PAINT: The First FAIR Database for Concentrating Solar Power Plants

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]

[1]Scientific Computing Center, Karlsruhe Institute of Technology (KIT), Hermann-von-Helmholtz-Platz 1, Eggenstein-Leopoldshafen, 76344, Baden-Württemberg, Germany.
[2]Helmholtz AI, Germany.
[3]Institute of Solar Research, German Aerospace Center (DLR), Linder Höhe, Cologne, 51147, North Rhine-Westphalia, Germany.
[4]Helmholtz Metadata Collaboration, Germany.
[5]Chair of Solar Technology, RWTH Aachen University, Templergraben 55, Aachen, 52062, North Rhine-Westphalia, Germany.

*Corresponding author(s). E-mail(s): kaleb.phipps@kit.edu;
Contributing authors: mathias.kuhl@dlr.de; marie.weiel@kit.edu; marlene.busch@dlr.de; jan.lewen@dlr.de; nicolas.blumenroehr@kit.edu; daniel.maldonadoquinto@dlr.de; charlotte.debus@kit.edu; felix.goehring@dlr.de; oliver.kaufhold@dlr.de; achim.streit@kit.edu; robert.pitz-paal@dlr.de; markus.goetz@kit.edu; max.pargmann@dlr.de;

Keywords: Concentrating solar power plant, Operational data, FAIR data

Concentrating solar power tower plants hold significant potential for sustainable energy generation but face challenges related to high operational costs, limited data availability, and slow adoption of novel technologies. To address these barriers, we introduce PAINT, the first FAIR open-access database for operational solar tower plant data. PAINT provides 849 GB of high-resolution operational data collected over multiple years from the Jülich solar tower plant, including heliostat properties, calibration and deflectometry measurements, and fine-grained weather data. Organized using the SpatioTemporal Asset Catalog metadata specification, PAINT supports the development of digital twins, AI-based calibration methods, predictive maintenance and fault detection, and improved solar flux prediction. It also introduces standardized benchmarks to promote reproducibility and fair comparisons. By democratizing access to high-quality concentrating solar power data, PAINT enables broader participation in solar research, accelerates innovation, and lays a foundation for data-driven improvements in the performance and reliability of solar tower power plants.

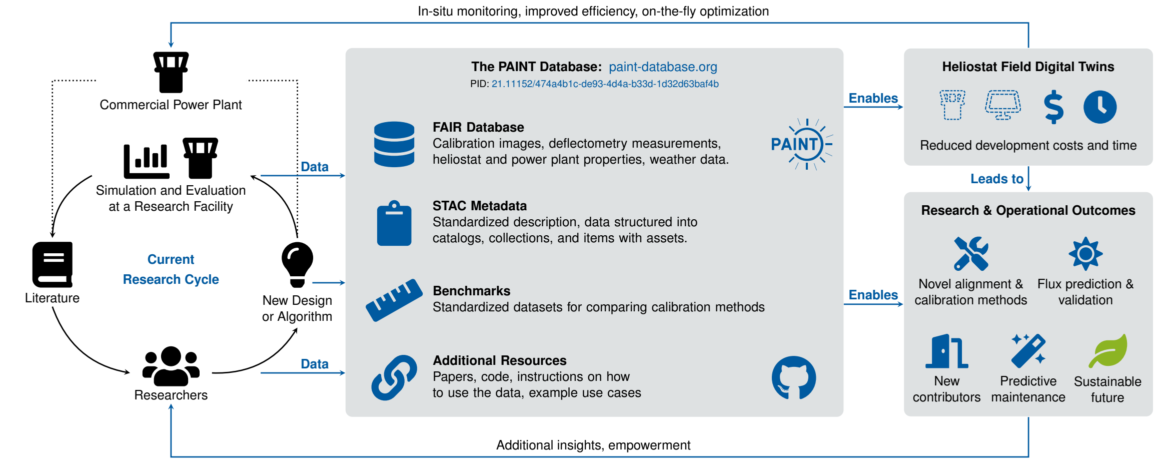


Fig. 1: **Advancing the research cycle for solar tower power plants.** An overview of the current research cycle for solar tower power plants and how the PAINT Database expands it by providing FAIR data, comprehensive metadata, standardized benchmarks, papers, and code. Black lines refer to the current research cycle, whilst blue lines highlight the inclusion of PAINT and insights gained from the data.

Introduction

Concentrating solar power (CSP) plants are one of the most promising options for renewable energy generation, specifically in the form of solar tower power plants [1]. These power plants use an array of up to ten thousand mirrors, known as heliostats, to focus sunlight onto a receiver and heat a medium to temperatures over 700 °C [2–4] with the resulting thermal energy converted into electricity, used directly in industrial processes, or stored in cheap thermal storage system [4–6]. However, there are still challenges limiting their widespread deployment, such as high initial deployment costs [7], limited geographical locations [8], heliostat alignment [9], heliostat imperfections [10], and the real-time coordination and optimization of heliostat aim points [11–14].

In the current research cycle, these challenges are addressed by researchers developing new designs or algorithms based on existing literature, simulating or testing these concepts at a research facility, and publishing their results to update the current literature (Figure 1). While some solutions reach commercial power plants, the process is hindered by operators' reluctance to adopt unproven technologies, largely due to the high operational and economic risk in adapting existing systems and procedures [9, 15]. 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 or these of other power plants [15–17].

Open data has already been identified in the roadmap for advanced heliostat technologies [18, 19] as a bottleneck hindering the realization of the true potential of CSP technology. Open data is particularly important for modern tools such as digital twins [20] and AI-based algorithms [21], which can revolutionize CSP research. AI-based methods have already been shown to outperform traditional methods in tasks such as heliostat calibration [10, 22, 23] and flux-density prediction [24, 25] but remain unfeasible without sufficient high-quality and well-documented data for training. Additionally, this data must be easily findable, accessible to researchers, interoperable across tools and platforms and reusable (FAIR) [26]. Despite its critical importance, operational data from solar tower plants remains largely inaccessible. Many factors, including commercial confidentiality, intellectual property concerns, and the absence of clear incentives or frameworks for data sharing, drive this scarcity. The limited open CSP data that is available, for example as a part of the OpenCSP project [27], is unstructured and fails to provide a complete overview of power plant operation, lacking the granularity to for data-driven solutions, such as digital twins.

Therefore, we introduce PAINT, the first FAIR database for operational CSP plant data in the world, providing 849 GB of operational data from a solar tower power plant in Jülich, Germany. The data from over 2,000 heliostats is supplemented with local weather data, extensive metadata in the SpatioTemporal Asset Catalog (STAC) specification, and software via a Python package. The PAINT database supports the development of heliostat field digital twins, which serve both as platforms for testing novel algorithms and as tools for in-situ monitoring of existing solar tower power plants

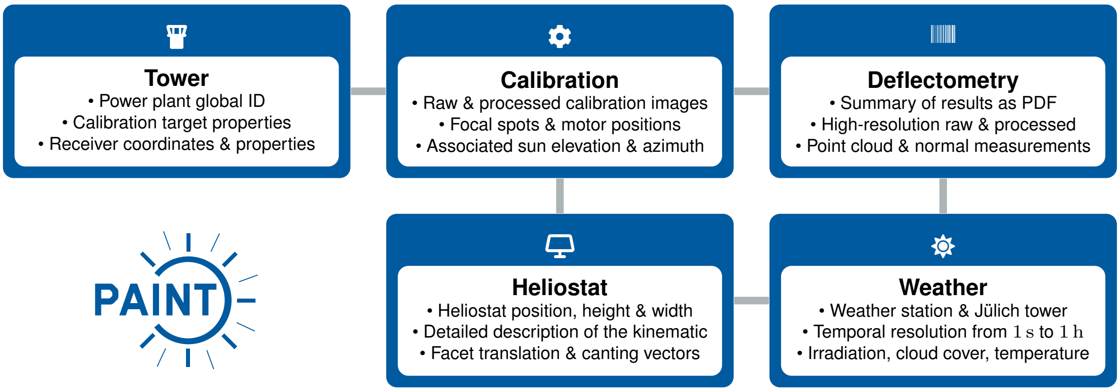


Fig. 2: **The data categories in the PAINT database.** Overview of the main categories of data provided via the database and a short description of the concrete data available.

(Figure 1). Integrated with a digital twin, PAINT enables all types of Beam Characterization Systems (BCS) advancements in heliostat calibration and alignment, surface reconstructions, improved flux-density prediction, and data-driven predictive maintenance e.g., soiling detection – each with the potential to reduce operating costs significantly. For the first time, standardized benchmarks will allow for 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 toward 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 (Figure 2). The tower measurements data contains information on the solar tower, including the global ID of the power plant, the properties of the various targets used for calibration measurements, and the properties and coordinates of the receiver. Weather data is collected at a one-second resolution from a weather tower at the Jülich tower as well as at a one-hour 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. This weather data includes 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 contains both raw and pre-processed solar flux images of the calibration target, taken during calibration measurements. Additionally, 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 comprises 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 pre-processed form, where missing vectors are filled with assumed ideal vectors. Heliostat properties data contains key information about each individual heliostat, including its location, geometric description, overview of the kinematic, and facet properties. Whilst properties data are available for all heliostats, calibration data is only available for 1,893 heliostats and deflectometry data for 471. The number of available calibration measurements varies per heliostat and PAINT includes calibration and deflectometry data from a range of heliostats throughout the field (Figure 3).

The collected data is combined and pre-processed to a consistent format before being included in the database[1]. Additionally, all metadata is 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 [28]. STAC organizes the data into a structure of *items*, *collections*, and *catalogs*, enabling users to efficiently browse and query the PAINT database by

[1] A detailed overview of the pre-processing applied is available in the supplementary material.

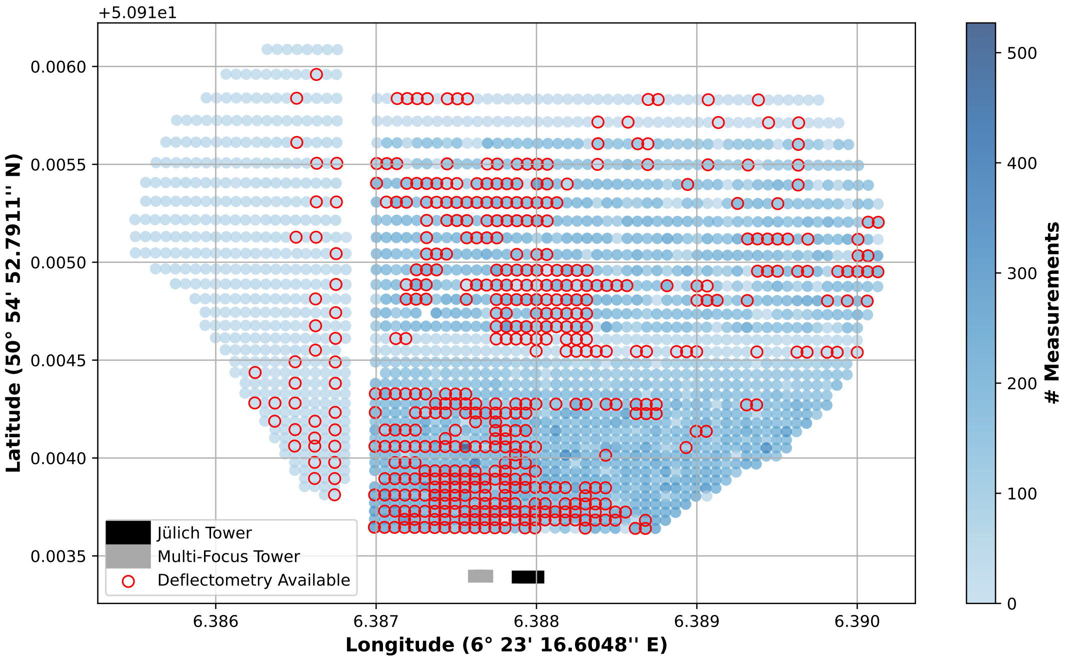


Fig. 3: **The heliostat positions and available data.** Overview of the latitude and longitude of all heliostats available in the PAINT database, colored based on the number of calibration measurements available for that heliostat. Heliostats marked by a red outline also include deflectometry measurements.

time and location, either with existing STAC supported APIs [29] or more efficiently with our own software package. In PAINT each heliostat is accompanied by a STAC *catalog*, whilst the properties, calibration, and deflectometry metadata for each heliostat are *collections* within this catalog. The weather data is a separate *collection* within the main power plant *catalog*, which also contains the tower properties *item*. A detailed overview of the STAC structure is provided in the supplementary material.

All resources are available online via the PAINT database (https://paint-database.org), which will be updated and maintained for the foreseeable future. The accompanying Python package can be installed via PyPi (https://pypi.org/project/paint-csp/) and allows easy access to the data and associated metadata. All code for pre-processing and preparing the database is documented via GitHub (https://github.com/ARTIST-Association/PAINT). These practices help ensure that PAINT adheres to the FAIR data principles [26].

Adherence to FAIR Data Principles

We strive to ensure that PAINT is, findable, accessible, interoperable, and reusable according to the FAIR principles [26]. This is achieved by inherent design and data curation decisions and by modeling the STAC metadata assets as FAIR Digital Objects for advanced machine-actionable decision making [30]. Findability is achieved with the previously mentioned website as well as a unique Persistent Identifier (PID) that resolves to the main, navigable STAC catalog (https://hdl.handle.net/21.11152/474a4b1c-de93-4d4a-b33d-1d32d63baf4b). In addition, the STAC specification enables the user to apply metadata-based filters on the data to easily find specific data, e.g. only for a certain time period.

As an open-source specification that is clearly documented via GitHub [28] and actively developed by the community, the STAC specification also ensures accessibility. STAC employs the JSON-format, which can be browsed in all common program languages and multiple tools exist, e.g. the Python PySTAC[2] package, 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

[2] https://github.com/stac-utils/pystac.

knowledge and the accompanying PID ensures the main STAC catalog 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) [31], and Portable Document Format (PDF). This data is clearly described through the STAC metadata, as well as on the website and in the supplementary of 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[3] – 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. Additionally the data is licensed under the Community Data License Agreement - Permissive - Version 2.0 (CDLA 2.0), whilst the associated software is available via an 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.

Example Uses of the PAINT Database

The key barriers hindering widespread commercial CSP deployment are the high initial deployment costs, unreliable performance, and operating and maintenance costs [19, 32]. In the following, we highlight various uses for the data that can contribute to solving these challenges.

Heliostat Field Digital Twins

A digital twin is a complete virtual description of a physical object facilitating bidirectional data exchange [33, 34]. They can be used in real-time to optimize CSP plant operation via in-situ monitoring to optimize performance [10] or in a research setting to evaluate novel solutions, and investigate critical scenarios without endangering the integrity of the plant or the researchers involved [13, 35, 36]. Furthermore, accurate heliostat field digital twins can also be used to simulate the performance of novel designs for receivers. Digital twin creation depends on extensive and accurate data. PAINT provides detailed information on heliostat locations, heliostat mirror imperfections, and kinematic design and parameters. The associated high-resolution weather data enables realistic simulations of challenging scenarios, such as fluctuating irradiance conditions. This data facilitates heliostat field digital twin creation, with the advantages already shown via ARTIST (https://github.com/ARTIST-Association/ARTIST), improving surface characterization and flux density prediction [10] and heliostat calibration [37]. These advantages are highlighted in the following uses.

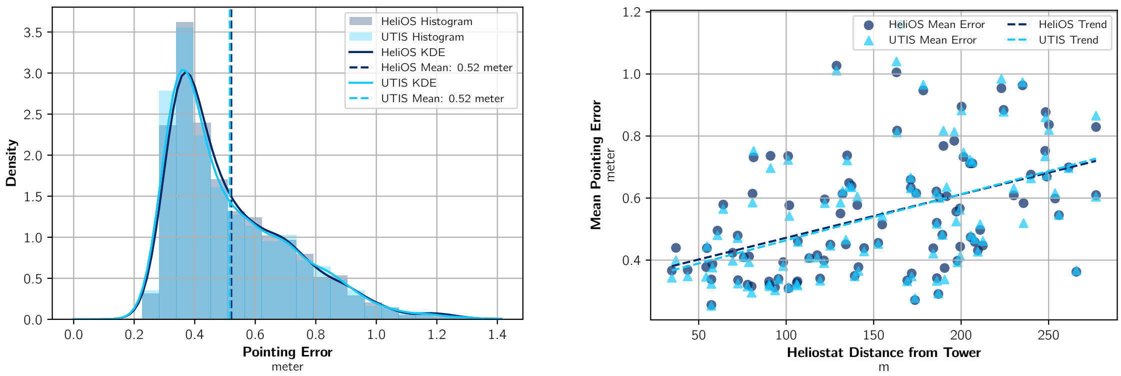
Focal Spot Centroid Detection

A crucial task to facilitate alignment and calibration is focal spot centroid detection in flux density target images [38]. Raw target images captured by target cameras often contain skewed orientations, variable lighting conditions, and off-center focal spots. Without detecting the focal spot centroid, there is no consistent frame of reference and any calibration or alignment algorithm based on this data will be unreliable. The large high-quality calibration image dataset provided via PAINT, enables AI-based image processing methods that can significantly improve focal spot centroid detection (Figure 4 (a)) [39].

Heliostat Calibration & Alignment

Due 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 kinematic. Furthermore, due mechanical wear, structural bending, or changing environmental conditions, the heliostats' conditions also gradually degrade over time. 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. PAINT provides over 218 thousand calibration images and associated metadata, enabling the robust evaluation of calibration methods,

[3] https://github.com/ARTIST-Association/ARTIST.



(a) **Error Distribution**

(b) **Distance & Error**

Fig. 4: **Focal Spot Detection and Heliostat Calibration & Alignment.** PAINT data enables novel focal spot centroid detection methods which improve calibration performance (a). The diverse calibration dataset enables robust analysis of calibration algorithms, here the calibration error compared to the distance of the heliostat from the tower (b).

e.g. a comprehensive analysis of the heliostat distance from the tower and the associated calibration error (Figure 4 (b)).

Solar Flux Prediction & 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, AI models can learn to predict flux distributions with greater accuracy (Figure 5) [24]. Furthermore, the vast amount of operational data available enables validation of these models against unseen data.

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, e.g. for low sun elevation when light reflects off the heliostat surface at a steeper angle and imperfections are amplified. Mirror 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. Recent approaches attempt to characterize heliostat mirror surfaces via deep learning and more widely available flux density calibration images [10, 40]. PAINT provides both calibration images and ground truth deflectometry measurements to enable the training and validation of data-driven approaches for mirror characterization (Figure 5).

Predictive Maintenance & 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 significantly reduce the efficiency of solar energy concentration thus increasing operational costs. By analyzing historical tracking data, flux measurements, and environmental conditions, AI 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 allows for such data-driven approaches to be developed and tested allowing, reducing unplanned downtime and extending the lifespan of heliostat components.

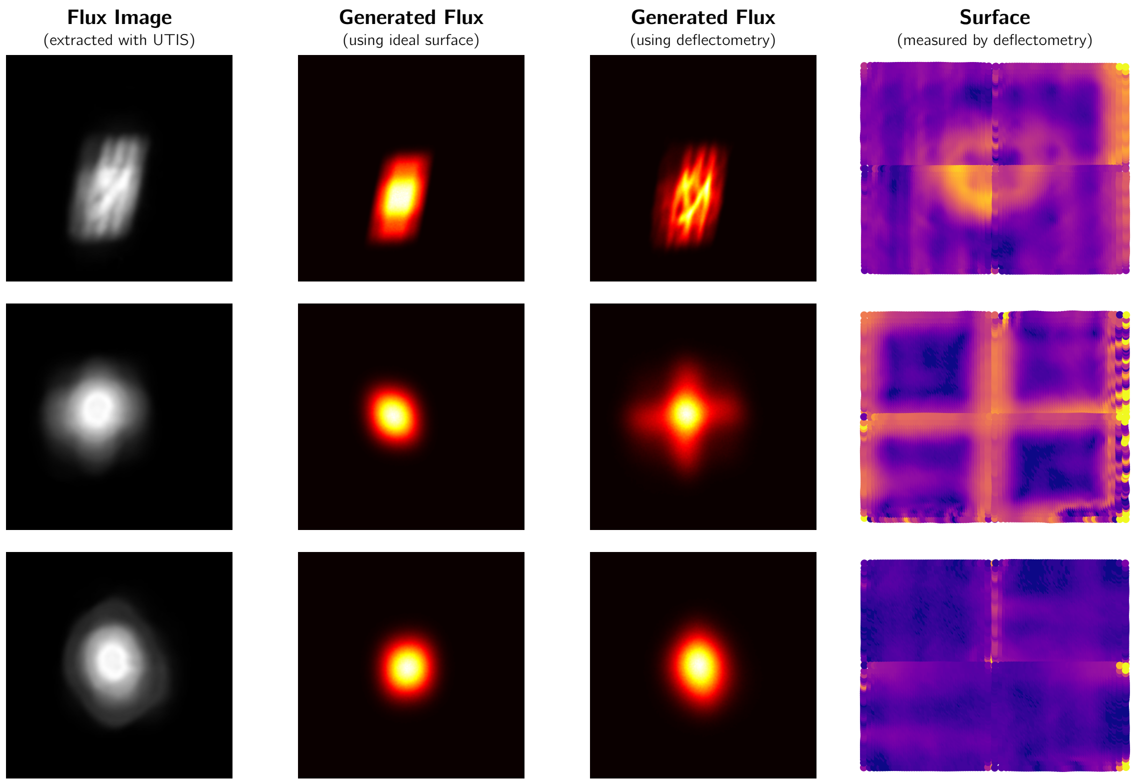


Fig. 5: **Solar Flux Prediction and Heliostat Mirror Characterization.** Solar flux prediction is improved by considering the heliostat mirror surface deformations. The extracted flux image (left), is poorly approximated with an ideal heliostat surface (middle left) and improves noticeably when the surface deformations are fitted (middle right). These deformations are highlighted by comparing the ideal surface with the measured surface from the deflectometry data (right).

Open Benchmarks for Solar Tower Power Plants

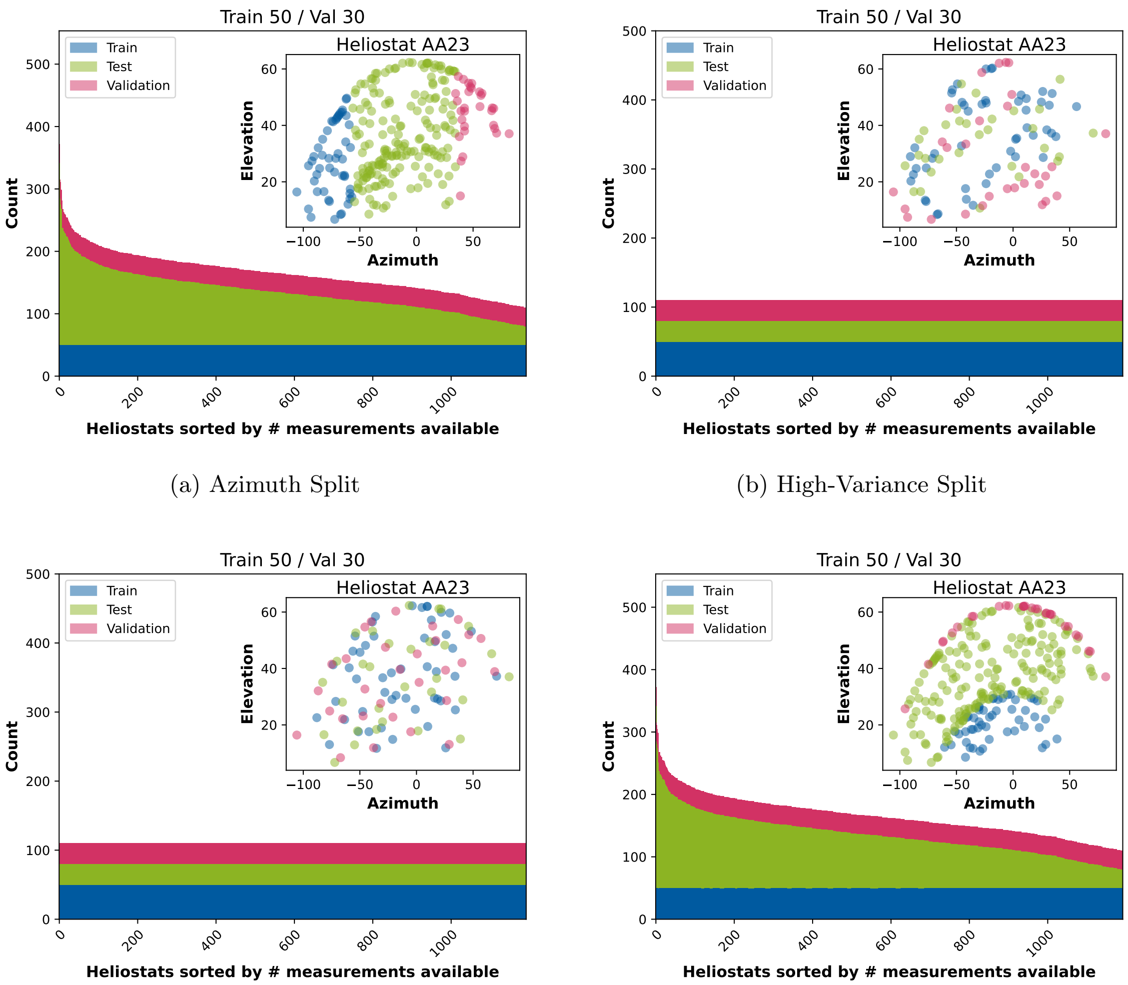
One key barrier hindering the adoption of novel methods and algorithms in commercial power plants is a lack of trust in research results. These results have consistently overestimated performance [16], been only applicable in specific conditions that are not comparable to commercial operation, and only 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 tasks such as heliostat calibration, flux prediction, control strategies, and operational optimization.

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 due 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 seasonal changes due to varying azimuth and elliptical longitudinal positions of the sun [16].

We address this challenge by providing four different methods for deriving calibration benchmarks from PAINT data. These methods split calibration data into training, validation, and test datasets, designed to represent distinct scenarios that are relevant for commercial power plants. The *Azimuth Split* method, splits the calibration data based on the azimuth position of the sun (Figure 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 [16]. The split is designed



(a) Azimuth Split (b) High-Variance Split

(c) Balanced Split (d) Solstice Split

Fig. 6: **The different split methods provided in PAINT.** Overview of the four different split methods used to create calibration benchmark datasets; (a) The Azimuth Split, based on the azimuth position of the sun, (b) The High-Variance Split designed to generate vastly different train and validation datasets, (c) The Balanced Split designed to generate balanced train and validation datasets, and (d) The Solstice Split based on the elliptical longitude of the sun.

to ensure that the training and validation datasets differ as much as possible by considering both the azimuth and elevation angles of the sun (Figure 6b) [16]. 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 (Figure 6c). This strategy ensures that each data split includes a broad and balanced mix of solar conditions, avoiding over-representation of any one region in the sun's path.

The *Solstice Split* is a seasonal split based on the elliptical longitude of the sun (Figure 6d). The split is designed so that the training data is comprised of calibration data recorded close to the winter solstice, whilst validation data is recorded close to the summer solstice. This ensures that calibration methods developed with this benchmark are robust to seasonal variations, i.e. the elliptical longitude of the sun.

Conclusions

The PAINT database represents a significant milestone in CSP research, offering the first 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 2014 heliostats includes 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, helping to overcome key barriers to the broader adoption of solar tower technology, and will be maintained and updated for the foreseeable future, defining community standards for open CSP research data and serving as the foundation for further innovation and collaboration.

Methods

We first collect operational data from the Jülich solar tower power plant between December 2020 and June 2024. The measurements of the solar tower, i.e. coordinates of all calibration targets and receiver, were achieved via laser measurements. The weather data from the Jülich weather station located at the power plant was recorded throughout this period using am array of meteorological sensors. The remainder of the data is heliostat specific and collected on a per-heliostat basis. The heliostat positions were also measured via laser, whilst all remaining information, i.e. the heliostat facet information including facet translation and canting vectors, and the heliostat kinematic information, are taken directly from manufacture specifications or CAD models of the heliostats.

The calibration data was obtained via the Camera-Target Method, often referred to as the Stone method [41]. This involves redirecting the heliostat's focal spot from the receiver to a Lambertian white target positioned close to the receiver. A camera is used to capture the target image, which is then saved, along with associated metadata. Our calibration data was obtained via the HeliOS.FDM measurement system.

Deflectometry measurements are obtained by projecting a known pattern of light onto the surface of a heliostat and analyzing the reflected pattern captured by a camera. By studying variations in intensity and phase shifts of the pattern, deflectometry can accurately map surface deformations. Our deflectometry data was obtained via the QDec_2014-101 measurement system, which in addition to the raw measurements provides a summary of the results in PDF format and a processed measurement, where missing values are filled with ideal normal vectors.

Given the raw data, multiple pre-processing steps were carried out. All data was sorted into the appropriate structure, grouping calibration, deflectometry, and heliostat properties data via heliostats. Furthermore, formats were unified, i.e. the deflectometry and weather data were saved in the HDF5 format and any properties data was converted to JSON format. In the case of the Jülich weather data, we also split data via months to generate multiple HDF5 files. The data from the DWD weather station was accessed via the Wetterdienst Python package (https://github.com/earthobservations/wetterdienst), and saved as an additional HDF5 file. All metadata was extracted from the available raw data and saved according to the STAC specification. This extraction process involved combining multiple sources (e.g. multiple CSV files or importing data from Excel workbooks) and conversions between coordinate systems (e.g. Gauss-Krüger to WGS84) were applied to confirm to the STAC specification and unify results. Additionally, certain values (e.g. sun azimuth and elevation) were derived. For the calibration images, additional pre-processed data is made available by cropping and focal spot centroid extraction using UTIS (https://github.com/DLR-SF/UTIS-HeliostatBeamCharacterization). We provide a more detailed description of the pre-processing and data structure in the supplementary material. Additionally, all code applied to sort the data, unify file formats, extract metadata, and apply conversions is available via GitHub (https://github.com/ARTIST-Association/PAINT) with the associated documentation on Readthedocs (https://paint.readthedocs.io/en/latest/).

To enable access to the data, we implement the PAINT website which enables users to browse the data hosted on the Large Scale Data Facility (LSDF) a the Karlsruhe Institute of Technology. We also implement a customized STAC client based on PySTAC (https://github.com/stac-utils/pystac) to enable access to both the data and metadata via Python code. Additionally, 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 (https://pypi.org/project/paint-csp/) or GitHub (https://github.com/ARTIST-Association/PAINT).

Data availability

All data presented in this article is available under the Community Data License Agreement - Permissive - Version 2.0 (CDLA 2.0), via the PAINT database: http://paint-database.org.

Code availability

All code used to generate the data, plots, and results in this article is available under a MIT license. The code for pre-processing the data and generating the plots is available in the PAINT GitHub: https://github.com/ARTIST-Association/PAINT. The code for the results presented in the example uses of the PAINT database is available in the ARTIST GitHub: https://github.com/ARTIST-Association/ARTIST/examples/paint_plots.

References

[1] Edenhofer, O. *et al.* *Renewable energy sources and climate change mitigation: Special report of the intergovernmental panel on climate change* (Cambridge University Press, 2011).

[2] 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. *Renewable Energy* **200**, 714–723 (2022).

[3] Zhang, H. L., Baeyens, J., Degrève, J. & Cáceres, G. Concentrated solar power plants: Review and design methodology. *Renewable and sustainable energy reviews* **22**, 466–481 (2013).

[4] Alexopoulos, S. & Hoffschmidt, B. Advances in solar tower technology. *Wiley Interdisciplinary Reviews: Energy and Environment* **6**, e217 (2017).

[5] Barlev, D., Vidu, R. & Stroeve, P. Innovation in concentrated solar power. *Solar energy materials and solar cells* **95**, 2703–2725 (2011).

[6] Sorgulu, F. & Dincer, I. Design and analysis of a solar tower power plant integrated with thermal energy storage system for cogeneration. *International Journal of Energy Research* **43**, 6151–6160 (2019).

[7] RENA, I. Renewable power generation costs in 2022 (2023).

[8] Spyridonidou, S. & Vagiona, D. G. A systematic review of site-selection procedures of pv and csp technologies. *Energy Reports* **9**, 2947–2979 (2023).

[9] Sattler, J. C. *et al.* Review of heliostat calibration and tracking control methods. *Solar Energy* **207**, 110–132 (2020).

[10] Pargmann, M. *et al.* Automatic heliostat learning for in situ concentrating solar power plant metrology with differentiable ray tracing. *Nature Communications* **15**, 6997 (2024).

[11] Oberkirsch, L., Zanger, D. & Hoffschmidt, B. Validation of a closed-loop aim point management system at the jülich solar tower. *Solar Energy* **264**, 111951 (2023).

[12] 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. *Solar Energy* **94**, 352–366 (2013).

[13] Carballo, J. *et al.* Reinforcement learning for heliostat aiming: Improving the performance of solar tower plants. *Applied Energy* **377**, 124574 (2025).

[14] 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 and AI* 100520 (2024).

[15] Ulmer, S., März, T., Prahl, C., Reinalter, W. & Belhomme, B. Automated high resolution measurement of heliostat slope errors. *Solar Energy* **85**, 681–687 (2011).

[16] 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. *Solar Energy* **286**, 113094 (2025).

[17] 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 Conference Proceedings* 110002 (2019).

[18] Zhu, G. *et al.* Roadmap to advance heliostat technologies for concentrating solar-thermal power. Tech. Rep., National Renewable Energy Lab.(NREL), Golden, CO (United States) (2022).

[19] Zhu, G. *et al.* Heliocon: A roadmap for advanced heliostat technologies for concentrating solar power. *Solar Energy* **264**, 111917 (2023).

[20] Grieves, M. & Vickers, J. Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems. *Transdisciplinary perspectives on complex systems: New findings and approaches* 85–113 (2017).

[21] Milidonis, K. *et al.* Review of application of ai techniques to solar tower systems. *Solar Energy* **224**, 500–515 (2021).

[22] 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. *Solar Energy* **287**, 113219 (2025).

[23] 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. *SolarPACES Conference Proceedings* (2022).

[24] Kuhl, M. *et al.* Flux density distribution forecasting in concentrated solar tower plants: A data-driven approach. *Solar Energy* **282**, 112894 (2024).

[25] 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. *Applied Artificial Intelligence* **38**, 2332114 (2024).

[26] Wilkinson, M. D. *et al.* The fair guiding principles for scientific data management and stewardship. *Scientific data* **3**, 1–9 (2016).

[27] OpenCSP Team. OpenCSP: An Environment for Collaborative CSP Optical Technology Development. https://opencsp.sandia.gov (2025).

[28] STAC Contributors. Spatiotemporal asset catalog (stac) specification (2021). URL https://stacspec.org. [Online; accessed April 11, 2025].

[29] STAC API Contributors. Spatiotemporal asset catalog (stac) api specification (2022). URL https://stacspec.org. [Online; accessed April 11, 2025].

[30] Blumenröhr, N., Ost, P.-J., Kraus, F. & Streit, A. Fair digital objects for the realization of globally aligned data spaces. *IEEE International Conference on Big Data (BigData)* 374–383 (2025).

[31] The HDF Group. Hierarchical Data Format, version 5. URL https://github.com/HDFGroup/hdf5.

[32] Zhu, R. *et al.* Heliostat field aiming strategy optimization with post-installation calibration. *Applied Thermal Engineering* **202**, 117720 (2022).

[33] Grieves, M. Digital twin: manufacturing excellence through virtual factory replication (2014).

[34] Grieves, M. *Digital twins: past, present, and future* (Springer, 2023).

[35] 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. *Solar Energy* **255**, 327–338 (2023).

[36] Wang, Y., Wu, Z. & Ni, D. Real-time optimization of heliostat field aiming strategy via an improved swarm intelligence algorithm. *Applied Sciences* **14**, 416 (2024).

[37] 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. *Solar Energy* **264**, 111962 (2023).

[38] Berenguel, M. *et al.* An artificial vision-based control system for automatic heliostat positioning offset correction in a central receiver solar power plant. *Solar energy* **76**, 563–575 (2004).

[39] Kuhl, M. *et al.* In-situ unet-based heliostat beam characterization method for precise flux calculation using the camera-target method. *Solar Energy* **279**, 112811 (2024).

[40] Lewen, J. *et al.* Inverse deep learning raytracing for heliostat surface prediction. *Solar Energy* **289**, 113312 (2025).

[41] Stone, K. W. Automatic heliostat track alignment method (1986).

Acknowledgments

We would like to thank Rainer Stotzka for tirelessly pushing for open data.

Author contribution

M.P. and M.G. envisioned and designed the project. M.W., M.P., K.P. mainly worked on code. All authors have participated in preparing the final draft of the paper.

Funding

This work is supported by the Helmholtz Association Initiative and Networking Fund through the Helmholtz AI platform, HAICORE@KIT and the ARTIST project under grant number ZT-I-PF-5-159.

Competing Interests

The authors declare no competing interests.