection loss calculations are shown in SI Section S2. The reflection-corrected POA (also known as absorbed or effective POA) G_{eff} is obtained as the sum of the reflection-corrected components (denoted by subscript rc):

$$G_{\mathrm{eff}} = B_{\mathrm{POA,rc}} + D_{\mathrm{POA,rc}} + R_{\mathrm{POA,rc}}. \tag{9}$$

The reflection-corrected values for measured POA (HEL, KUO, SOT-90 systems) were calculated as follows:

$$G_{\mathrm{eff}}^{\mathrm{meas}} = G_{\mathrm{POA}}^{\mathrm{meas}} \frac{G_{\mathrm{eff}}^{\mathrm{modelled}}}{G_{\mathrm{POA}}^{\mathrm{modelled}}}, \tag{10}$$

where $G_{\mathrm{POA}}^{\mathrm{meas}}$ is the measured POA value, $G_{\mathrm{eff}}^{\mathrm{modelled}}$ is the reflection-corrected POA calculated as above in Eq. (9), and $G_{\mathrm{POA}}^{\mathrm{modelled}}$ is the POA calculated as in Eq. (1).

Cell temperatures for HEL, KUO, SOT-20, and SOT-90 systems were calculated using the Sandia model [42], based on panel temperatures measured with thermocouples from the rear side of the panels:

$$T_{\mathrm{cell}} = T_{\mathrm{m}} + \frac{G_{\mathrm{eff}}}{G_{\mathrm{ref}}} \cdot \Delta T, \tag{11}$$

where ΔT corresponds to the temperature difference between the panel and cell at the reference irradiance of 1000 W/m^2. For the HEL system, Varjopuro et al. [6] calculated $\Delta T \approx 1$, which was also used for KUO, SOT-20, and SOT-90 systems. Cell temperatures for the TKU system were calculated with the Sandia model [42], which was modified as follows:

$$T_{\mathrm{cell}} = G_{\mathrm{eff}} \cdot \exp(a_{\mathrm{cell}} + b_{\mathrm{cell}} \cdot WS) + T_{\mathrm{air}}, \tag{12}$$

where parameter values $a_{\mathrm{cell}} = -3.43$ and $b_{\mathrm{cell}} = -0.125$ were used, as suggested by Varjopuro et al. for silicon panels in a high-latitude location [6]. Panel temperatures for the TKU system were then calculated from Eq. (11):

$$T_{\mathrm{m}} = T_{\mathrm{cell}} - \frac{G_{\mathrm{eff}}}{G_{\mathrm{ref}}} \cdot \Delta T, \tag{13}$$

where the value $\Delta T = 1$ was also used.

2.3. Data filters

The following filters are applied to remove nighttime, erroneous, and outlier datapoints, as well as datapoints when panels are under snow cover:

1. Daytime filter ($G_{\mathrm{POA}} \geq 5$ W/m^2)
2. Threshold filters suggested by the International Electrotechnical Commission (IEC) [43]
3. Clipping filter (constant threshold)
4. Snow cover and quality filter by Karhu et al. [7] applied to FMI datasets (later referred to as QC filter)
5. Outlier filter (± 2 standard deviations (SD) from the mean of P/G_{eff} of a rolling horizon over 800 datapoints sorted by irradiance)
6. Cut-outlier filter for SOT-20 system (removing datapoints based on line-fitting)

The IEC threshold filters remove unrealistic values, and are shown in Table 2. For HEL and KUO systems, a clipping threshold value of 20 kW for DC power was used, while other systems were not impacted by inverter limitations. In addition, the datapoints from TKU data where $P \leq 100$ W and $G_{\mathrm{eff}} \geq 100$ W/m^2 were removed because of the unrealistic combination of low power and high irradiance.

A QC filter was applied to FMI's datasets HEL, KUO, SOT-20, and SOT-90. The filter excludes datapoints with snow cover and quality issues based on manual classification from photos and data. The QC filter was applied based on the following variables and their values: 'dataQC = 1' and 'vis_SnoP = 0'. These variables are readily available in the datasets for HEL, KUO, SOT-20, and SOT-90 systems, and are described in more detail by Karhu et al. [7]. Since these variables are missing in the TKU dataset, the QC filtering is omitted for TKU.

Outliers were filtered out by including only datapoints that are within 2SD from the mean of 800 rolling horizon P/G_{eff} points sorted by irradiance. The rolling horizon of 800 removed the fewest data points while still maintaining the same performance compared to other tested rolling horizon lengths, as shown in SI Figure S1. In addition, this outlier filter was extended for the SOT-20 system by implementing a line-fitting to low power/irradiance points and removing points below that line. A detailed description of the outlier filter and cut-outlier filter is provided in SI Section S1.1. After filtering, the datasets were aggregated from the original resolution (one minute with FMI data, five minutes with TUAS data) to 10 min resolutions to unify the datasets (see SI Section S1.2).

2.4. Performance loss rate

Aging and soiling decrease the power output of the PV system over time. Power decay of the system is usually expressed by the performance loss rate (PLR), which is provided in the unit of %/year [44]. If performance losses are omitted when predicting power, predictions will be greater than the actual output. In addition, performance losses affect model training when historical data are used. The performance losses are taken into account by calculating what the power would be without performance losses for every datapoint using the following equation:

$$P_{\mathrm{without\_plr}} = \frac{P_{\mathrm{with\_plr}}}{(1 - PLR \cdot \Delta t)}, \tag{14}$$

where Δt is years from the initial date. PLR value of -0.33%/year is used for all systems, since it is reported to be the median degradation rate for c-Si technologies in cold and snowy climate [45]. The first datapoint of filtered data was used as the initial date. The models were trained using $P_{\mathrm{without\_plr}}$ as the output and predictions of the models were scaled back to powers with PLR losses using the following equation:

$$P_{\mathrm{with\_plr}} = (1 - PLR \cdot \Delta t) \cdot P_{\mathrm{without\_plr}}, \tag{15}$$

which makes the output comparable with the actual system output power.

Table 2
IEC filters, where G_{POA} refers to POA irradiance, T_{air} refers to air temperature, WS refers to wind speed, P_{AC} refers to AC power after inverter, and P_0 refers to rated DC power.

Variable	Lower limit	Upper limit
G_{POA} (W/m^2)	−6	1500
T_{air} (°C)	−20	50
WS (m/s)	0	32
P_{AC} (W)	$-0.01 \cdot P_0$	$1.02 \cdot P_0$

2.5. Power models

Empirical models

The Huld model is a regression-based empirical model introduced by Huld et al. [26]. It models the power using POA irradiance and module temperature as follows:

$$P(G_{\mathrm{POA}},T_{\mathrm{m}},P_0) = G'_{\mathrm{POA}}(P_0 + k_1 \ln(G'_{\mathrm{POA}}) + k_2 \ln(G'_{\mathrm{POA}})^2 + k_3 T' + k_4 T' \ln(G'_{\mathrm{POA}}) + k_5 T' \ln(G'_{\mathrm{POA}})^2 + k_6 T'^2), \tag{16}$$

where

$$G'_{\mathrm{POA}} = G_{\mathrm{POA}}/G_{\mathrm{ref}}$$
$$T' = T_{\mathrm{m}} - T_{\mathrm{ref}},$$

and $k_1,\dots,k_6$ are fitting coefficients. The reflection-corrected POA irradiance G_{eff} is used as the input in Eq. (16), and in the following models, instead of G_{POA}. The Huld model has been shown to work well for outdoor power modelling with $G_{\mathrm{POA}} > 300\ \mathrm{W/m^2}$ [26], and it is used in power modelling and forecasting studies in high-latitude conditions [8,9,11,20]. The coefficients and the output can be scaled with the nominal power of the system, which will result in notations $k'_1 = k_1/P_0,\dots,k'_6 = k_6/P_0$ and P/P_0. Huld et al. provide default values for these coefficients $k'_1,\dots,k'_6$ [26], which can be used to construct a general model. The coefficients can also be fine-tuned to a specific system by using historical data.

The Photovoltaics for Utility Scale Applications (PVUSA) project developed a regression-based empirical model for power modelling [46]. It models power by using POA irradiance, air temperature, and wind speed as follows:

$$P(G_{\mathrm{POA}},T_{\mathrm{air}},WS) = G_{\mathrm{POA}} \cdot \left(a + b \cdot G_{\mathrm{POA}} + c \cdot T_{\mathrm{air}} + d \cdot WS\right), \tag{17}$$

where a, b, c, and d are fitting coefficients. Similar to the Huld model, these coefficients must be fitted in the PVUSA model before it can be used. However, the PVUSA model lacks default values for the coefficients. To use it as a general model, the same training method is used as the method used with ML models as described in Section 2.6. The coefficients are scaled similarly to the scaling used with the Huld model, which results in notations $a' = a/P_0,\dots,d' = d/P_0$. The PVUSA model is shown to have poor performance with POA irradiances below 500 $\mathrm{W/m^2}$ [46]. It also returned unrealistic results in high-latitude conditions [11].

PVWatts is a web application for estimating the power production and energy yields of a PV system developed by the National Renewable Energy Laboratory (NREL) [47]. It uses a simple empirical model that calculates power using POA irradiance and cell temperature as follows:

$$P(G_{\mathrm{POA}},T_{\mathrm{cell}},\gamma,P_0) = \frac{C_T \cdot P_0 \cdot G_{\mathrm{POA}}}{1000\ \mathrm{W/m^2}}, \tag{18}$$

where C_T is a temperature correction term $C_T = 1 + \gamma(T_{\mathrm{cell}} - 25\ ^\circ\mathrm{C})$ and γ is the temperature coefficient of the panel. Unlike previous models, PVWatts has no fitting coefficients, which means that the model can be used directly as a general model to calculate power without fitting. It is described as a tool for quick and easy estimation of PV production, but annual energy yield errors can be up to 20% and monthly energy yield errors up to 40% [47].

Machine learning models

In addition to the empirical models presented, the MLP and GBM machine learning models are used. MLP is an artificial neural network consisting of multiple layers, whereas GBM is an ensemble of weak predictors (usually decision trees) [48]. Both models have been used successfully in PV power forecasting [19,49]. In order to keep the model usage similar to the empirical models, the same input variables are used with the ML models. These include reflection-corrected POA irradiance, air temperature, module temperature, and wind speed. The models were implemented using the Python library scikit-learn version 1.7.2 [50]. Because of the large size of the datasets used, the histogram-based implementation of GBM provided by the class `HistGradientBoostingRegressor` was used. Detailed descriptions of the ML model implementations are in SI Section S4. The ML models were trained using the supercomputer Puhti provided by CSC. All codes used for this work are available in the repository [51].

2.6. Model training

PVWatts can be used to directly estimate the power of a PV system without historical data due to the lack of fitting coefficients. Additionally, the Huld model can be used directly since default values are provided for its coefficients. The Huld model with default coefficients and the PVWatts model will be referred to later as default models due to the lack of fitting requirements. To construct the general Huld, PVUSA, MLP, and GBM models, TL is utilised by using data from four systems to build the general model and testing the model on the data from the system that was left out. LOSO-CV is used to validate the TL approach by holding each system as a test set once and training the models with the data from the rest of the systems. The illustration of training the general model is shown in the upper part of Fig. 3.

The Huld, PVUSA, MLP, and GBM models were also fine-tuned for each system using the data from the first 365 days, which is a common length for the training set [11]. By using one year of data for training, the models have the opportunity to learn seasonal characteristics that are highly prevalent in high-latitude conditions. Using shorter training sets than a full year would likely lead to worse performance of the models, since during the testing phase the models would encounter unseen seasons. Thus, it can be expected that a comparison of the general models to fine-tuned models with a full year of training would show the most difference, and shorter training sets would lead to similar performance between these models. For the Huld and PVUSA models, the coefficients of the general models were used as starting values, and for the MLP model, the weights of the general MLP model were used as starting values. For the GBM model, a completely new model was created for fine-tuning using the same hyperparameters as with the general GBM model, because fine-tuning the existing general model showed decreased performance when new weak learners were added. The lower part of Fig. 3 illustrates the fine-tuning process.

Hyperparameters of the general ML models were tuned using grid searches. A detailed description of the tuning process is available in SI Section S4. The following MLP model parameters were selected for each system: hidden layer size (200, 200, 200), an initial learning rate of 0.0001, a batch size of 32, the ReLu activation function, and Adam as the algorithm. The following GBM model parameters were selected for each system: learning rate of 0.1, maximum number of leaf nodes for each tree of 32, and maximum number of trees of 400. The possibility of overfitting is mitigated by using early stopping if the validation score does not improve for 10 epochs for both ML models.

The outlier and cut-outlier filters were estimated using the full time series. Since the general models are constructed using a transfer learning approach utilising data from different systems than the target system, there is no information leakage, as the data of each system are processed independently. The avoidance of information leakage also applies to the default Huld and PVWatts models, as no training is done with them.

With fine-tuned models, a year of filtered data is used to train the model and the rest of the data is used for testing. When performing the outlier filtering, data is sorted by irradiance before the rolling window, so the training set and testing set are mixed. Thus, there is a possibility that the testing set affects the training and vice versa. However, since there is no long time-dependent rolling window and since the outlier

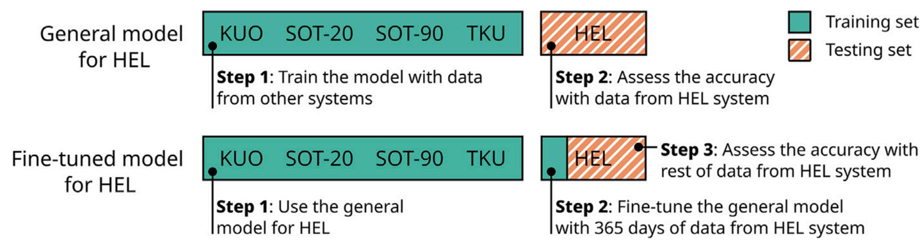

Fig. 3. Illustration of the training process for models for the HEL system. General models for each system were trained by leaving the data from that system aside, and training the general model using data from other systems. Fine-tuned models are trained from general models by performing additional training with data from the first 365 days of the given system.

and cut-outlier filters remove a small portion of the data (up to 5%), we expect the possible data leakages to be minor, thus retaining the validity of the results.

3. Results and discussion

3.1. Impact of filtering

To reliably compare the accuracy of the models, the data used should be of high quality. The quality of the data is enhanced with various filters. The performance of the filters was assessed by plotting the actual output power against reflection-corrected POA irradiance, as shown in Fig. 4. Since the power of the system is highly correlated with the irradiance, erroneous datapoints, caused by, for example, shading and snow cover, can usually be seen from the scatter plot as outliers. Outlier values can occur, for example, if panels are shaded while the irradiance measurement instrument is not (or vice versa), creating a power–irradiance mismatch.

Each system in Fig. 4 shows varying levels of variance. Data in which IEC filters, clipping filters (for HEL and KUO), and QC filters (for HEL, KUO, and SOT systems) have been applied are referred to as threshold filtered data and are shown in Fig. 4 as green dots. They removed 2% to 34% of the daytime datapoints (yellow dots) in HEL, KUO, SOT-20, and SOT-90 systems. The outlier filter (blue dots) removed 5%-point, at most, of additional datapoints but made the resulting dataset visually more cohesive.

In the SOT-20 system, the outlier filtered datapoints form a “hook” on irradiance values over 800 W/m^2 (as shown in SI Figure S4F). This pattern might be a result of a rolling-window outlier filter and low data density with high irradiance values. Because these datapoints deviate from the expected highly linear relationship between irradiance and power, they were removed. Since the cut-outlier filter removes 0.2% of the data (Fig. S4G), the effect on the model performance is likely small. Removal is done by fitting a line to low power/irradiance points and removing points below that line. This method is referred to as a cut-outlier filter and is explained in more detail in SI Section S1.1.

There are no similar QC filters for TKU data that filter out snow cover and erroneous data similar to Karhu QC for FMI’s systems, due to different weather sensing availability in FMI and TUAS. In the TKU system, the threshold filters filtered 4% of the daytime data. After outlier filtering, the TKU system also showed high scattering, similar to the SOT-20 system. Scattering in these systems might be caused by the lack of POA measurements, which results in the use of modelled POA values that may introduce inaccuracies to the G_{eff} values. In addition, the irradiance measurement devices for GHI, DHI, and DNI are located outside the SOT systems, which can introduce scattering to SOT-20 data (the SOT-90 system has direct on-site G_{POA} measurements), especially during rapidly changing weather. Furthermore, in practice, the SOT-90 system prevents the accumulation of snow (as explained in Section 2.1), which likely contributes to a reduction of scattered data.

The filters remove the majority of the scattering in each system, while retaining a high proportion of daytime data (over 63%). In combination with the preprocessing steps (reflection corrections, DNI quality control, handling of performance loss rates, and aggregating to 10 min resolution), filtering is used to minimise the noise and erroneous datapoints, and to unify the data across systems. If such processing is omitted, the noise will affect the training of the models, resulting in decreased learning of underlying patterns related to weather conditions and power output. The trained models were also evaluated using datasets with only daytime filtering in order to analyse the impact of filtering on the model accuracy.

3.2. Model performance across the whole dataset

The overall accuracies of the models were assessed by plotting modelled power output against actual filtered power output and calculating the corresponding coefficient of determination (R^2) value, which is a metric often used in PV power forecasting and represents how well the model fits the data [52]. It is described further in SI Section S3. Fig. 5 shows scatter plots for each model (def = default; gen = general; fit = fine-tuned) and system with all filters. The default PVWatts model tends to overestimate the power values in each system, as the majority of the points are above the diagonal lines. On the other hand, the default and general Huld models overestimate the power values in SOT-20 and SOT-90 systems, whereas general PVUSA, MLP, and GBM models seem to overestimate power values only in the SOT-20 system; in other systems, the points follow the diagonal line well.

R^2 values are high (from 0.980 up to 0.997) in the HEL and KUO systems with all models, indicating good accuracy of the models. However, the lowest R^2 values are in the SOT-20 system, likely due to the highly scattered data. Default Huld and PVWatts models have R^2 values of 0.883 and 0.869, while general MLP, GBM, Huld, and PVUSA have higher R^2 values of 0.974, 0.972, 0.944, and 0.966. The differences in R^2 values between default and general models in SOT-20, SOT-90, and TKU systems indicate that utilising TL with data from other systems to construct the general models could increase model accuracy.

Fine-tuning the models generally increases the R^2 values and reduces over-/underestimation with empirical and ML models, which was an expected result. Even with the highest scattered data from the SOT-20 system, fine-tuned models achieve high R^2 values from 0.966 to 0.991, indicating that the models captured the power generation patterns well. However, the general MLP, GBM, Huld, and PVUSA models have the same or almost the same R^2 value (a difference smaller than 0.01) as their fine-tuned version in HEL, KUO, SOT-90, and TKU, meaning that the general models captured the variability of the data almost as well as the fine-tuned models. Thus, the overall accuracy of the models across all datasets is good based on the R^2 values, especially when the models are fine-tuned.

Fig. 6 shows the scatter plots and R^2 values when only a daytime filter was applied. Compared to the results when all filters are applied in Fig. 5, the data exhibit more scattering and lower R^2 values as expected due to noisier data. Overall trends are still similar to those with all filters:

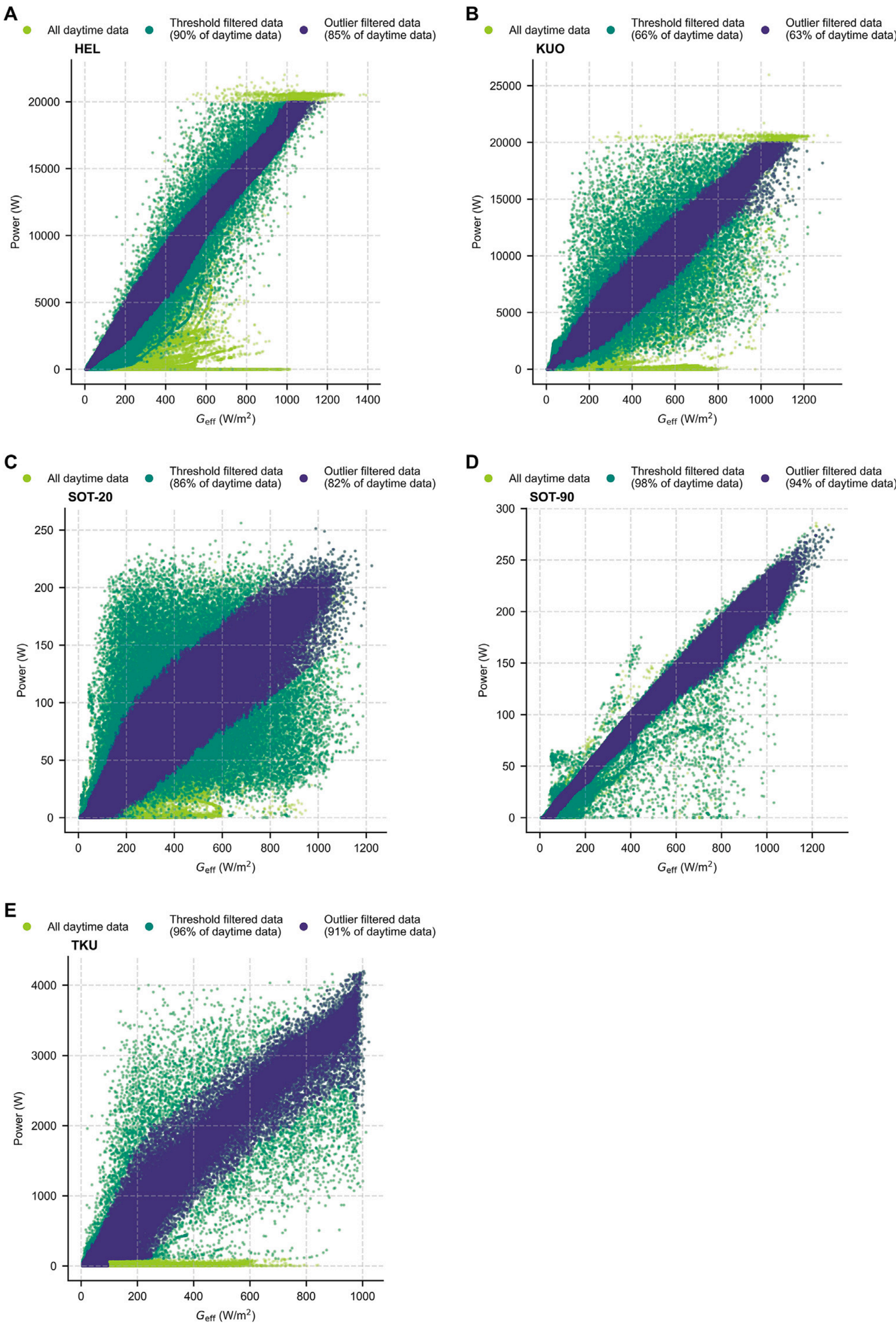

Fig. 4. Comparison of power outputs at given reflection-corrected POA irradiance values with different filters for HEL (A), KUO (B), SOT-20 (C), SOT-90 (D), and TKU (E) systems.

default models have the lowest R^2 values and fine-tuned models have the highest. However, in the KUO system, general Huld and ML models have higher R^2 values than the fine-tuned models. Additionally, the R^2 values in the SOT-90 system show only slight differences between the filtering, which is likely the result of lower data noise due to periods of low snow cover. The following sections show results when all filters are applied to the datasets and the results for only daytime filtered data are shown in SI Sections 5, 7, and 8.

3.3. Model performance at different irradiance levels

To get a better understanding of the performance of the models at different irradiance levels, the data are split into irradiance bins using the same bin ranges as Karttunen et al. [11]. Fig. 7 shows the ratio of reflection-corrected POA irradiance for each irradiance bin in each system when all filters are applied. The majority of the irradiance is under 200 W/m^2 for every system. The ratios are similar to the ratios reported by Karttunen et al. [11], except for the SOT-90 system in Fig. 7,

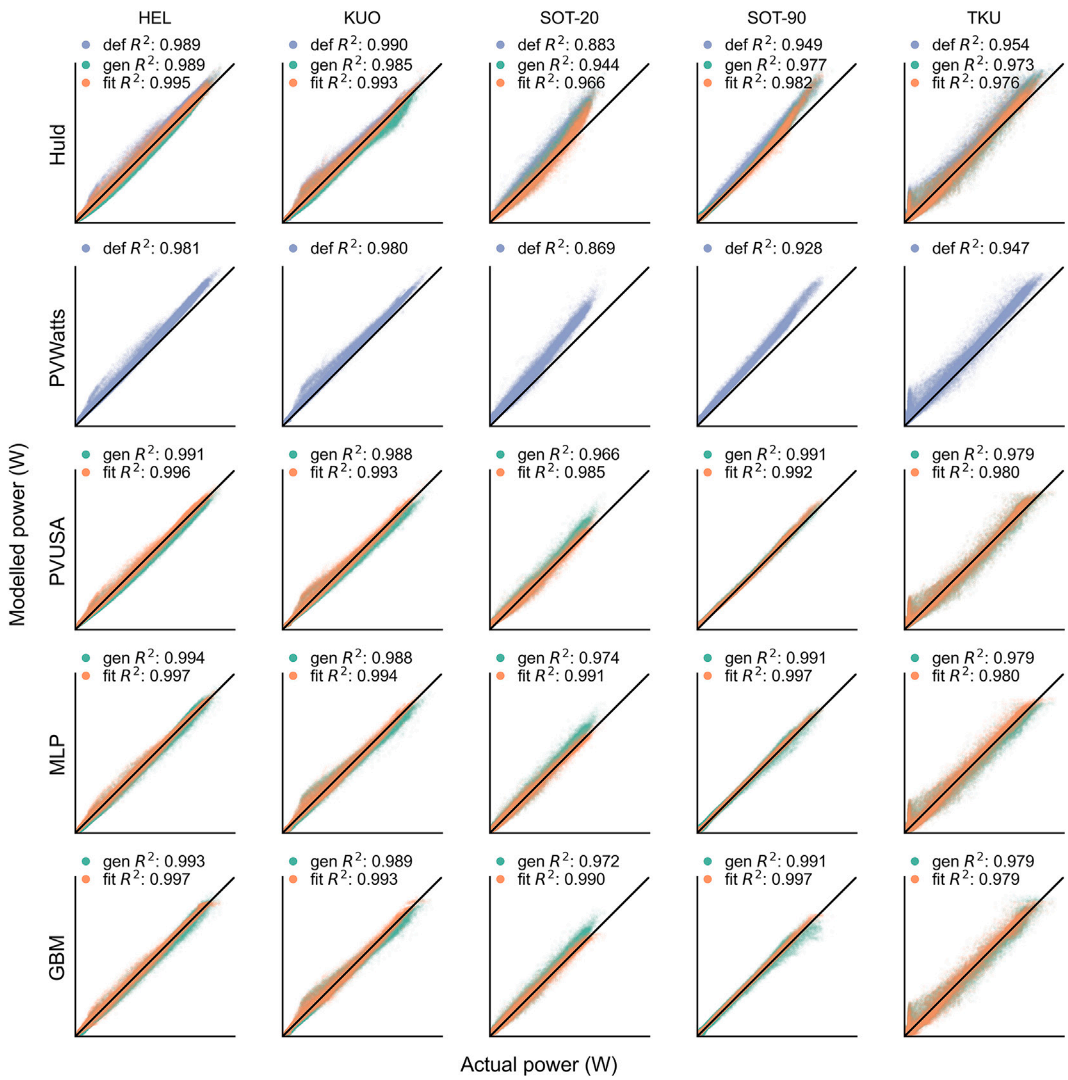

Fig. 5. Actual vs. modelled power in watts for every system and model, as well as the corresponding R^2 value when all filters are applied. PVWatts lacks the ability to fine-tune, hence the Orange and green dots are missing. The axis limits for each system are chosen so that all datapoints are visible. Each system plot shares the same axis limits. The diagonal line represents the perfect match between modelled and actual power. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

which shows an increased share of data in the range of 20–100 W/m^2. These results show that the majority of the data lie in low irradiance conditions, thus highlighting the importance of model validation under these low irradiances. This result also agrees with the recommendation of Øgaard et al., suggesting that common irradiance threshold filters, such as 200 W/m^2, should be avoided in high-latitude conditions due to excessive removal of data [14].

Fig. 8 shows normalised root mean squared error (nRMSE) values for each model and irradiance bin for each system. The nRMSE metric is described further in SI Section S3. Generally, nRMSE values decrease as irradiance increases or the model is fine-tuned, indicating improved accuracy. In HEL and KUO systems, most models produce nRMSE values below 0.2 down to an irradiance of 50 W/m^2. In SOT-20 and SOT-90 systems, most models produce increased nRMSE values up to an irradiance of 100 W/m^2. In the TKU system, most models showed similar nRMSE values regardless of whether the model was default, general or fine-tuned, but in other systems, fine-tuning generally lowered the nRMSE errors. In HEL and KUO systems, the default Huld model produced lower nRMSE values than the general Huld model in the majority of the irradiance bins, especially in the lowest irradiance bin of 5–20 W/m^2, where it produced the lowest nRMSE values out of all the models, including fine-tuned models. However, in the rest of the systems, the general and fine-tuned Huld models have higher accuracy in the majority of the irradiance bins, indicating that HEL and KUO systems resemble the systems for which the default Huld coefficients were obtained.

Both SOT systems showed a rapid increase of errors with the lowest irradiance bin, reaching nRMSE values over 20. These high errors differ from other systems, which have error values consistently under 2. The poor accuracy in low irradiances of the SOT systems could arise from poor low-irradiance performance of the micro-inverters, or low data counts in these irradiance bins in the SOT systems, for example. Generally, high errors can originate from noisy data caused by unsteady efficiencies of panels and inverters in low irradiance conditions [14]. The efficiency of the solar panels decreases and becomes non-linear under low irradiance conditions due to the increase in shunt resistance [53], and the uncertainty increases with larger shares of low light conditions [54]. In addition, increased relative errors, even with low absolute errors, can increase nRMSE values in low irradiance conditions since the errors are divided by the mean of the power, which is small in low irradiance conditions. In addition to high error values in the low

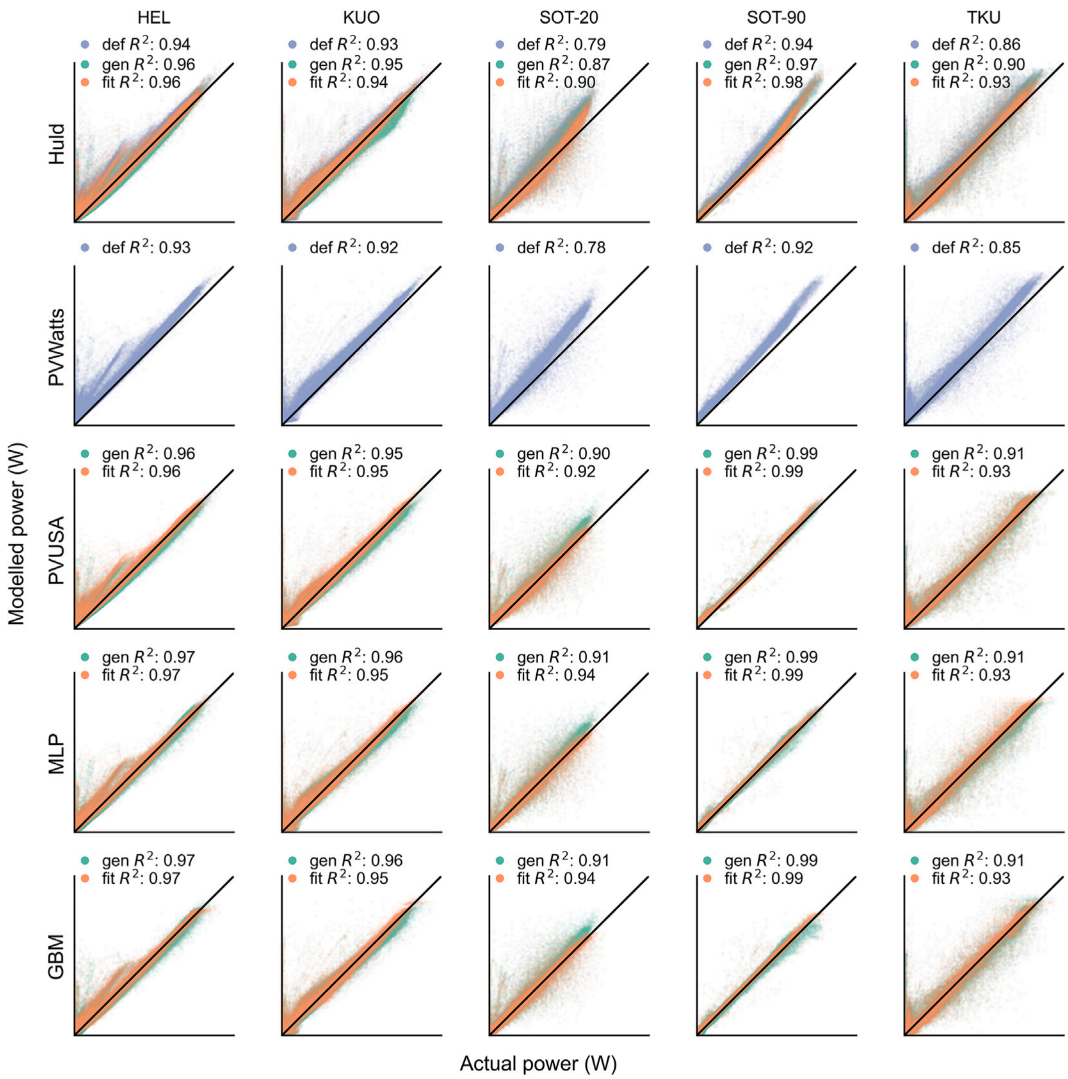

Fig. 6. Actual vs. modelled power in watts for every system and model, as well as the corresponding R^2 value when only the daytime filter is applied. PVWatts lacks the ability to fine-tune, hence the Orange dots are missing. The axis limits for each system are chosen so that all datapoints are visible. Each system plot shares the same axis limits. The diagonal line represents the perfect match between modelled and actual power. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

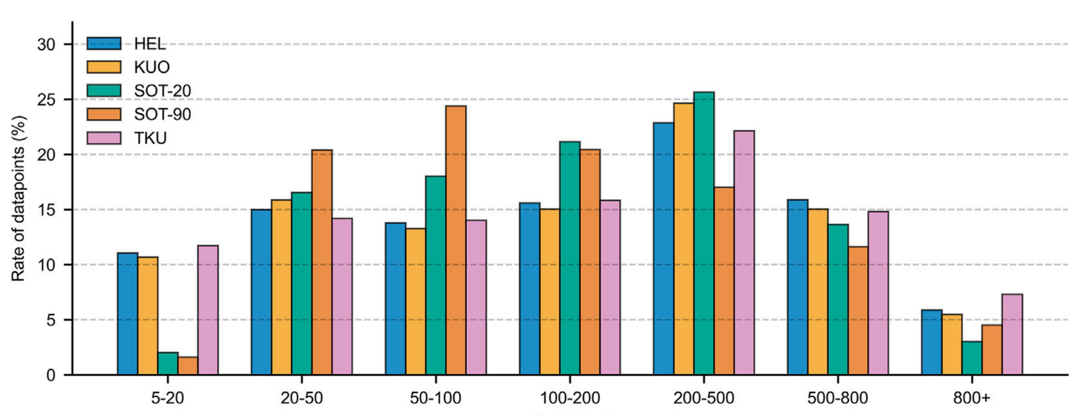

Fig. 7. Share of reflection-corrected POA irradiance for different irradiance bins of each system when all filters are applied.

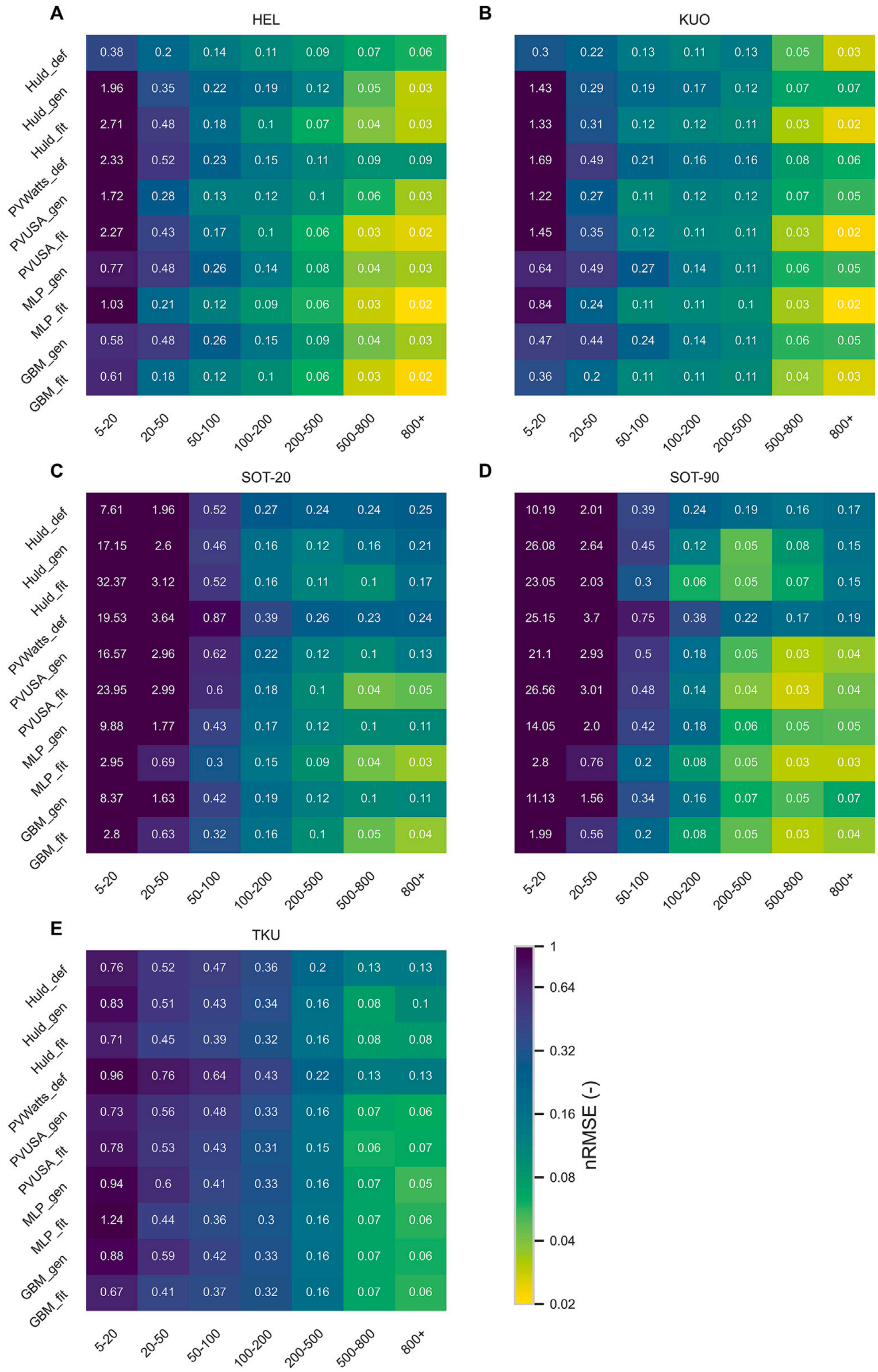

Fig. 8. nRMSE values for each model and irradiance bin for HEL (A), KUO (B), SOT-20 (C), SOT-90 (D), and TKU (E) systems when all filters are applied. The colouring is logarithmic and same with values over 1 to emphasize the comparison of irradiance levels with smaller nRMSE values.

irradiance conditions in SOT systems, the error values tended to increase in the last irradiance bin (irradiance over 800 W/m^2), especially with the fine-tuned Huld model. This increase of errors could be caused by the low data count (as seen in Fig. 7). nRMSE values for the data where only daytime filtering is applied are shown in SI Figure S8. The accuracy differences are smaller than with datasets where all filters are applied. However, the high low-irradiance accuracies of the fine-tuned MLP and GBM models are visible, as well as the high low-irradiance accuracy of the default Huld model on the HEL and KUO systems.

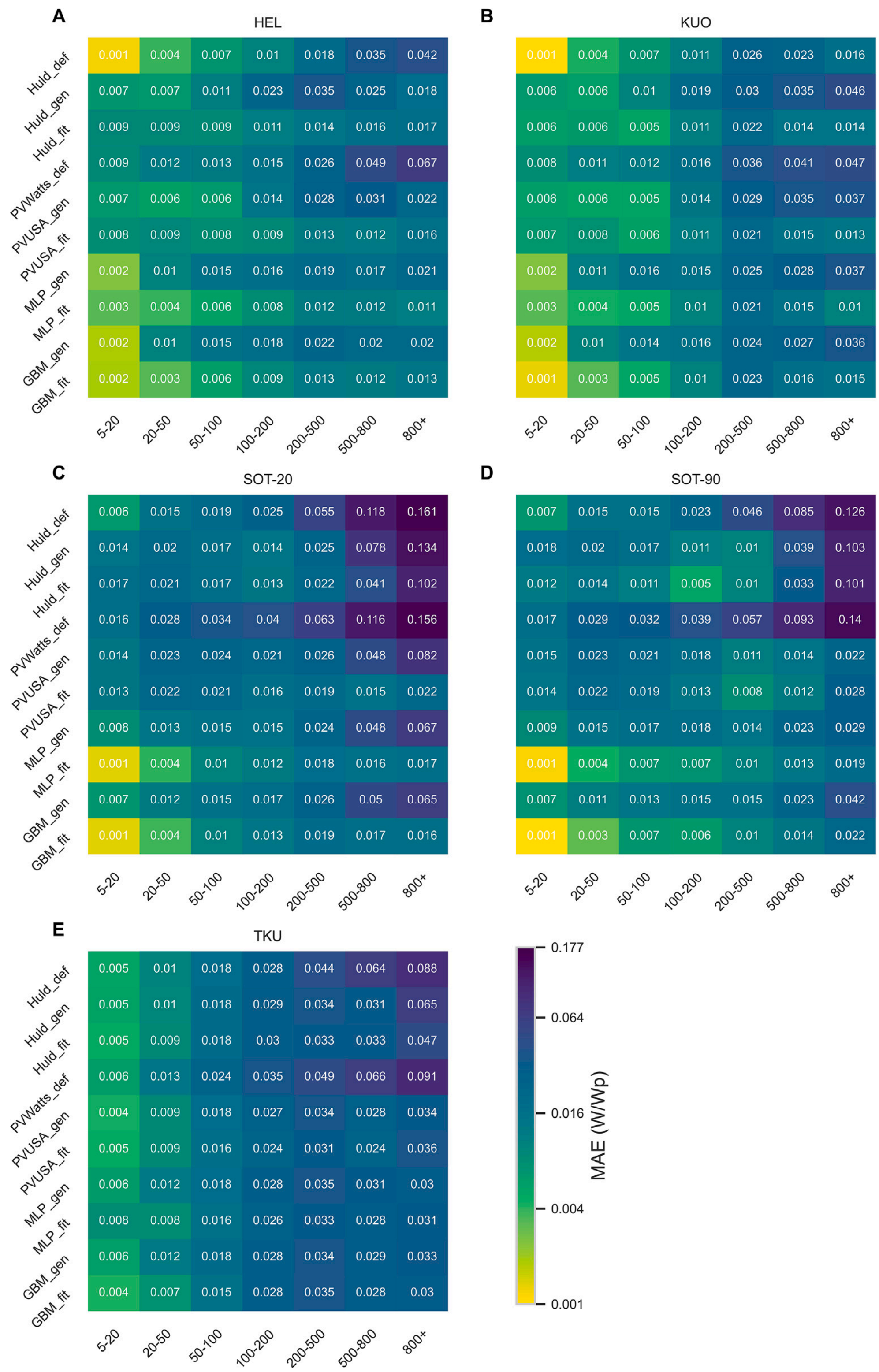

Fig. 9. MAE values scaled with the rated power of the system for each model and irradiance bin for HEL (A), KUO (B), SOT-20 (C), SOT-90 (D), and TKU (E) systems when all filters are applied. The colouring is logarithmic.

Because low power values can inflate the nRMSE values at low irradiance levels due to the division by mean power, mean absolute error (MAE) values are also calculated for the irradiance bins. MAE values that are scaled with the rated power of the system are shown in Fig. 9. MAE values for the data where only daytime filtering is applied are shown in SI Figure S9. Compared to Fig. 8, MAE values are low during low irradiance conditions due to low absolute errors in that region. Despite the metric used, fine-tuning generally lowers the errors, and fine-tuned ML models have lower low-light errors compared to empirical models. Additionally, the increased errors under high-irradiance conditions in

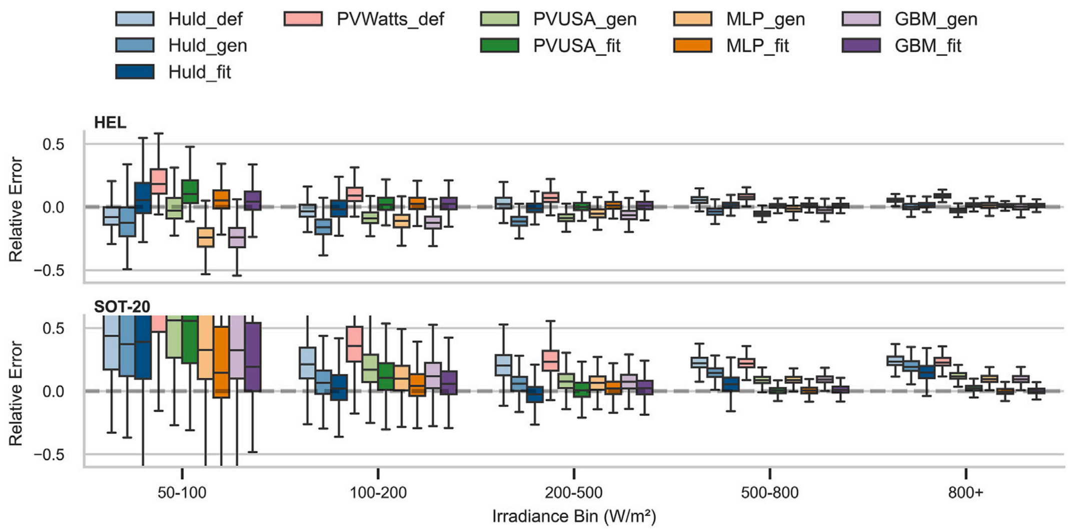

Fig. 10. Boxplots of relative errors ((modelled – actual) / actual) for HEL and SOT-20 systems when all filters are applied. The irradiance bins under 50 W/m^2 are omitted to focus on conditions in which the models demonstrate sufficient performance.

the SOT-20 and SOT-90 systems are visible as darker MAE values in Fig. 9(C) and (D). Better performance under specific conditions, such as in low-light, of the ML models is likely due to the more complex and less limiting structure of the ML models compared to the empirical models, which are more limited by the formulation of the used equation and the number of coefficients.

Fig. 10 shows the relative errors of the models for HEL and SOT-20 systems with irradiance bins over 50 W/m^2. The lowest irradiance levels are omitted due to high errors (as shown in Fig. 8). These two systems had the highest and lowest overall R^2 values, as shown in Fig. 5. Additionally, they represent geographically varying conditions, as the HEL system is located on the coast of Southern Finland with a latitude of 60.20° N, whereas the SOT-20 system is located in snowy Northern Finland with a latitude of 67.37° N. The relative errors of other systems and low irradiance values are shown in SI Figure S10. SI Figure S11 shows the relative errors when only the daytime filter is applied. The boxplots represent error distributions of all datapoints within the given irradiance bin, the box indicates the first and third quartiles, the line within the box displays the median, and the whiskers indicate the furthest points within 1.5 interquartile range (IQR) from the first and third quartiles. Fig. 10 shows that the variance of errors decreases with irradiance, as shown by the decreasing sizes of the boxplots at higher irradiances. However, even with the low variance of errors (short boxplots), the default PVWatts model has a tendency to overestimate the power across irradiance levels and systems (as shown in SI Figure S10), because the medians of the boxes are greater than 0. The default Huld model lacks this tendency, since the medians are close to 0 in the HEL system (Fig. 10A), but systematically over 0 in the SOT-20 system (Fig. 10B). The error trends of the general models are similar to each other.

Figs. 8 and 10 show that the accuracy of the models changes across different irradiance levels. These patterns are harder to spot from Fig. 5, highlighting the need for a more detailed irradiance-based comparison rather than overall R^2 values. The accuracy differences between the systems imply that the system characteristics affect the power generation patterns. These characteristics include different installations, such as varying system sizes, azimuth and tilt angles, but also geographical locations, local climate, and shading patterns, for example. However, the number of available systems with long-term data limits how deep an analysis can be made. Still, errors are generally smaller with general models than with default Huld and PVWatts models, showing that even with domain shifts across systems, the used TL method is beneficial even with empirical models.

3.4. Example power profiles

To visualise the accuracy of the models during a single day, power profiles for a sunny day (18/05/2019) in Helsinki are plotted in Fig. 11. The upper figure shows the profiles and the lower figure shows the relative errors of the models. Power profiles for other systems showed patterns similar to the HEL system, and are shown in SI Section S6. All models followed the shape of the profile quite well, but mornings and evenings showed distinct relative error patterns due to small production values. Empirical models showed increased relative errors for each model under low irradiance conditions, for example, before 6:00 and after 18:00 in the UTC+2 time (Fig. 11A), except for the default Huld model, which had relative errors close to 0 around 20:00. A positive increase in relative errors indicates overestimation of power under low irradiance conditions. Fine-tuning decreases the relative errors for empirical models for most of the day, but after 18:00, the fine-tuned Huld and PVUSA models showed larger relative errors than their general versions.

In contrast, the relative errors of ML models (Fig. 11B) are similar to each other and stay closer to 0 for longer during the day. However, general ML models showed signs of systematic underestimation of power on the example day from 15:00 to 18:00 (Fig. 11B), which is effectively corrected by fine-tuning. The relative errors also oscillated between positive and negative, indicating a lack of systematic error trends under low irradiance conditions and consequently preventing any easy way of correcting them.

3.5. Model performance at monthly and yearly levels

To investigate how errors accumulate to less granular levels and how well the models handle the non-stationarity caused by seasonality and system aging, the energy yield errors are compared at monthly and annual levels. Correct yield estimation is crucial when estimating the viability of a new system, for example. Fig. 12 shows errors in the HEL system and Fig. 13 shows errors in the SOT-20 system for empirical models (A) and ML models (B). Other systems are shown in SI Figures S16–S18, and results with only the daytime filter applied are shown in SI Figures S19–S23. Both systems in Figs. 12 and 13 showed an increase in relative errors during winter, but because the production is low during

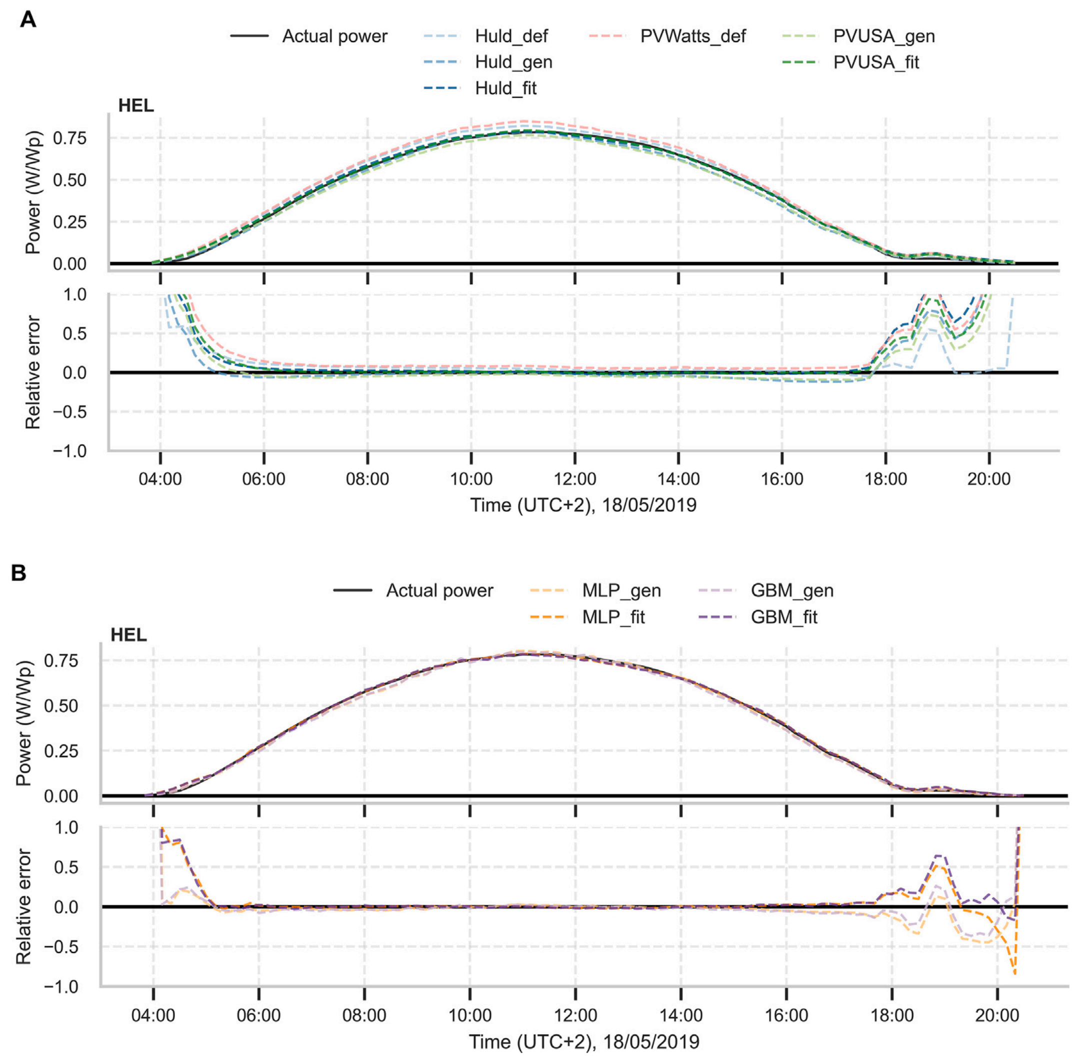

Fig. 11. Power profiles for empirical models (A) and ML models (B) and their relative errors in Helsinki for a sunny day (18/05/2019) when all filters are applied.

winter, even small absolute errors can produce high relative errors. Due to this phenomenon, this study focused primarily on months with high production, and ignored the clipping of relative errors in low production months. Fig. 13(A) shows how the systematic overestimation of power accumulates to modelled energy yields: the modelled monthly productions are greater than the grey bars, which display the actual production. The default PVWatts model produced the largest monthly energy yield errors in all systems. On the other hand, the default Huld model often had the lowest errors during winter in the HEL system, whereas in the SOT-20 system, the default Huld model produced high errors during winter along with the PVWatts model. These conflicting results indicate a lack of robustness in the default Huld model.

The largest errors were in SOT-20 and SOT-90 systems where the default PVWatts model systematically overpredicted the energy yield by around 25%, while general MLP and GBM models overpredicted energy yield by around 10% during the highest production months. This overestimation with general models suggests the possibility of underperformance of SOT-20 and SOT-90 systems, possibly due to poor efficiency of the micro-inverters. Even though poor inverter efficiency mostly affects the AC readings of the power, DC readings can also be affected by the inverter [55].

In the HEL and KUO systems, general models systematically underestimate energy yields by around 5%. Part of this underestimation may be caused by using data from possibly underperforming SOT-20 and SOT-90 systems to train general models for the HEL and KUO systems. The magnitudes of monthly energy yield errors are generally smaller with general models than with default Huld and PVWatts models, except for the HEL and KUO systems, where the default Huld model generally performs better.

Fine-tuning reduces energy yield errors, which results in all models being consistently within 10% of the actual yield, including in the SOT-20 system, during high production months. For the HEL system, errors of the fine-tuned Huld, PVUSA, GBM and MLP models are often below 5%, even under 1%. In the SOT-20 system, the fine-tuned Huld and PVUSA models had around 5% errors, whereas the fine-tuned GBM and MLP models produced fewer errors, often around 2%. Numerical values for monthly errors are shown in SI Section S7.

With all models and systems, errors increase during winter and decrease during summer. This pattern indicates that the models face challenges when dealing with non-stationary seasonality. However, this pattern is practically inevitable, since even small differences between the actual and predicted values lead to large relative errors when the production values are small.

The errors in monthly levels directly influence annual energy yield errors, as shown in Fig. 14. Because the HEL data begin in December 2015 and fine-tuned models require one year of data, the years from 2017 onwards are shown in Fig. 14(A) to focus on errors of full years. Similarly, Fig. 14(B) discards the year 2018. All years and systems are

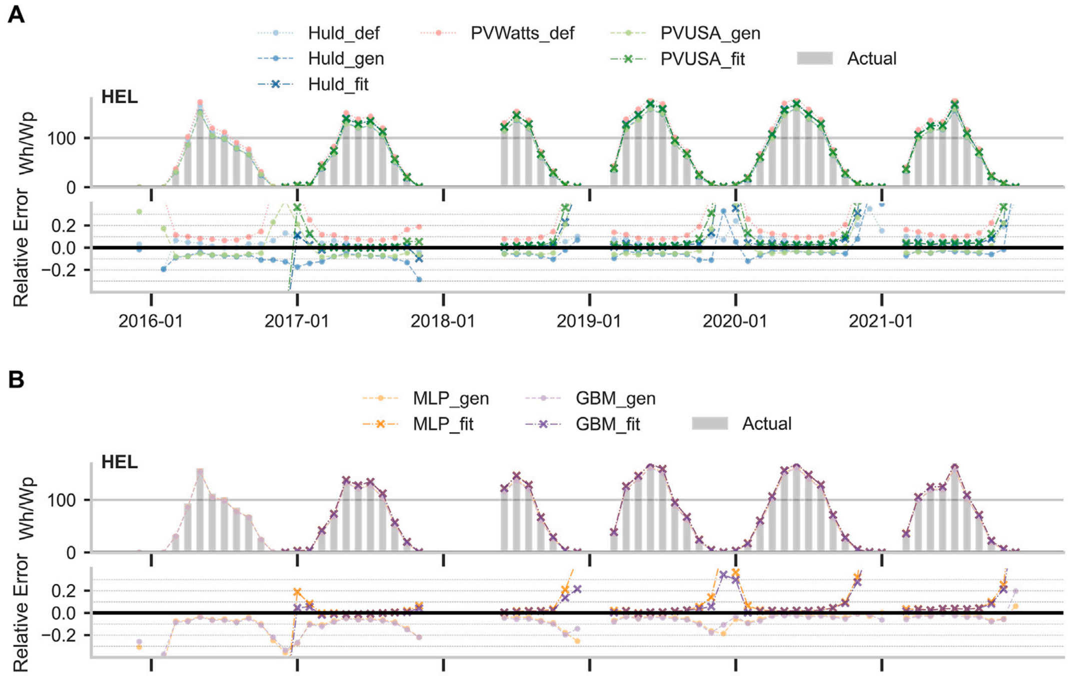

Fig. 12. Monthly energy yields (upper figure) and their relative errors (lower figure) for empirical models (A) and ML models (B) for the HEL system when all filters are applied. Actual energy yields are marked with grey bars. The relative error graph has lines for ±10%, ±20%, and ±30% errors.

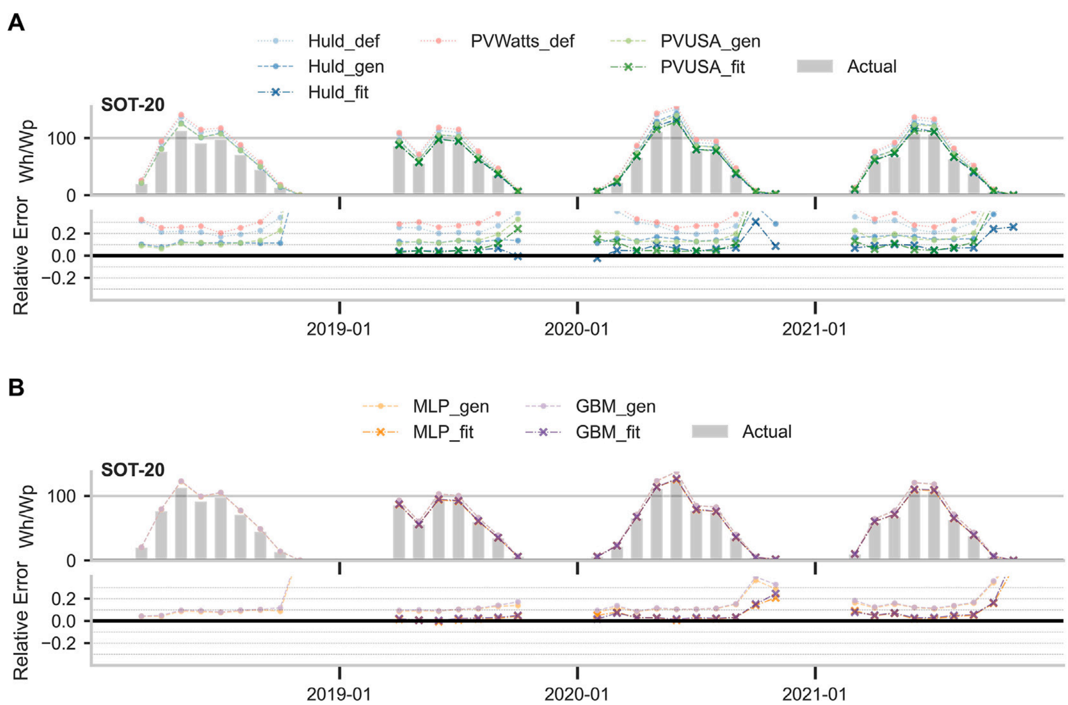

Fig. 13. Monthly energy yields (upper figure) and their relative errors (lower figure) for empirical models (A) and ML models (B) for the SOT-20 system when all filters are applied. Actual energy yields are marked with grey bars. The relative error graph has lines for ±10%, ±20%, and ±30% errors.

shown in SI Figure S24 and results with only the daytime filter are shown in SI Figure S25.

Fig. 14 shows that the overestimation of power from the default PVWatts model resulted in a consistent annual energy yield overestimation. For the HEL system, this error is close to 10%, whereas for the SOT-20 system, this error is around 30%. These errors are significant, and would lead to either false-positive fault detection in the monitoring context, or large economic losses, if used to predict the energy yield of a new system. On the other hand, general Huld, PVUSA and ML models have different error trends between the systems. While these models underestimate the energy yield by around 5% in the HEL system, they overestimate the energy yield by 10% to 15% in the SOT-20 system. The default Huld model shows errors around 5% for the HEL system and around 25% for the SOT-20 system. Fine-tuning lowers the errors.

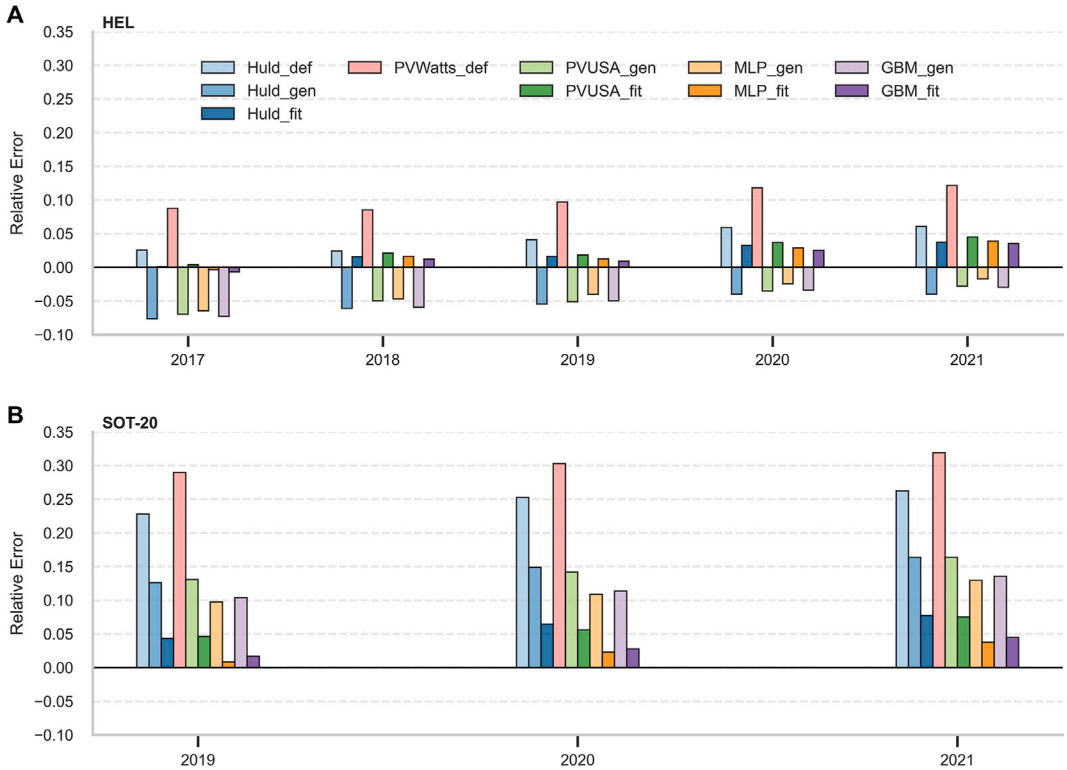

Fig. 14. Annual energy yield errors for HEL (A) and SOT-20 (B) systems when all filters are applied. Positive values represent overestimation of energy, and negative values represent underestimation of energy.

With the SOT-20 system, fine-tuned ML models result in the lowest errors, while fine-tuned Huld and PVUSA show noticeably larger errors. A similar trend is missing in the HEL system, in which all fine-tuned models produced similar errors under 5%. The errors seem to shift towards positive values over the years in all systems, suggesting challenges regarding the non-stationarity of the time-series. The reason is likely a higher PLR than the used −0.33%/year, which causes the models to overestimate the power as time passes.

These differences in errors between models highlight the previous observation, that training a general model affects its accuracy. The general Huld and PVUSA models were trained similarly to the ML models, and the error trends were similar to the ML models in both systems. In addition, the differences in errors between the systems highlight the previous observation that the systems might have characteristics, such as shading or a low-performing inverter. To minimise inverter performance issues, this study focused on DC power output rather than AC power output. However, the inverter might still affect the measured DC power by, for example, clipping the measured DC power output [55]. These site-specific characteristics amplify the errors of the default and general models, underlining the importance of fine-tuning.

To reduce the effects of low data quality, strict filtering was used. Measurements from local weather stations were used to avoid the uncertainty associated with satellite-based measurements. These factors might limit the applicability of the results for smaller-scale PV systems, for example at the household level. For example, the snow filtering used for HEL, KUO, SOT-20, and SOT-90 utilised validation from system photographs [7]. However, similar differences between models were observed with only daytime filtered data. This result means that the findings are likely applicable with different filtering possibilities.

The training length for fine-tuned models was one year in order to take into account the strong seasonality of high-latitude conditions. Even though the length of training data could limit the applicability of the results to newly constructed systems, the analysed default and general models offer good overall performance, thus they can be used when there is not enough historical data for fine-tuning.

4. Conclusions

The aim of this work was to compare the accuracy of different power models in high-latitude conditions. To achieve that aim, the objectives were to analyse how the models perform as general and fine-tuned models, and how the accuracy of the models changes across irradiance levels in various locations and with varying levels of filtering. Despite the challenges in power modelling in high-latitude conditions, high accuracies were obtained with a novel combination of filtering and training: high-quality data with transfer learning resulted in general models with R^2 values of 0.944–0.994, and superior accuracy compared to the default models with only daytime filtering applied, which had R^2 values even as low as 0.78. Fine-tuning with site-specific high-quality data increased the R^2 values to 0.966–0.997. Filtering had little effect on the SOT-90 system, possibly due to vertical installation, which effectively removes snow cover and the resulting noise.

The default and general models had different error trends: all general models showed high R^2 values over 0.944 when all filters were applied, whereas default Huld and PVWatts models showed varying R^2 values across systems, from 0.869 to 0.990. In addition, the default PVWatts model consistently overestimated the power in all irradiance ranges, which also accumulated the monthly energy yield and resulted in the overestimation of the annual energy yield by up to 30%. On the other hand, the default Huld model had high accuracy on the HEL and KUO systems, especially at low irradiances, but the lowest accuracy at irradiance levels over 100 W/m^2 in the SOT-20, SOT-90, and TKU systems, apart from the PVWatts model. All general models avoided such systematic over-/underestimation trends across the systems, indicating that transfer learning results in robust general models. Thus, using data from other systems to construct general power models proved to be beneficial.

Fine-tuning increased R^2 values and generally lowered errors in all models, making their overall accuracy comparable. Even when the overall accuracy of the fine-tuned models showed only a minor increase in R^2 values, the errors in specific irradiance levels decreased by up to 80%. This result underscores the need for more detailed analysis than merely reporting error metrics over the whole dataset when comparing model accuracy. The empirical PVUSA had consistently similar accuracy to ML models, showing that the increased complexity did not offer a significant increase in accuracy. However, fine-tuned ML models had lower nRMSE errors than empirical models in low irradiance levels under 100 W/m^2, highlighting their potential for low-light monitoring.

CRediT authorship contribution statement

Väinö Anttalainen: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Lauri Karttunen:** Writing – review & editing, Visualization, Software, Methodology, Data curation, Conceptualization. **Sami Jouttijärvi:** Writing – review & editing, Supervision, Conceptualization. **Anders V. Lindfors:** Writing – review & editing, Resources. **Juha A. Karhu:** Writing – review & editing, Resources, Data curation. **Hugo Huerta:** Writing – review & editing, Resources, Data curation. **Samuli Ranta:** Writing – review & editing, Resources. **Tarmo Lipping:** Writing – review & editing, Supervision, Methodology. **Kati Miettunen:** Writing – review & editing, Supervision, Funding acquisition.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered potential competing interests:

Co-author served as an Associate Editor for Solar Energy during 2020–2023 (A. V. Lindfors). Given his previous role as Associate Editor of Solar Energy, he had no involvement in the peer review of this article and had no access to information regarding its peer review. Full responsibility for the editorial process for this article was delegated to another journal editor. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The work was funded by the University of Turku and the city of Salo (project HEMS) (LK, VA), and the Strategic Research Council within the Research Council of Finland, Decision No. 358542 (SJ, KM), Decision No. 358543 (JK, AL), and Decision No. 359141 (SR, HH). LK is also grateful for the funding from the University of Turku Graduate School (UTUGS). The authors acknowledge CSC – IT Center for Science, Finland, for computational resources.

Appendix A. Supplementary data

Supplementary data for this article can be found online at doi:10.1016/j.solener.2026.114785.

Data availability

FMI's data (HEL, KUO, SOT) [33] and TUAS's data (TKU) [34] are openly available.

References

[1] International Renewable Energy Agency, Renewable Power Generation Costs in 2024, International Renewable Energy Agency IRENA, 2025.

[2] Suomen uusiutuvat, Suunnittelussa olevat aurinkovoimahankkeet, 2026, https://suomenuusiutuvat.fi/aurinkovoima/aurinkovoimahankkeet-ja-voimalat-suomessa/suunnittelussa-olevat-aurinkovoimahankkeet/ (Accessed: 28 April 2026).

[3] S. Jouttijärvi, L. Karttunen, S. Ranta, K. Miettunen, Techno-economic analysis on optimizing the value of photovoltaic electricity in a high-latitude location, Appl. Energy 361 (2024) 122924, https://doi.org/10.1016/j.apenergy.2024.122924

[4] V. Shekar, D. Abraham, A. Caló, E. Pongrácz, Determining the optimal azimuth for solar-ready buildings: simulating for maximising the economic value of solar PV installations in Lapland, Finland, Renew. Energy 241 (2025) 122357, https://doi.org/10.1016/j.renene.2025.122357

[5] S. Jouttijärvi, J. Thorning, M. Manni, H. Huerta, S. Ranta, M. Di Sabatino, G. Lobaccaro, K. Miettunen, A comprehensive methodological workflow to maximize solar energy in low-voltage grids: a case study of vertical bifacial panels in Nordic conditions, Sol. Energy 262 (2023) 111819, https://doi.org/10.1016/j.solener.2023.111819

[6] J. Varjopuro, A. Kamppinen, A. Poskela, J.A. Karhu, A.V. Lindfors, K. Miettunen, Computational simulation of perovskite and silicon solar panel operating temperatures in varying ambient conditions, Sol. Energy Mater. Sol. Cells 290 (2025) 113657, https://doi.org/10.1016/j.solmat.2025.113657

[7] J.A. Karhu, A.V. Lindfors, W. Wandji Nyamsi, T. Salola, A. Poikonen, M.R.A. Pitkänen, T. Mielonen, O. Mantikka, Photovoltaic power and meteorological datasets with snow detection from the outdoor solar power laboratories of the Finnish meteorological institute, Geosci. Data J. 13 (2026), https://doi.org/10.1002/gdj3.70039

[8] H. Böök, A. Poikonen, A. Aarva, T. Mielonen, M.R.A. Pitkänen, A.V. Lindfors, Photovoltaic system modeling: a validation study at high latitudes with implementation of a novel DNI quality control method, Sol. Energy 204 (2020) 316–329, https://doi.org/10.1016/j.solener.2020.04.068

[9] C. Brester, V. Kallio-Myers, A.V. Lindfors, M. Kolehmainen, H. Niska, Evaluating neural network models in site-specific solar PV forecasting using numerical weather prediction data and weather observations, Renew. Energy 207 (2023) 266–274, https://doi.org/10.1016/j.renene.2023.02.130

[10] H. Böök, A.V. Lindfors, Site-specific adjustment of a NWP-based photovoltaic production forecast, Sol. Energy 211 (2020) 779–788, https://doi.org/10.1016/j.solener.2020.10.024

[11] L. Karttunen, S. Jouttijärvi, A. Poskela, H. Palonen, H. Huerta, M. Todorović, S. Ranta, K. Miettunen, Comparing methods for the long-term performance assessment of bifacial photovoltaic modules in Nordic conditions, Renew. Energy 219 (2023) 119473–. https://doi.org/10.1016/j.renene.2023.119473

[12] R. Platon, J. Martel, N. Woodruff, T.Y. Chau, Online fault detection in PV systems, IEEE Trans. Sustain. Energy 6 (4) (2015) 1200–1207, Conference Name: IEEE Transactions on Sustainable Energy. https://doi.org/10.1109/TSTE.2015.2421447

[13] A. Livera, A. Phinikarides, G. Makrides, J. Sutterlueti, G. Georghiou, Advanced failure detection algorithms and performance decision classification for grid-connected PV systems, 33rd European Photovoltaic Solar Energy Conference and Exhibition; 2358-2363 (2017), https://doi.org/10.4229/EUPVSEC20172017-6BV.2.13

[14] M.B. Øgaard, H.N. Riise, H. Haug, S. Sartori, J.H. Selj, Photovoltaic system monitoring for high latitude locations, Sol. Energy 207 (2020) 1045–1054, https://doi.org/10.1016/j.solener.2020.07.043

[15] S. Silvestre, A. Chouder, E. Karatepe, Automatic fault detection in grid connected PV systems, Sol. Energy 94 (2013) 119–127.

[16] A. Chouder, S. Silvestre, Automatic supervision and fault detection of PV systems based on power losses analysis, Energy Convers. Manag. 51 (10) (2010) 1929–1937, https://doi.org/10.1016/j.enconman.2010.02.025, https://www.sciencedirect.com/science/article/pii/S0196890410000919.

[17] I. de la Parra, M. Muñoz, E. Lorenzo, M. García, J. Marcos, F. Martínez-Moreno, PV performance modelling: a review in the light of quality assurance for large PV plants, Renew. Sustain. Energy Rev. 78 (2017) 780–797, https://doi.org/10.1016/j.rser.2017.04.080

[18] T. Yildiz, M. Gol, A malfunction detection method for PV systems, in: 2019 IEEE Milan PowerTech, 2019, pp. 1–6, https://doi.org/10.1109/PTC.2019.8810946

[19] B.D. Dimd, S. Voller, U. Cali, O.-M. Midtgard, A review of machine learning-based photovoltaic output power forecasting: Nordic context, IEEE Access 10 (2022) 26404–26425, https://doi.org/10.1109/ACCESS.2022.3156942

[20] S. Jouttijärvi, M. Tok, L. Karttunen, H. Huerta Medina, S. Ranta, K. Miettunen, Modeling and measuring the power output of vertical bifacial solar panels in Nordic conditions, 8th World Conference on Photovoltaic Energy Conversion; 507-511 (2022), https://doi.org/10.4229/WCPEC-82022-3BO.14.2

[21] S. Al-Dahidi, M. Alrbai, H. Alahmer, B. Rinchi, A. Alahmer, Enhancing solar photovoltaic energy production prediction using diverse machine learning models tuned with the chimp optimization algorithm, Sci. Rep. 14 (1) (2024) 18583, https://doi.org/10.1038/s41598-024-69544-8

[22] A. Keddouda, R. Ihaddadene, A. Boukhari, A. Atia, M. Arıcı, N. Lebbihiat, N. Ihaddadene, Solar photovoltaic power prediction using artificial neural network and multiple regression considering ambient and operating conditions, Energy Convers. Manag. 288 (2023) 117186, https://doi.org/10.1016/j.enconman.2023.117186

[23] M. Abdelsattar, A. AbdelMoety, A. Emad-Eldeen, Advanced machine learning techniques for predicting power generation and fault detection in solar photovoltaic systems, Neural Comput. Appl. (Feb 2025), https://doi.org/10.1007/s00521-025-11035-6

[24] A. Alcañiz, A.V. Lindfors, M. Zeman, H. Ziar, O. Isabella, Effect of climate on photovoltaic yield prediction using machine learning models, Glob. Chall. 7 (1) (2023) 2200166, _eprint: https://doi.org/10.1002/gch2.202200166, https://onlinelibrary.wiley.com/doi/pdf/10.1002/gch2.202200166.

[25] C. Hajjaj, A. Alami Merrouni, A. Bouaichi, M. Benhmida, S. Sahnoun, A. Ghennioui, H. Zitouni, Evaluation, comparison and experimental validation of different PV power prediction models under semi-arid climate, Energy Convers. Manag. 173 (2018) 476–488, https://doi.org/10.1016/j.enconman.2018.07.094

[26] T. Huld, G. Friesen, A. Skoczek, R.P. Kenny, T. Sample, M. Field, E.D. Dunlop, A power-rating model for crystalline silicon PV modules, Sol. Energy Mater. Sol. Cells 95 (12) (2011) 3359–3369, https://doi.org/10.1016/j.solmat.2011.07.026
[27] E. Sarmas, N. Dimitropoulos, V. Marinakis, Z. Mylona, H. Doukas, Transfer learning strategies for solar power forecasting under data scarcity, Sci. Rep. 12 (1) (2022) 14643, https://doi.org/10.1038/s41598-022-18516-x
[28] A. Bellagarda, D. Grassi, A. Aliberti, L. Bottaccioli, A. Macii, E. Patti, Effectiveness of neural networks and transfer learning to forecast photovoltaic power production, Appl. Soft Comput. 149 (2023) 110988, https://doi.org/10.1016/j.asoc.2023.110988
[29] G. Anamiati, L. Landberg, G. Guerra, P.M. Ruiz, Benchmarking photovoltaic plant performance: a machine learning model using multi-dimensional neighbouring plants, EPJ Photovolt. 15 (2024) 27, https://doi.org/10.1051/epjpv/2024023
[30] L. Deotti, I. Silva Júnior, L. Honório, A. Marcato, Empirical models applied to distributed energy resources—an analysis in the light of regulatory aspects, Energies 14 (2) (2021) 326, https://doi.org/10.3390/en14020326
[31] A. Livera, M. Theristis, G. Makrides, S. Ransome, J. Sutterlueti, G.E. Georghiou, Optimal development of location and technology independent machine learning photovoltaic performance predictive models, in: 2019 IEEE 46th Photovoltaic Specialists Conference (PVSC), 2019, pp. 1270–1275, https://doi.org/10.1109/PVSC40753.2019.8980474
[32] V. Anttalainen, Aurinkoenergiajärjestelmän tehon mallintaminen: fysikaalisten ja koneoppimispohjaisten tehomallien vertailu (Master's thesis), Tampere University, 2025, https://trepo.tuni.fi/handle/10024/225143.
[33] J.A. Karhu, A.V. Lindfors, PV production data with ancillary PV and meteorological data including solar radiation measurements from FMI's outdoor solar laboratories in Helsinki, Kuopio and Sodankylä (Finland) starting from August 2015 and ending dec 2021, 2025, https://doi.org/10.57707/fmi-b2share.c7b3ec9f19324a6497639d6d8c0cde84
[34] New Energy Research Center Turku, NERC's data sharing and document library – new energy research center Turku, 2025, https://nerc.turkuamk.fi/data-portal/ (Accessed: 9 October 2025).
[35] L. Karttunen, V. Anttalainen, S. Jouttijärvi, J.A. Karhu, H. Huerta, S. Ranta, A.V. Lindfors, K. Miettunen, Robust performance loss rate calculation for photovoltaic systems in high latitude locations, Accepted for publication in Progress in Photovoltaics (2026), https://doi.org/10.1002/pip.70130
[36] Finnish Meteorological Institute, FMI open PV forecast, 2025, https://github.com/fmidev/fmi-open-pv-forecast (Accessed: 12 March 2025).
[37] A.S. Gardner, M.J. Sharp, A review of snow and ice albedo and the development of a new physically based broadband albedo parameterization, J. Geophys. Res. Earth Surf. 115 (F1) (2010), https://doi.org/10.1029/2009JF001444
[38] Finnish Meteorological Institute, Havaintojen lataus, 2025, https://www.ilmatieteenlaitos.fi/havaintojen-lataus (Accessed: 24 September 2025).
[39] R. Perez, P. Ineichen, R. Seals, J. Michalsky, R. Stewart, Modeling daylight availability and irradiance components from direct and global irradiance, Sol. Energy 44 (5) (1990) 271–289, https://doi.org/10.1016/0038-092X(90)90055-H
[40] D.G. Erbs, S.A. Klein, J.A. Duffie, Estimation of the diffuse radiation fraction for hourly, daily and monthly-average global radiation, Sol. Energy 28 (4) (1982) 293–302, https://doi.org/10.1016/0038-092X(82)90302-4
[41] N. Martin, J.M. Ruiz, Calculation of the PV modules angular losses under field conditions by means of an analytical model, Sol. Energy Mater. Sol. Cells 70 (1) (2001) 25–38, https://doi.org/10.1016/S0927-0248(00)00408-6
[42] J. Kratochvil, W. Boyson, D. King, Photovoltaic array performance model, Tech. Rep. SAND2004-3535, 919131, Sandia National Laboratories (Aug 2004, https://doi.org/10.2172/919131
[43] IEC 61724-3, Photovoltaic system performance - Part 3: Energy evaluation method, 2016.
[44] S. Lindig, D. Moser, A.J. Curran, K. Rath, A. Khalilnejad, R.H. French, M. Herz, B. Müller, G. Makrides, G. Georghiou, A. Livera, M. Richter, J. Ascencio-Vásquez, M. van Iseghem, M. Meftah, D. Jordan, C. Deline, W. van Sark, J.S. Stein, M. Theristis, B. Meyers, F. Baumgartner, W. Luo, International collaboration framework for the calculation of performance loss rates: data quality, benchmarks, and trends (towards a uniform methodology), Prog. Photovolt. Res. Appl. 29 (6) (2021) 573–602, https://doi.org/10.1002/pip.3397
[45] E.M. Tonita, D.C. Jordan, S. Ovaitt, H. Toal, K. Hinzer, C. Pike, C. Deline, Long-term photovoltaic system performance in cold, snowy climates, Prog. Photovolt. Res. Appl. 33 (11) (2025) 1180–1197, https://doi.org/10.1002/pip.70014
[46] R. Steele, PVUSA progress report, 1989–1990, 1990, https://doi.org/10.2172/10163193
[47] A. Dobos, PVWatts version 5 manual, Tech. Rep. NREL/TP-6A20-62641, 1158421, NREL (Sep 2014, https://doi.org/10.2172/1158421
[48] T.J. Hastie, R. Tibshirani, J.H. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer series in statistics, second ed, Springer, New York, 2009.
[49] C. Persson, P. Bacher, T. Shiga, H. Madsen, Multi-site solar power forecasting using gradient boosted regression trees, Sol. Energy 150 (2017) 423–436, https://doi.org/10.1016/j.solener.2017.04.066
[50] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-Learn: machine learning in Python, J. Mach. Learn. Res. 12 (85) (2011) 2825–2830.
[51] V. Anttalainen, L. Karttunen, PV Power Modelling in High Latitudes, 2026, Available: https://github.com/energy-systems-team/pv-power-modelling-in-high-latitudes (Accessed: 11 June 2026).
[52] A. Alcañiz, D. Grzebyk, H. Ziar, O. Isabella, Trends and gaps in photovoltaic power forecasting with machine learning, Energy Rep. 9 (2023) 447–471, https://doi.org/10.1016/j.egyr.2022.11.208
[53] F. Mavromatakis, F. Vignola, B. Marion, Low irradiance losses of photovoltaic modules, Sol. Energy 157 (2017) 496–506, https://doi.org/10.1016/j.solener.2017.08.062, https://www.sciencedirect.com/science/article/pii/S0038092X17307430.
[54] C. Reise, B. Müller, D. Moser, G. Belluardo, P. Ingenhoven, Uncertainties in PV system yield predictions and assessments, Tech. rep., Report IEA-PVPS T13-12:2018, IEA-PVPS (Apr 2018.
[55] J. Yuventi, A method for evaluating the influence of wiring on the performance of components in a photovoltaic system design, Sol. Energy 86 (10) (2012) 2996–3003, https://doi.org/10.1016/j.solener.2012.07.007