nature energy

Resource

https://doi.org/10.1038/s41560-026-02070-1

The PAINT database for operational concentrating solar power plant data following FAIR data principles

Received: 16 October 2025

Accepted: 20 April 2026

Published online: 16 June 2026

Check for updates

Kaleb Phipps[1,2] ✉, Mathias Kuhl[3], Marie Weiel[1,2], Marlene Busch[3], Jan Lewen[3], Nicolas Blumenröhr[1,4], Daniel Maldonado Quinto[3], Charlotte Debus[1], Felix Göhring[3], Oliver Kaufhold[3], Achim Streit[1], Robert Pitz-Paal[3,5], Markus Götz[1,2] & Max Pargmann[3]

Integrating renewable energy is critical for maintaining grid stability while striving towards global climate goals. Concentrating solar power tower plants offer a promising solution but their competitiveness is currently hindered by high operational costs, limited data availability and slow adoption of emerging technologies. To address these barriers, here we introduce PAINT, a FAIR (findable, accessible, interoperable and reusable) open-access database for operational solar tower plant data. PAINT provides 849 GB of high-resolution data collected over multiple years from the Jülich solar tower plant, including heliostat properties, calibration and deflectometry measurements and fine-grained weather data. The database is organized using the SpatioTemporal Asset Catalog metadata specification and supports the development of digital twins, artificial intelligence-based calibration methods, predictive maintenance and improved solar flux prediction. We also introduce standardized benchmarks to promote reproducibility and fair comparisons. Providing access to high-quality data, PAINT enables broader participation in solar research, accelerates innovation and facilitates data-driven solutions in solar tower power plant research.

The transition to a renewable-based energy system is essential to meeting the United Nations' climate mandates[1–3]. Concentrating solar power (CSP) plants, especially in the solar tower configuration, are one promising solution for realizing these goals[4]. These power plants use an array of up to 10,000 mirrors, known as heliostats, to focus sunlight onto a receiver and heat a medium to temperatures over 700 °C (refs. 5–7), with the resulting thermal energy converted into electricity, used directly in industrial processes or stored in cheap thermal storage systems[7–9]. However, there are still challenges limiting their widespread deployment, such as high initial deployment costs[10], limited geographical locations[11], heliostat alignment[12], heliostat imperfections[13] and the real-time coordination and optimization of heliostat aim points[14–17].

In the current research cycle, these challenges are addressed by researchers developing new designs or algorithms on the basis of existing literature, simulating or testing these concepts at a research facility and publishing their results to update current literature. While some solutions reach commercial power plants, the process is hindered by operators' reluctance to adopt unproven technologies, largely owing to the high operational and economic risk in adapting existing systems

[1]Scientific Computing Center, Karlsruhe Institute of Technology, Eggenstein-Leopoldshafen, Germany. [2]Helmholtz AI, Karlsruhe, Germany. [3]Institute of Solar Research, German Aerospace Center, Cologne, Germany. [4]Helmholtz Metadata Collaboration, Karlsruhe, Germany. [5]Chair of Solar Technology, RWTH Aachen University, Aachen, Germany. ✉e-mail: kaleb.phipps@kit.edu

https://doi.org/10.1038/s41560-026-02070-1

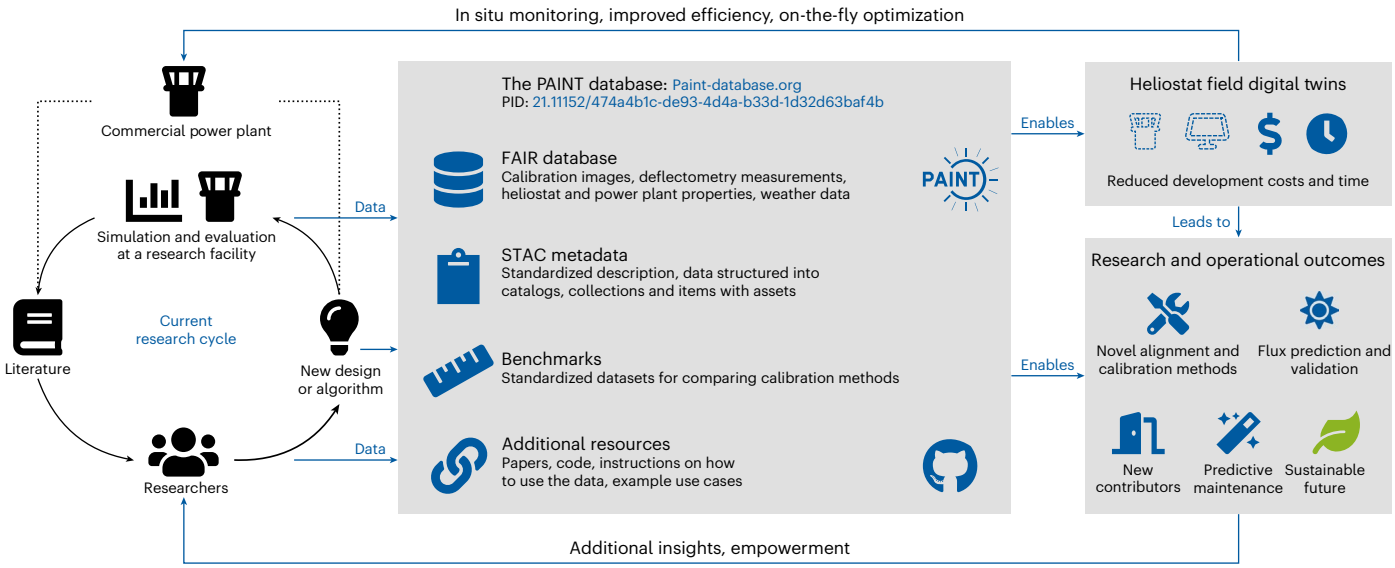

Fig. 1 | Advancing the research cycle for solar tower power plants. An overview of the current research cycle for solar tower power plants and how the PAINT database expands it by providing FAIR data, comprehensive metadata, standardized benchmarks, papers and code. Black lines refer to the current research cycle, while blue lines highlight the inclusion of PAINT and insights gained from the data. Credit: icons, Font Awesome under the SIL OFL 1.1 license.

and procedures[12,18]. This lack of trust frequently stems from research based on limited, cherry-picked data from a single measurement campaign that fails to accurately reflect real-world operational conditions of these or other power plants[18–20].

Open data have already been identified in the roadmap for advanced heliostat technologies[21,22] as a bottleneck hindering the realization of the true potential of CSP technology. Open data are particularly important for modern tools such as digital twins[23] and artificial intelligence (AI)-based algorithms[24], which can revolutionize CSP research. AI-based methods have already been shown to outperform traditional methods in tasks such as heliostat calibration[13,25,26] and flux-density prediction[27,28] but remain unfeasible without sufficient high-quality and well-documented data for training. In addition, these data must be easily findable, accessible to researchers, interoperable across tools and platforms and reusable (FAIR)[29]. Despite their critical importance, operational data from solar tower plants remain largely inaccessible. Many factors, including commercial confidentiality, intellectual property concerns and the absence of clear incentives or frameworks for data sharing, drive this scarcity. One initiative for accessible data is the OpenCSP project[30,31], which provides a collaborative environment for pooling open-source code, tools and mechanical documentation to support CSP research and education. While OpenCSP serves as a valuable repository for diverse community resources, it does not provide a unified data architecture; instead, it aggregates distinct datasets from various sources. Each dataset utilizes unique schemas and documentation standards suited to specific project goals[31], which complicates their utility for data-driven applications. Furthermore, OpenCSP currently relies on file-hosting services (for example, Box) that require manual interaction, making programmatic access and automated data loading challenging.

To address these limitations, we introduce PAINT, a FAIR database for operational CSP plant data. PAINT provides 849 GB of operational data from a solar tower power plant in Jülich, Germany. These data, covering over 2,000 heliostats, are supplemented by local weather records, extensive metadata adhering to the Spatiotemporal Asset Catalog (STAC) specification and software to enable programmatic access via a Python package. Unlike previous initiatives, PAINT is designed as a systematic, temporally continuous database specifically engineered for data-driven applications. The PAINT database supports the development of heliostat field digital twins, which serve both as platforms for testing algorithms and as tools for in situ monitoring of existing solar tower power plants (Fig. 1). Integrated with a digital twin, PAINT enables all types of beam characterization system advancements in heliostat calibration and alignment, surface reconstructions, improved flux-density prediction and data-driven predictive maintenance, for example, soiling detection—each with the potential to reduce operating costs considerably. As a result, standardized benchmarks will permit an unbiased comparison of algorithms developed across institutions and equipment, allowing cost-effective and trustworthy benchmarking that will lower the barrier for commercial integration. Moreover, by providing access to high-quality data, PAINT opens the field to a wider pool of researchers, including those without access to experimental facilities, thereby strengthening the international CSP research community and accelerating progress towards a sustainable energy future.

Details of the PAINT database

We collect operational data at the Jülich solar tower power plant from December 2020 to June 2024, comprising measurements from 2,014 heliostats used during this period. The PAINT database contains five main categories of data: tower measurements, calibration data, deflectometry data, heliostat properties and weather data (Fig. 2). The tower measurements data contain information on the solar tower, including the unique global identifier associated with the Jülich power plant, the properties of the various targets used for calibration measurements and the properties and coordinates of the receiver. Weather data are collected at a 1-s resolution from a weather tower at the Jülich tower as well as at a 1-h resolution from the Aachen-Orsbach German Weather Service (DWD) weather station, the closest DWD station to the Jülich tower with data available for the desired period. These weather data include variables relevant to solar tower operation, including cloud cover, irradiation and temperature values. The calibration, deflectometry and heliostat properties data are heliostat-specific data. Calibration data contain both raw and preprocessed solar flux images of the calibration target, taken during calibration measurements. In addition, we extract the focal spots and associated motor positions and provide the azimuth and elevation of the sun at the time of the calibration measurement. The deflectometry data comprise high-resolution point cloud and normal vector measurements obtained during deflectometry measurement campaigns. These measurements are supported by a summary file and provided as both raw measurements and in a preprocessed form, where missing vectors

https://doi.org/10.1038/s41560-026-02070-1

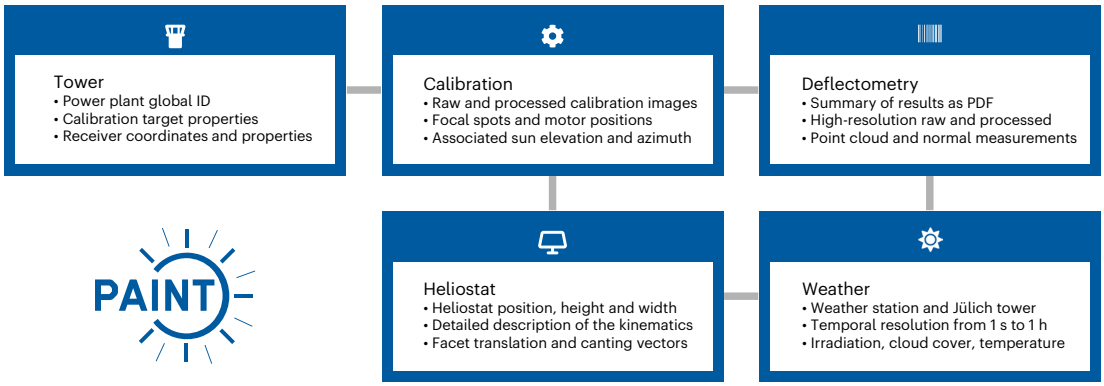

Fig. 2 | The data categories in the PAINT database. Overview of the main categories of data provided via the database and a short description of the concrete data available. Credit: icons, Font Awesome under the SIL OFL 1.1 license.

are filled with assumed ideal vectors, that is, vectors perpendicular to the surface signifying no deformation. Heliostat properties data contain key information about each individual heliostat, including its location, geometric description, overview of the kinematics and facet properties. The data coverage in PAINT is comprehensive: distinct property data are available for the entire field, and calibration data cover 1,893 heliostats (approximately 94%). With deflectometry measurement available for a subset of 471 heliostats, this represents an unprecedented volume of open high-fidelity surface data for the CSP community, particularly given the resource-intensive nature of deflectometry campaigns. The number of available calibration measurements varies per heliostat and PAINT includes calibration and deflectometry data from a range of heliostats throughout the field (Fig. 3).

The collected data are combined and preprocessed to a consistent format before being included in the database (for detailed information, see the Methods section). In addition, all metadata are collated into the widely used STAC specification. The STAC specification is an open standard designed to make geospatial data more accessible, searchable and interoperable across different platforms. It applies a lightweight, JSON-based format to describe the metadata and is already widely adopted[32]. STAC organizes the data into a structure of 'items', 'collections' and 'catalogues', enabling users to efficiently browse and query the PAINT database by time and location, either with existing STAC-supported APIs[33] or more efficiently with our own software package. In PAINT, each heliostat is accompanied by a STAC catalogue, while the properties, calibration and deflectometry metadata for each heliostat are collections within this catalogue. The weather data are a separate collection within the main power plant catalogue, which also contains the tower properties item. A detailed overview of the STAC structure is provided in Supplementary Note 2.

All resources are available online via the PAINT database, which will be updated and maintained for the foreseeable future. The accompanying Python package can be installed via PyPI and allows easy access to the data and associated metadata. All code for preprocessing and preparing the database is documented via GitHub. These practices help ensure that PAINT adheres to the FAIR data principles[29].

FAIR data compliance and future roadmap

We are committed to ensuring that PAINT remains findable, accessible, interoperable and reusable, in strict adherence to the FAIR data principles[29]. This section details the specific measures implemented to satisfy these standards and outlines our strategic roadmap for the database's long-term sustainability, maintenance and future expansion.

Adherence to FAIR principles

PAINT's adherence to FAIR principles is achieved by inherent design and data curation decisions and by modelling the STAC metadata assets as FAIR digital objects for advanced machine-actionable decision making[34]. Findability is achieved with the previously mentioned website as well as a unique persistent identifier (PID) that resolves to the main, navigable STAC catalogue. In addition, the STAC specification enables the user to apply metadata-based filters on the data to easily find specific data, for example, only for a certain time period.

As an open-source specification that is clearly documented via GitHub[32] and actively developed by the community, the STAC specification also ensures accessibility. STAC uses the JSON-format, which can be browsed in all common programme languages and multiple tools exist, for example, the Python PySTAC package[35], to automate access. Furthermore, the accompanying PAINT software provides a customized STAC client designed for the database, which enables access with only a few lines of code. PAINT is also resolvable via URL, enabling users to access files without any programming knowledge, and the accompanying PID ensures the main STAC catalogue is accessible even if this URL should change.

Interoperability is achieved through the use of common and broadly used formats such as JSON, Portable Network Graphics (PNG), Hierarchical Data Format version 5 (HDF5)[36] and Portable Document Format (PDF). The resulting datasets are clearly described through the STAC metadata, as well as on the website and in Supplementary Note 1 for this paper. Furthermore, the supporting software ensures PAINT data can be easily integrated into any Python-based application, demonstrated already with an integration in ARTIST[37]—a differentiable ray tracer and solar tower digital twin developed using PyTorch in Python.

All of these aspects enable PAINT data to be inherently reusable. In addition, the data are licensed under the Community Data License Agreement Permissive Version 2.0 (CDLA 2.0), while the associated software is available via a Massachusetts Institute of Technology (MIT) license, which allows the community to freely reuse both data and code. We hope these decisions will help to establish community standards for FAIR data in the CSP research community and foster a collaborative ecosystem supporting novel ideas and applications.

Roadmap for sustainable data management and expansion

While the initial release of PAINT marks a notable step for open CSP research, we envision it as a living resource that will evolve over time. Our long-term goal is for PAINT to serve as a standardized hub for operational data from multiple power plants globally, bridging the gap between research-grade data and commercial operations. To ensure this sustainability and scalability, we have defined the following roadmap:

- Stable infrastructure: The database is hosted via the Large Scale Data Facility (LSDF)[38], a centralized infrastructure designed for the long-term preservation of scientific data. The LSDF provides state-of-the-art redundancy and security, ensuring that the dataset remains persistently accessible as it grows in volume.

https://doi.org/10.1038/s41560-026-02070-1

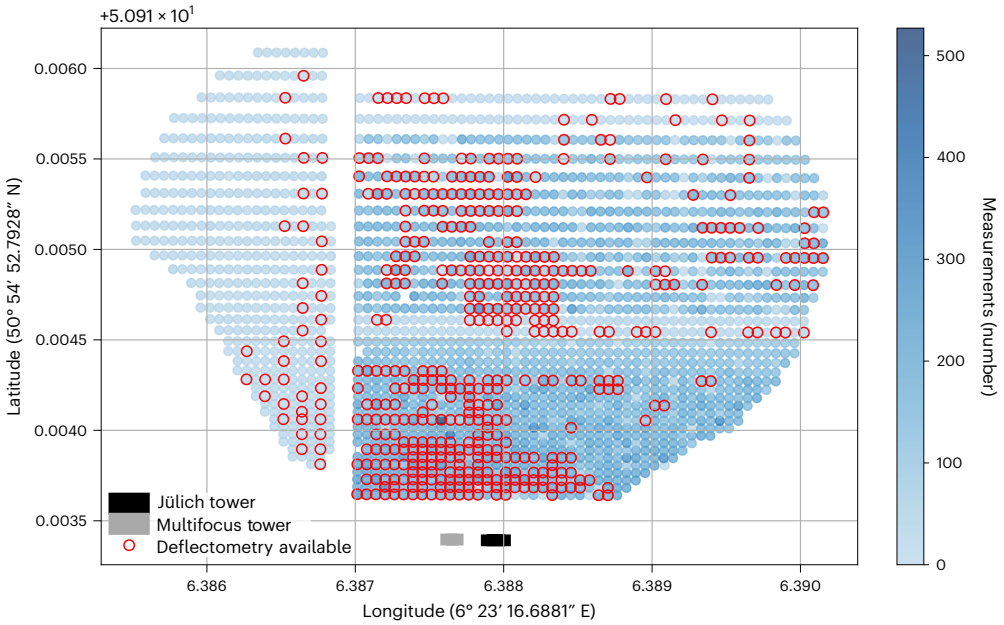

Fig. 3 | The heliostat positions and available data. Overview of the latitude and longitude of all heliostats available in the PAINT database, coloured on the basis of the number of calibration measurements available for that heliostat. Heliostats marked by a red outline also include deflectometry measurements.

- Versioning and reproducibility: We use a combination of STAC architecture and versioned PIDs to manage updates. As STAC decouples metadata (catalogues) from the underlying data assets, we can release new versions of the database by simply publishing updated catalogues that link to new or existing files. This permits efficient, incremental updates without data duplication. Crucially, each version is assigned a unique PID, ensuring that previous versions of the catalogue remain accessible to guarantee the reproducibility of earlier research. For technical details, see Supplementary Note 1.
- Internal expansion: As an active research facility, the Jülich solar tower continues to generate valuable data. We plan to perform periodic updates to PAINT to include an extended temporal scope by adding measurements from ongoing calibration and deflectometry campaigns to fill coverage gaps and capture long-term operational trends. In addition, future research projects may also enable the integration of new modalities and heterogeneous data sources such as solar flux maps or thermal imagery. The flexible nature of the STAC standard allows these diverse data types to be ingested without disrupting the structure of existing assets.
- External integration: We aim to expand the PAINT ecosystem to include data from other research sites and commercial power plants. To maintain strict quality standards and consistency, we do not support automatic uploads. Instead, we invite researchers and plant operators to collaborate with us directly. We will assist external contributors in mapping their data to the PAINT standard, ensuring that the database remains a unified, high-quality resource rather than a fragmented repository.
- Open code ecosystem: The PAINT software is fully open source and hosted on GitHub. We view this as a community-driven tool and explicitly welcome contributions. Users can submit pull requests to improve data access tools or propose features for handling new data types. While the core package is focused on data access, we encourage the community to build and share compatible analysis plugins (for example, for hot-spot detection or advanced calibration) within this ecosystem.

We believe that these governance structures will ensure PAINT remains a cornerstone for CSP research, capable of adapting to future data needs while maintaining rigorous scientific standards.

Example uses of the PAINT database

The key barriers hindering widespread commercial CSP deployment are high initial deployment costs, unreliable performance and operating and maintenance costs[22,39]. The PAINT database addresses these challenges by providing a standardized resource for CSP research. In this section, we outline multiple use cases to demonstrate the dataset's value and versatility. For several of these use cases, we provide sample implementations and preliminary results. These examples are not intended as exhaustive studies but rather as concrete evidence of the dataset's suitability for addressing problems in CSP research and starting points to enable further in-depth exploration by the community. Further technical details on these use cases are provided in the 'Uses of PAINT Data' section of Supplementary Methods.

Heliostat field digital twins

A digital twin is a complete virtual description of a physical object facilitating bidirectional data exchange[40,41]. Accurate heliostat field digital twins can be used in real-time to optimize CSP plant operation via in situ monitoring to maximize performance[13] or in a research setting to evaluate novel solutions and investigate critical scenarios without endangering the integrity of the plant or the researchers involved[16,42,43]. Furthermore, they can be used to simulate the performance of different receiver designs. Digital twin creation depends on extensive and accurate data. PAINT provides detailed information on heliostat locations, heliostat mirror imperfections and kinematics design and parameters. The associated high-resolution weather data enable realistic simulations of challenging scenarios, such as fluctuating irradiance conditions. These data facilitate the creation of heliostat field digital twins with a higher level of precision than previously possible, while also enabling the integration of machine learning models. An illustrative implementation of this is ARTIST[37], which was constructed using PAINT data. This tool has been used to demonstrate improvements in surface characterization and flux-density prediction[13] as well as heliostat calibration[44].

Focal spot centroid detection

A crucial task to facilitate alignment and calibration is focal spot centroid detection in flux-density target images[45]. Raw target images captured by target cameras often contain skewed orientations, variable lighting conditions and off-centre focal spots. Without detecting the focal spot

https://doi.org/10.1038/s41560-026-02070-1

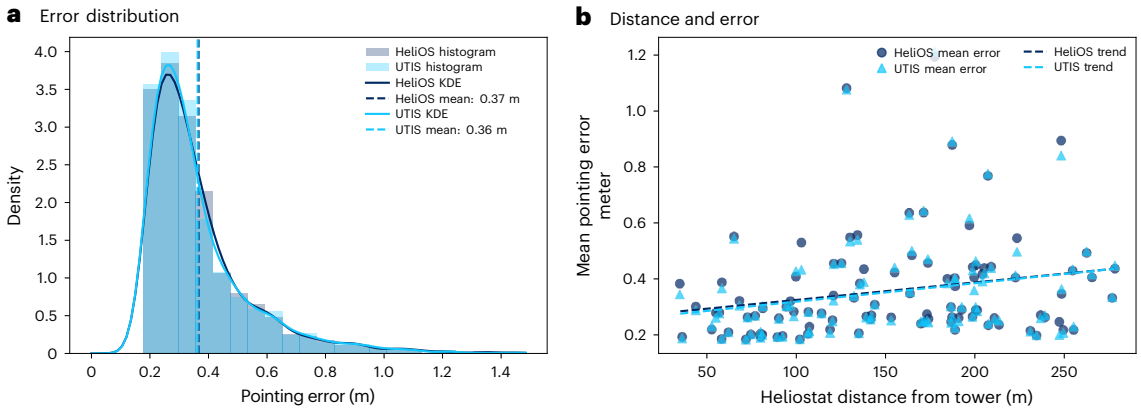

Fig. 4 | Focal spot detection and heliostat calibration and alignment. PAINT provides a large database of high-quality calibration images and associated metadata that can be used to develop, validate and benchmark methods for focal spot detection and heliostat calibration. **a**, The data-driven focal spot centroid detection method UTIS[54] developed with PAINT data outperforms current operational solutions. KDE with a Gaussian kernel is applied to visually simplify the comparison. **b**, The diverse calibration dataset enables robust analysis of calibration error compared with the distance of the heliostat from the tower. Linear regression lines are included to visualize the trend.

centroid, there is no consistent frame of reference and any calibration or alignment algorithm based on these data will be unreliable. PAINT provides a substantial collection of high-quality calibration images, enabling the development, testing and benchmarking of data-driven focal spot centroid detection methods. One example is the AI-based image processing method UTIS[46]. Trained on PAINT data, UTIS demonstrates improved centroid detection accuracy compared with the current operational solution provided by HeliOS[47,48]. This comparison is shown in Fig. 4a, which displays histograms of the pointing errors for both detection methods. To facilitate comparison, we overlay smooth probability density functions generated via kernel density estimation (KDE) using a Gaussian kernel with Scott's rule for bandwidth selection.

Heliostat calibration and alignment

Owing to cost-effective design and production, heliostats are plagued by multiple imperfections and sources of error including optical mirror deformations, facet alignment issues and inaccuracies in the kinematics. Furthermore, the heliostats' conditions gradually degrade over time owing to mechanical wear, structural bending or changing environmental conditions. As a result, heliostats must be constantly calibrated to account for these offsets and optimally aligned to the receiver throughout the day. Improving heliostat calibration and alignment with novel methods, often via digital image processing and AI-integration, is a key focus of CSP research. With over 218,000 calibration images and detailed metadata, PAINT provides the scale necessary for robust development and evaluation of calibration algorithms. This extensive dataset permits unprecedented depth of analysis; for example, we can now quantify the dependency between heliostat-to-tower distance and calibration error. We show this in Fig. 4b, plotting the mean pointing error per heliostat against its distance from the tower. Linear regression lines are included to visualize the trend.

Solar flux prediction and validation

Accurate solar flux prediction is crucial for the efficient and safe operation of solar tower power plants. Precise forecasting of solar flux—the intensity and distribution of sunlight concentrated on the receiver—enables effective thermal management, prevents localized overheating and ensures consistent power generation. Traditional methods for predicting solar flux often rely on complex simulations and are limited in their ability to adapt to dynamic environmental conditions. Recently, AI-based data-driven approaches have enhanced solar flux prediction. By leveraging the calibration images, high-resolution weather data and heliostat position information available in PAINT, researchers

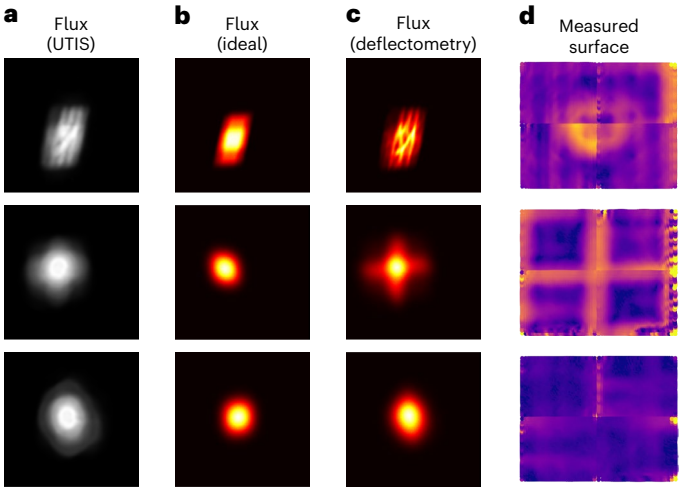

Fig. 5 | Solar flux prediction and heliostat mirror characterization. **a**–**d**, Solar flux prediction is improved by considering the heliostat mirror surface deformations, and PAINT data enable accurate surface reconstructions. An example using ray-tracing with ARTIST[37] shows that the extracted flux image (**a**) is poorly approximated with an ideal heliostat surface (**b**) and improves noticeably when the surface deformations are fitted (**c**). These deformations are highlighted by comparing the ideal surface with the measured surface from the deflectometry data (**d**).

can develop and benchmark data-driven flux prediction models with high precision. The scale of the dataset permits robust validation against extensive unseen test sets. One example of this data-driven approach is the combination of differentiable ray-tracing in ARTIST[37] with high-fidelity surface reconstructions from PAINT deflectometry data to generate accurate flux predictions as shown in Fig. 5.

Heliostat mirror characterization

Correctly characterizing heliostat mirror deformations is crucial for effective CSP tower plant operation. Even minute deformations can have a noticeable impact on the concentrated sunlight on the receiver, particularly during challenging conditions, for example, for low sun elevation when light reflects off the heliostat surface at a steeper angle and imperfections are amplified. These deformations are shown in Fig. 5 (right), which highlights deviations for three different heliostats from an ideal surface, that is, a surface without deformations, using PAINT deflectometry measurements. As deformations vary per heliostat, mirror

https://doi.org/10.1038/s41560-026-02070-1

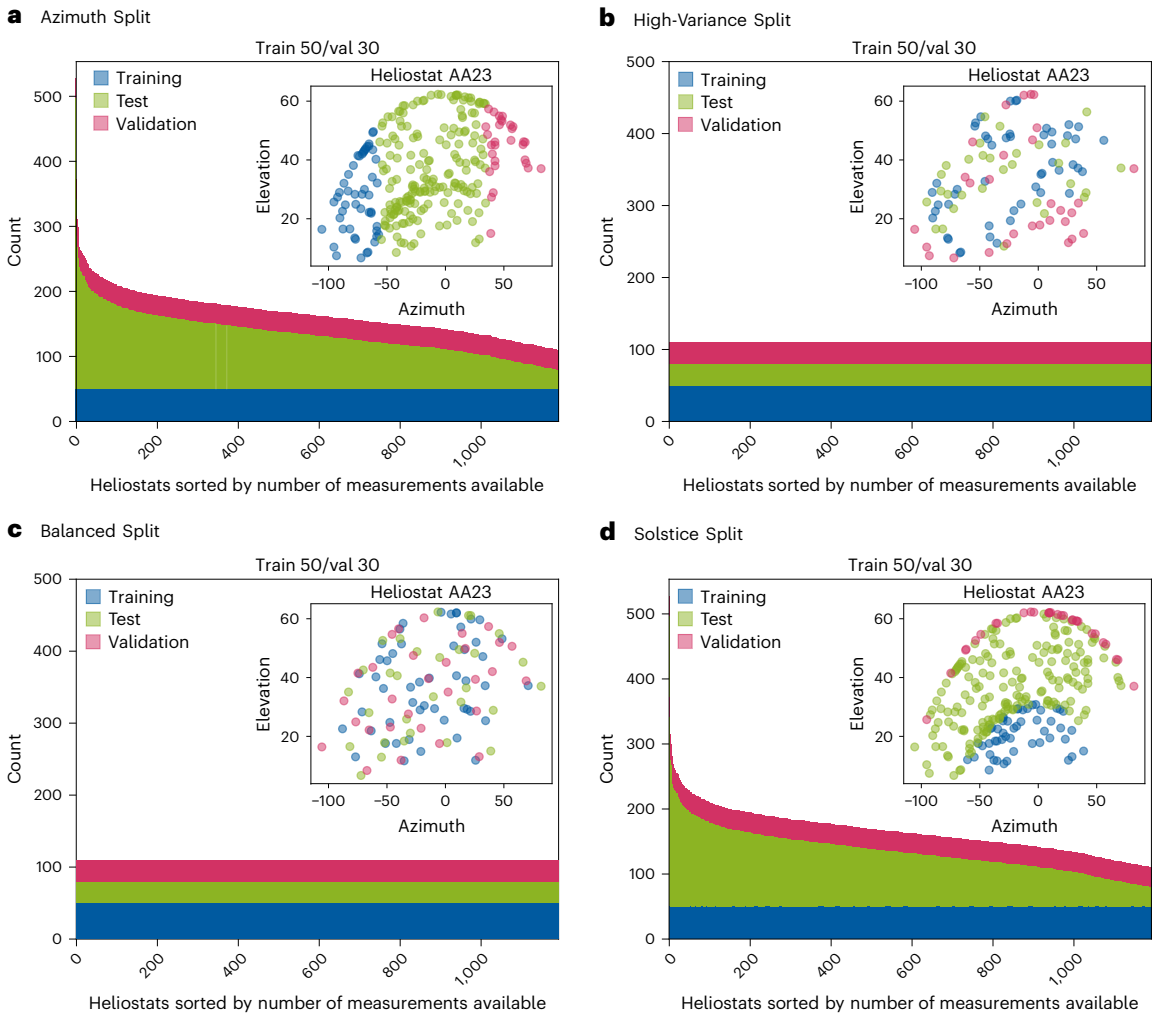

Fig. 6 | The different calibration split methods provided in PAINT. **a**–**d**, Overview of the four different split methods used to create calibration benchmark datasets. The Azimuth Split, based on the azimuth position of the sun (**a**), the High-Variance Split designed to generate vastly different train and validation datasets (**b**), the Balanced Split designed to generate balanced train and validation datasets (**c**) and the Solstice Split based on the ecliptic longitude of the sun (**d**). Insets detail the spatial distribution of training (blue), validation (red) and test (green) samples across azimuth and elevation for a representative individual heliostat (AA23) under each splitting method.

characteristics are usually obtained via expensive deflectometry or photogrammetry measurement campaigns. However, these measurements are only a snapshot of a single point in time and fail to account for later deformations. While recent approaches attempt to characterize heliostat mirror surfaces via deep learning and more widely available flux-density calibration images[13,49], validating these methods without ground-truth data remains difficult. PAINT addresses this gap by providing paired calibration images and ground-truth deflectometry, enabling the rigorous training and validation of data-driven mirror characterization algorithms.

Predictive maintenance and fault detection

Heliostats are continuously exposed to environmental stressors such as thermal cycling, wind loads, dust accumulation and mechanical wear. These factors can lead to gradual deformations in mirror surfaces, misalignments in tracking mechanisms and degradation of optical components. Over time, such imperfections can substantially reduce the efficiency of solar energy concentration, thus increasing operational costs. By analysing historical tracking data, flux measurements and environmental conditions, data-driven models can be trained to identify patterns indicative of impending faults or performance degradation to mitigate these issues. For instance, changes in the distribution of reflected sunlight captured in calibration images can signal mirror surface deformations or tracking errors. PAINT data enable development and testing of such data-driven approaches, reducing unplanned downtime and extending the lifespan of heliostat components.

Open benchmarks for solar tower power plants

One key barrier hindering the adoption of newly developed methods and algorithms in commercial power plants is a lack of trust in research results. These results have consistently overestimated performance[19], are applicable only under specific conditions that are not comparable to commercial operation and have only been validated for a single type of heliostat or field design. PAINT provides a unique opportunity to derive open and standardized benchmarks to enable fair, reproducible comparisons of different methodologies and tasks such as heliostat calibration, flux prediction, control strategies and operational optimization. To demonstrate this potential, we define one possible calibration benchmark based on PAINT in the next section.

Standardized calibration benchmarks from PAINT

Accurate heliostat tracking as a result of effective calibration is crucial for optimal solar tower power plant operation. Current research fails to fairly compare calibration methods owing to the absence of a standardized benchmark and the use of varying heliostat designs and

technologies in each publication. Furthermore, methods are often developed and validated on data from a single measurement campaign and fail to account for seasonal changes due to varying azimuth and ecliptic longitudinal positions of the sun[19].

We address this challenge by providing four different methods for deriving calibration benchmarks from PAINT data. These methods split calibration data into the three data partitions required for data-driven model development: a training set for optimizing model parameters, a validation set for hyperparameter tuning and model selection and a hold-out test set that is completely unseen during the training phase, to provide an unbiased evaluation of the model's final performance. The splits presented here are designed to represent distinct scenarios that are relevant for commercial power plants, and we provide more information on them in the 'Standardized Calibration Benchmarks' section of Supplementary Methods. The Azimuth Split method splits the calibration data on the basis of the azimuth position of the sun (Fig. 6a). The split is designed so that the azimuth positions of the sun in the training dataset are vastly different from those in the validation dataset to encourage the development of calibration methods that can generalize to previously unseen or out-of-distribution conditions.

The High-Variance Split method promotes generalization to unseen or out-of-distribution conditions by leveraging a distance metric that serves as a quality indicator[19]. The split is designed to ensure that the training and validation datasets differ as much as possible by considering both the azimuth and elevation angles of the sun[19] (Fig. 6b). This approach creates a deliberately challenging scenario for learning where the model must generalize across a wide range of sun positions.

The Balanced Split approach uses both the azimuth and elevation of the sun to create evenly distributed training, validation and test datasets (Fig. 6c). This strategy ensures that each data split includes a broad and balanced mix of solar conditions, avoiding overrepresentation of any one region in the sun's path.

The Solstice Split is a seasonal split based on the ecliptic longitude of the sun (Fig. 6d). The split is designed so that the training data comprise calibration data recorded close to the winter solstice, while validation data is recorded close to the summer solstice. This ensures that calibration methods developed with this benchmark are robust to seasonal variations, that is, the ecliptic longitude of the sun.

Conclusions

The PAINT database represents a notable milestone in CSP research, offering a publicly available FAIR dataset of operational data from a solar tower power plant. By providing 849 GB of operational data and accompanying metadata in the STAC format, PAINT enables the development of advanced digital twins, AI-driven calibration and alignment methods, improved solar flux prediction and predictive maintenance tools for research and operators. The data from 2,014 heliostats include detailed heliostat properties, calibration data, deflectometry measurements and weather conditions. The inclusion of standardized benchmarks further fosters reproducibility and fair comparison of research outcomes across institutions and technologies. PAINT democratizes access to high-quality CSP operational data, supports robust algorithm development and accelerates innovation, directly addressing key barriers to the broader adoption of solar tower technology. We envision this release as the catalyst for a paradigm shift in open CSP research. By establishing community standards, PAINT provides a scalable framework for integrating additional data sources and modalities. This future expansion is critical to enhancing the dataset's applicability and transferability across diverse plant designs, serving as a cornerstone for sustained global collaboration.

Methods

The methods applied are divided in three sections, first focusing on the measurement techniques applied to obtain the data before describing the preprocessing and finally how we make our database accessible.

Measurement methods

We collected operational data from the Jülich solar tower power plant between December 2020 and June 2024. The measurements of the solar tower, that is, coordinates of all calibration targets and the receiver, were achieved via laser measurements. The weather data from the Jülich weather station located at the power plant were recorded throughout this period using an array of meteorological sensors. The remainder of the data are heliostat specific and collected on a per-heliostat basis. The heliostat positions were also measured via laser, while all remaining information, that is, the heliostat facet information including facet translation and canting vectors and the heliostat kinematics information, is taken directly from manufacture specifications or computer-aided design (CAD) models of the heliostats. The heliostat calibration and deflectometry data were obtained via measurement campaigns. We explain the setup and methods used to acquire these measurements in detail hereafter.

Heliostat calibration. Heliostat calibration at the Jülich facility is performed using the camera-target method, commonly known as the Stone method[50]. This approach redirects the heliostat's focal spot from its intended receiver onto a Lambertian white calibration target placed near the receiver structure. A camera system captures images of the reflected focal spot on the target, enabling analysis of the heliostat's optical performance. The Jülich facility includes three distinct calibration targets: the Solar Tower Jülich Upper, the Solar Tower Jülich Lower and the Multi Focus Tower target (see Supplementary Fig. 1 for further information). Alongside the captured images, the system records the heliostat's motor positions—describing its kinematics configuration—as well as the solar position at the time of measurement, including azimuth and elevation angles. This combination of image data, actuator states and solar geometry constitutes the raw calibration data. These raw calibration data were obtained via the HeliOS.FDM[47,48] measurement system. The images are subsequently preprocessed, including cropping and identification of the focal spot centroid, as detailed in the 'Preprocessing methods' section below.

Deflectometry measurement. Deflectometry is an optical measurement technique used to characterize the surface shape and quality of heliostats. It operates by projecting a known light pattern, typically a fringe or grid, onto a screen and using a calibrated camera system to record the heliostat reflecting this pattern. Deviations in the reflected pattern are analysed to infer surface slope and curvature, allowing for precise reconstruction of the mirror's surface geometry. This noncontact method is highly sensitive and capable of detecting small deformations such as warping, waviness or other optical defects. In the context of solar tower power plants, deflectometry has become the most widely recommended method for assessing the optical performance of heliostat mirrors[51]. The deflectometry measurements included in the PAINT database were performed with the QDec_2014-101 software[52] from CSP Services GmbH[18]. In addition to the raw measurements, this software provides a summary of the results in PDF format and a processed measurement, where missing values are filled with ideal normal vectors, that is, vectors signifying a surface without deformation. Although the specific deflectometry algorithm is proprietary, the provider operates in full compliance with the SolarPACES Heliostat Measurement Guidelines. This adherence ensures that the resulting surface measurements satisfy the established accuracy requirements of the field.

Preprocessing methods

Before inclusion in the PAINT database, the raw data collected at the Jülich solar tower underwent a series of preprocessing steps. These included organizing unstructured datasets, unifying data formats, performing coordinate transformations and applying image processing techniques to extract relevant features. We describe these steps in more detail below.

Data organization and format standardization. The original data were sourced from various unstructured dumps, distributed across inconsistent directory hierarchies. A critical early step was to reorganize these datasets into a standardized structure aligned with the PAINT database schema. This involved grouping calibration, deflectometry and heliostat properties data via heliostats. For example, calibration images were linked to specific heliostats on the basis of a mapping provided in Excel spreadsheets; the images were then relocated into folders organized by heliostat ID. Additional preprocessed data are made available by cropping and focal spot centroid extraction (see below for further details).

Deflectometry measurements were originally stored in a proprietary binary format. These were converted to HDF5 to ensure consistency across the database. The heliostat ID, extracted during conversion, was also used to place the data into the appropriate directory structure, mirroring the organization scheme used for calibration images.

Heliostat property data were compiled from multiple heterogeneous sources, including CSV and Excel files detailing locations, facet configurations and kinematics parameters. These were merged into unified JSON-based property files, each stored according to heliostat ID.

Weather data from the Jülich tower were initially provided as raw text files. These were parsed and converted into HDF5 format. To facilitate efficient access and download, the data were partitioned into separate HDF5 files, each representing a single month. The data from the DWD weather station were accessed via the Wetterdienst Python package[53] and saved as an additional HDF5 file.

Coordinate conversion. An essential component of the preprocessing pipeline involved transforming spatial coordinates from the Gauss–Krüger (GK) coordinate system to the globally recognized WGS84 reference system. Much of the original data from the Jülich solar tower were recorded in GK zone 2 (EPSG:31466), a projected coordinate system commonly used in Germany for engineering and topographic applications. Unlike WGS84, which represents geographic positions using latitude, longitude and optionally elevation, the GK system is based on the Bessel ellipsoid and uses a transverse Mercator projection, expressing positions in metric units (easting and northing) within defined 3-degree longitudinal zones.

While the GK system offers high local precision, it lacks interoperability with global geospatial standards such as those used in the STAC specification. Therefore, consistent and accurate coordinate transformation was necessary across all spatial metadata. This transformation was implemented using the pyproj library, which performs both projection and datum conversions. A `Transformer` object was instantiated with EPSG:31466 as the source and EPSG:4326 (WGS84) as the target. This transformation was applied uniformly across all metadata referencing spatial locations—including heliostat positions, calibration targets and tower coordinates—to ensure geospatial consistency throughout the dataset.

Image processing and feature extraction. To facilitate the use of calibration data and support algorithm development without requiring raw image processing, we also included preprocessed versions of the calibration images, along with extracted features. Image cropping was performed using a template-matching algorithm based on known markers on the calibration targets. Once cropped, further processing was conducted using a pretrained deep learning model, UNet-Based Target Image Segmentation (UTIS), available in ref. 54. UTIS was used to extract the focal spot by learning spatial features from the images. The resulting preprocessed images—converted to greyscale and centred on the focal spot—as well as the extracted focal spot centroids were stored in the database. A detailed description of the UTIS model and its training methodology is provided in ref. 46. Alternatively, we also include the focal spot measured via the HeliOS system during the calibration process.

STAC file generation. In addition to the organizational and analytical steps described above, another central task was the generation of STAC-compliant metadata files. This process was conducted in parallel with data conversion and feature extraction. Relevant metadata were aggregated from various sources and structured into the STAC specification using the standard hierarchy of catalogues, collections and items. The generated STAC files are available as part of the PAINT package, and the schema used for generating these STAC files is available via GitHub. Extracting these metadata involved combining multiple sources (for example, multiple CSV files or importing data from Excel workbooks), and conversions between coordinate systems were applied to conform to the STAC specification and unify results. In addition, certain values (for example, sun azimuth and elevation) were derived.

Database access

To enable access to the data, we implement the PAINT website, which enables users to browse the data hosted on the LSDF[38]. We also implement a customized STAC client based on PySTAC[35] to enable access to both the data and metadata via Python code. In addition, we implement functionality to create benchmark datasets and load these as PyTorch datasets. All this functionality is available as a part of the PAINT software package via PyPI or GitHub.

Data availability

All data presented in this article are available under the Community Data License Agreement Permissive Version 2.0 (CDLA 2.0), via the PAINT database https://paint-database.org and the PID 21.11152/474a4b1c-de93-4d4a-b33d-1d32d63baf4b. To replicate the illustrative examples, the exact data used can be downloaded with the scripts provided via the ARTIST Github at https://github.com/ARTIST-Association/ARTIST/examples/paint_plots.

Code availability

All code used to generate the data, plots and results in this article is available under a Massachusetts Institute of Technology (MIT) license. The code for preprocessing the data and generating the plots is available via the PAINT GitHub at https://github.com/ARTIST-Association/PAINT. The code for the illustrative examples included in the example uses of the PAINT database section is available via the ARTIST GitHub at https://github.com/ARTIST-Association/ARTIST/examples/paint_plots.

References

1. United Nations Framework Convention on Climate Change (UNFCCC). *The Paris Agreement* (United Nations, 2015); https://unfccc.int/sites/default/files/english_paris_agreement.pdf
2. Conference of the Parties serving as the meeting of the Parties to the Paris Agreement (CMA). *Glasgow Climate Pact* (United Nations Framework Convention on Climate Change, 2021); https://unfccc.int/documents/310497
3. Conference of the Parties serving as the meeting of the Parties to the Paris Agreement (CMA). *Sharm el-Sheikh Implementation Plan* (United Nations Framework Convention on Climate Change, 2022); https://unfccc.int/sites/default/files/resource/cop27_auv_2_cover%20decision.pdf
4. Edenhofer, O. et al. *Renewable Energy Sources and Climate Change Mitigation: Special Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, 2011).
5. Wang, W.-Q., He, Y.-L. & Jiang, R. A multi-scale solar receiver with peak receiver efficiency over 90% at 720 °C for the next-generation solar power tower. *Renew. Energy* **200**, 714–723 (2022).
6. Zhang, H. L., Baeyens, J., Degrève, J. & Cáceres, G. Concentrated solar power plants: review and design methodology. *Renew. Sustain. Energy Rev.* **22**, 466–481 (2013).

7. Alexopoulos, S. & Hoffschmidt, B. Advances in solar tower technology. *WIREs Energy Environ.* **6**, e217 (2017).
8. Barlev, D., Vidu, R. & Stroeve, P. Innovation in concentrated solar power. *Sol. Energy Mater. Sol. Cells* **95**, 2703–2725 (2011).
9. Sorgulu, F. & Dincer, I. Design and analysis of a solar tower power plant integrated with thermal energy storage system for cogeneration. *Int. J. Energy Res.* **43**, 6151–6160 (2019).
10. *Renewable Power Generation Costs in 2022* (International Renewable Energy Agency, 2023); https://www.irena.org/-/media/Files/IRENA/Agency/Publication/2023/Aug/IRENA_Renewable_power_generation_costs_in_2022.pdf
11. Spyridonidou, S. & Vagiona, D. G. A systematic review of site-selection procedures of PV and CSP technologies. *Energy Rep.* **9**, 2947–2979 (2023).
12. Sattler, J. C. et al. Review of heliostat calibration and tracking control methods. *Sol. Energy* **207**, 110–132 (2020).
13. Pargmann, M. et al. Automatic heliostat learning for in situ concentrating solar power plant metrology with differentiable ray tracing. *Nat. Commun.* **15**, 6997 (2024).
14. Oberkirsch, L., Zanger, D. & Hoffschmidt, B. Validation of a closed-loop aim point management system at the Jülich solar tower. *Sol. Energy* **264**, 111951 (2023).
15. Salomé, A., Chhel, F., Flamant, G., Ferrière, A. & Thiery, F. Control of the flux distribution on a solar tower receiver using an optimized aiming point strategy: application to THEMIS solar tower. *Sol. Energy* **94**, 352–366 (2013).
16. Carballo, J. et al. Reinforcement learning for heliostat aiming: improving the performance of solar tower plants. *Appl. Energy* **377**, 124574 (2025).
17. Alcántara, A., Diaz-Cachinero, P., Sánchez-González, A. & Ruiz, C. Leveraging neural networks to optimize heliostat field aiming strategies in concentrating solar power tower plants. *Energy AI* **21**, 100520 (2024).
18. Ulmer, S., März, T., Prahl, C., Reinalter, W. & Belhomme, B. Automated high resolution measurement of heliostat slope errors. *Sol. Energy* **85**, 681–687 (2011).
19. Pargmann, M., Leibauer, M., Nettelroth, V., Quinto, D. M. & Pitz-Paal, R. Questioning the reliability of open-loop calibration methods: introducing a robust data sampling for year-round high accuracy. *Sol. Energy* **286**, 113094 (2025).
20. Offergeld, M., Röger, M., Stadler, H., Gorzalka, P. & Hoffschmidt, B. Flux density measurement for industrial-scale solar power towers using the reflection off the absorber. *AIP Conf. Proc.* **2126**, 110002 (2019).
21. Zhu, G. et al. *Roadmap to Advance Heliostat Technologies for Concentrating Solar-Thermal Power* (National Renewable Energy Laboratory, 2022).
22. Zhu, G. et al. HelioCon: a roadmap for advanced heliostat technologies for concentrating solar power. *Sol. Energy* **264**, 111917 (2023).
23. Grieves, M. & Vickers, J. in *Transdisciplinary Perspectives on Complex Systems: New Findings and Approaches* (eds Kahlen, J. et al.) 85–113 (Springer, 2017).
24. Milidonis, K. et al. Review of application of AI techniques to solar tower systems. *Sol. Energy* **224**, 500–515 (2021).
25. Sievers, L. T. E., Pargmann, M., Quinto, D. M. & Hoffschmidt, B. End-to-end sensitivity analysis of a hybrid heliostat calibration process involving artificial neural networks. *Sol. Energy* **287**, 113219 (2025).
26. Pargmann, M., Ebert, J., Kesselheim, S., Maldonado Quinto, D. & Pitz-Paal, R. In situ enhancement of heliostat calibration using differentiable ray tracing and artificial intelligence. In *SolarPACES Conference Proceedings* Vol. 1 (2022).
27. Kuhl, M. et al. Flux density distribution forecasting in concentrated solar tower plants: a data-driven approach. *Sol. Energy* **282**, 112894 (2024).
28. Xu, F., Wang, J., Guo, M. & Wang, Z. Prediction of solar concentration flux distribution for a heliostat based on lunar concentration image and generative adversarial networks. *Appl. Artif. Intell.* **38**, 2332114 (2024).
29. Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).
30. OpenCSP: an environment for collaborative CSP optical technology development. *OpenCSP Team* https://opencsp.sandia.gov (2025).
31. Brost, R. et al. OpenCSP: collaborative code and data for CSP. *US Department of Energy* https://www.osti.gov/biblio/2999387 (2024).
32. SpatioTemporal Asset Catalog (STAC) specification. *STAC* https://stacspec.org/en/about/stac-spec/ (2021).
33. Spatiotemporal Asset Catalog (STAC) API specification. *GitHub* https://github.com/radiantearth/stac-api-spec (2022).
34. Blumenröhr, N., Ost, P.-J., Kraus, F. & Streit, A. FAIR Digital Objects for the realization of globally aligned data spaces. In *IEEE International Conference on Big Data (BigData)* 374–383 (IEEE, 2025).
35. PySTAC Contributors. PySTAC: Python library for working with SpatioTemporal Asset Catalogs (STAC). *GitHub* https://github.com/stac-utils/pystac (2024).
36. The HDF Group. Hierarchical Data Format, version 5. *GitHub* https://github.com/HDFGroup/hdf5 (2026).
37. ARTIST Association. ARTIST—AI-enhanced differentiable ray tracer for irradiation prediction in solar tower digital twins. *GitHub* https://github.com/ARTIST-Association/ARTIST (2025).
38. Large Scale Data Facility (LSDF). *Karlsruhe Institute of Technology* https://www.scc.kit.edu/en/research/11843.php (2025).
39. Zhu, R. et al. Heliostat field aiming strategy optimization with post-installation calibration. *Appl. Therm. Eng* **202**, 117720 (2022).
40. Grieves, M. *Digital Twin: Manufacturing Excellence Through Virtual Factory Replication* White Paper 1.2014: 1–7 (2014); https://www.3ds.com/fileadmin/PRODUCTS-SERVICES/DELMIA/PDF/Whitepaper/DELMIA-APRISO-Digital-Twin-Whitepaper.pdf
41. Grieves, M. in *The Digital Twin* (eds Crespi, N. et al.) 97–121 (Springer, 2023).
42. Oberkirsch, L., Zanger, D. A. V., Quinto, D. M., Schwarzbözl, P. & Hoffschmidt, B. Static optimal control: real-time optimization within closed-loop aim point control for solar power towers. *Sol. Energy* **255**, 327–338 (2023).
43. Wang, Y., Wu, Z. & Ni, D. Real-time optimization of heliostat field aiming strategy via an improved swarm intelligence algorithm. *Appl. Sci.* **14**, 416 (2024).
44. Pargmann, M., Leibauer, M., Nettelroth, V., Quinto, D. M. & Pitz-Paal, R. Enhancing heliostat calibration on low data by fusing robotic rigid body kinematics with neural networks. *Sol. Energy* **264**, 111962 (2023).
45. Berenguel, M. et al. An artificial vision-based control system for automatic heliostat positioning offset correction in a central receiver solar power plant. *Sol. Energy* **76**, 563–575 (2004).
46. Kuhl, M. et al. In-situ UNet-based heliostat beam characterization method for precise flux calculation using the camera-target method. *Sol. Energy* **279**, 112811 (2024).
47. Geiger, M., Gross, F. & Buck, R. HeliOS control system virtually operates a 100 MW molten salt tower. *AIP Conf. Proc.* **2033**, 210006 (2018).
48. Gross, F., Geiger, M. & Buck, R. A universal heliostat control system. *AIPConf. Proc.* **1850**, 030022 (2017).
49. Lewen, J. et al. Inverse deep learning raytracing for heliostat surface prediction. *Sol. Energy* **289**, 113312 (2025).
50. Stone, K. W. Automatic heliostat track alignment method. US patent US4564275A (1986).
51. März, T., Prahl, C., Ulmer, S., Wilbert, S. & Weber, C. Validation of two optical measurement methods for the qualification of the shape accuracy of mirror panels for concentrating solar systems. *J. Sol. Energy Eng* **133**, 031022 (2011).

52. Quality control in manufacturing. Raise your standards!. *CSP Services* https://www.cspservices.de/quality-control (2026).
53. Gutzmann, B. & Motl, A. Wetterdienst: open weather data for humans. *GitHub* https://github.com/earthobservations/wetterdienst (2024).
54. Kuhl, M. & Pargmann, M. UTIS-HeliostatBeamCharacterization. *GitHub* https://github.com/DLR-SF/UTIS-HeliostatBeamCharacterization/tree/main (2024).

Acknowledgements

This work was supported by the Helmholtz Association Initiative and Networking Fund through the Helmholtz AI platform (to K.P., M.G. and M.W.), HAICORE@KIT and the ARTIST project under grant no. ZT-I-PF-5-159 (to K.P., M.W., M.B., D.M.Q., M.G. and M.P.). This work was performed with the help of the Large Scale Data Facility at the Karlsruhe Institute of Technology funded by the Ministry of Science, Research and the Arts Baden-Württemberg and by the Federal Ministry of Education and Research. We also thank R. Stotzka for tirelessly pushing for open data.

Author contributions

M.P. and M.G. acquired funding for the project. M.P., D.M.Q., C.D., A.S., R.P.-P. and M.G. contributed to the conceptualization and design of the study. K.P. managed and coordinated the project, developed the core codebase and wrote the original draft of the manuscript. M.K., M.W., M.B., J.L. and M.G. contributed to software and code development, with M.G. additionally assisting with web design. O.K. and F.G. provided primary data resources, with F.G. conducting data preprocessing. N.B. implemented FAIR data standards and managed hosting infrastructure, while C.D. provided additional infrastructural support. M.W. made substantial contributions to manuscript editing and structural refinement. All authors except J.L. and O.K. contributed to reviewing and editing the final manuscript.

Funding

Open access funding provided by Karlsruher Institut für Technologie (KIT).

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41560-026-02070-1.

Correspondence and requests for materials should be addressed to Kaleb Phipps.

Peer review information *Nature Energy* thanks Javier Bonilla, Devon Kesseli, Guangdong Zhu and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.