

Questioning the reliability of open-loop calibration methods: Introducing a robust data sampling for year-round high accuracy

Max Pargmann^{a,1,*}, Moritz Leibauer^{b,1}, Vincent Nettelroth^b, Daniel Maldonado Quinto^a, Robert Pitz-Paal^a

^a Institute of Solar Research, German Aerospace Center (DLR), Linder Höhe, Köln, 51147, Germany

^b Synhelion Germany GmbH, Am Brainery Park 1, Jülich, 52428, Germany

ARTICLE INFO

Keywords:

Concentrating solar power
Solar tower power plant
Heliostat aiming
Data set sampling
Nearest neighbors

ABSTRACT

Heliostat calibration in solar tower plants is critical for optimizing plant efficiencies through precise solar tracking. Current practices often assume heliostat precision degrades over time, leading to the development of new calibration procedures utilizing time-dependent data sets. However, contrary to this prevailing assumption, our study demonstrates the consistency of tracking accuracy over extended periods when appropriate calibration points are selected. We introduce a novel data sampling method that uses sun positions in Euler angles as relevancy scores, enabling higher accuracy with a reduced data requirement. Our thorough analysis challenges the common belief that time significantly impacts calibration accuracy. Furthermore, we unveil an overlooked relationship between prediction accuracy and solar position coverage, raising legitimate concerns about the reliability of reported accuracies in previous publications. To promote transparency, we present clear data and advocate for improved reporting practices in future publications. Applying the new data set sampling to a non optimized data set we archive a year-round stable accuracy below 1.5 mrad with as little as 27 calibration points.

1. Introduction

The efficiency of power plants relies heavily on the ability of heliostats to accurately track the sun. Karellas and Roupedakis [1] estimates the maximum achievable heliostat field efficiency to 70% under the assumption of ideal heliostat tracking but further stresses the trade off between receiver efficiency losses due to aperture size on the one hand and spillage losses on the other hand. The focal spot size of each heliostat can be minimized by its concentrator's design. However, the combined focal spot size of the heliostat field results from the superimposition of all individual focal spots. The minimum spillage can thus be obtained via heliostat aim point optimization which requires minimal heliostat tracking errors. Heliostats must therefore meet strict tracking accuracy requirements, while also being cost-effective and able to withstand various external factors, like wind or even sandstorm [2].

Commonly open-loop heliostat control, for instance, without continuous feedback regarding its current alignment is applied for cost-efficiency and ease-of-implementation. Because of the lack of real-time alignment measurements the open-loop control applies a predictive model to estimate the current alignment from observed actuator stepper

motor positions. Open-loop methods therefore require discrete measurements of the heliostats' behavior and modeling of their characteristics. During field operation the derived model is fit to the measurements by applying machine learning algorithms. The tracking accuracy thus depends on the quality of model training. Heliostat calibration is therefore key to maintaining high tracking accuracies. The most commonly used method is the camera-target method [3] (also known as beam characterization system).

This involves moving a single heliostat from the receiver to a target located close by (compare Fig. 1. Due to individual errors of the heliostat, the intended focal spot position of the heliostat may deviate from its measured position. By analyzing this deviation, along with the heliostat's position in the field and the sun's position, a heliostat kinematic model can be fitted by regression. This model can then be used to adjust the orientation of the heliostat and minimize its sun tracking error. The method is highly automated and stable, although its accuracy typically falls in the range of 1–5 milliradians. For comparison, the sun's divergence is approximately 3 mrad. These high error margins can lead to focal spots missing the target for heliostats located at high distance to the tower. This enforces a conservative operation

* Corresponding author.

E-mail address: max.pargmann@dlr.de (M. Pargmann).

¹ Contributed equally to this work.

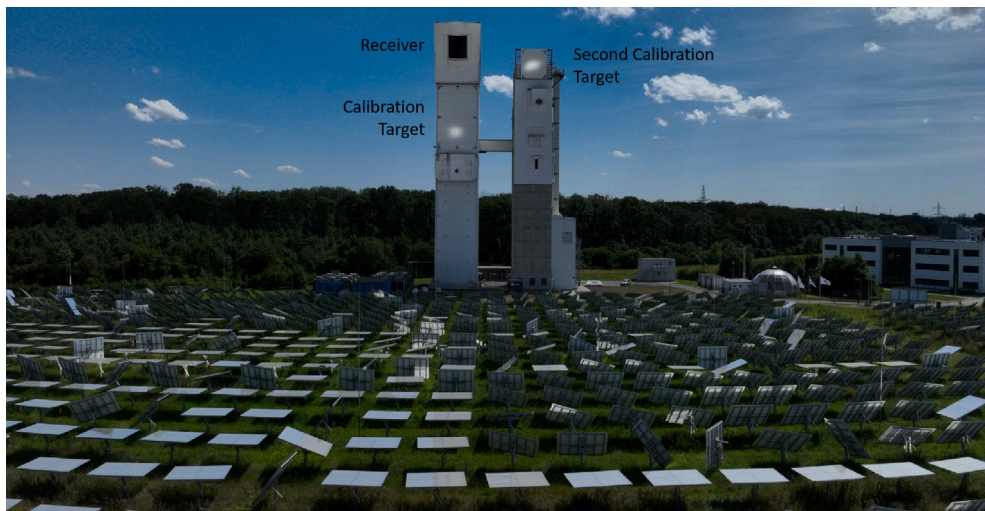


Fig. 1. The solar tower in Jülich shown on the left and multi-focus tower shown at the right. On top of the solar tower is the receiver, where all heliostat focal spots are superimposed. Below is the calibration target, showing a single focal spot. On top of the multi-focus tower is the secondary calibration target, also showing a single focal spot.

and can cost power plant operators millions of dollars per year [4]. Moreover, the heliostats are calibrated regularly since it is assumed that the sun tracking accuracy deteriorate over time. This assumption is partly responsible for the fact that time is included as a major factor in the calibration of the heliostats [5,6]. Furthermore, based on this assumptions old calibration data is often regarded as deprecated and dismissed.

The camera-target method was taken up and improved by a multitude of more advanced approaches, like heliostat alignment via lasers [7] and cameras [8] or on the heliostats at night using stars or moonlight [9], measured directly on the receiver [10] or drone flights [11]. In general, the number of published calibration methods is vast [12], indicating a significant research effort towards developing and validating analytical procedures for accurate measurements and reliable results.

Despite good published focal spot position measuring accuracies below 1 mrad (see [12]), none of these methods were adapted in commercial power plants, so far. The camera-target method is still the most commonly used method for heliostat calibration in commercial power plants.

This can be explained by the risk, which power plant operators are taking, when exchanging the calibration procedure and the well established reliability of measurement accuracies by the camera target methods. However, this comes at the cost of slow measuring speeds per data point. Consequently, data point acquisition holds a high amount of opportunity costs about deciding which heliostat to calibrate next. However, until now, little analysis into the applied data points' relevancies was pursued.

In this paper, we investigate the influence of different data splitting techniques on tracking accuracies. In order to investigate their impact on resulting calibration accuracies, a large data set gathered over two years of operation from the solar tower in Jülich was divided into various train/test/validation sets. The results demonstrate that for the given data set the distribution and balance of the calibration data set with respect to the sun angles have a more significant impact on calibration accuracy than time or data set size, as commonly assumed in the literature. According to our observation, we introduce a simple metric for a conservative assumption, which assigns relevancies to all data points within a data set based on the solar distance between measurements calculated by the sun's Euler angles (Azimuth, Elevation). The metric can be used as a quality metric for prediction accuracy and to make data collection more efficient. Using this metric, a tracking accuracy below 1 mrad with year round stability was achieved using less than 30 data points and a rudimentary heliostat model. Thus reduce the

required data set for a calibration accuracy smaller than to less than 30 data points for each training and validation.

Moreover, the findings of this investigation have implications for numerous preceding publications, raising questions about the reliability of their conclusions. Especially, since information about the applied data sets is often provided sparsely.

This paper's objective is to push the scientific community into reconsidering the impact of applied data sets on heliostat calibration and also to provide advice for future publication, how to publish results and data sets transparently and comprehensible, accelerating the deployments of new approaches.

2. Theoretical outline

2.1. The effect of data set sampling on stated Heliostat tracking accuracies

To understand the effect of data set sampling a simple thought experiment is conducted. First, a heliostat is considered, which does not change over time. Second, a heliostat fitting model is used in training, which does not perfectly cover all deficiencies of the real heliostat. Now, considering a data set "T" from May 21 used for training the heliostat model. It is tested with two different data sets. Test set "A" is from June 21. Test set "B" is from July 21. Both data sets contain similar amounts of data collected at similar times of the day. Since our training data set "T" is one month before the solstice, it has very similar solar altitudes to test data set "B", which is one month after the solstice and therefore requires the heliostat to perform almost equal sun tracking movements. Under these circumstances, it is very likely that the trained model will achieve a higher test accuracy when tested with the test data set "B" than with the test data set "A", even though it is the same model. Most of the information about the test data set is already included in the training. Moreover, if only data from test data A is available it can (wrongly) be assumed that the accuracy of a heliostat decays over time. Thus the distance e.g. measured in euclidean coordinates between the sun positions has an effect on the outcome of the fit. In the conducted thought experiment, we find that the distance between training and test set B is close to zero, while the distance to test set A is greater. Vice versa it can be concluded, that if a test data set achieves high accuracies but does not contain data with high distances to the training data set, the accuracies are not trustworthy, since the model has to interpolate or extrapolate less, or in a simplified assumption, the model can use the known data as a look up table for its tracking movements. A more detailed examination of this example will be provided in a subsequent section.

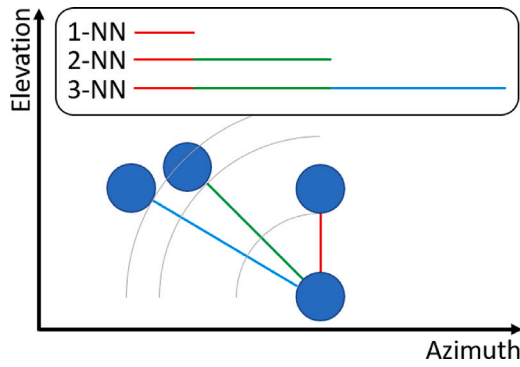


Fig. 2. Schematic drawing of the proposed nearest neighbor metric. The shortest distance is measured by the sun position given in Euler coordinates. The heliostat motor positions can be used as well, but do not provide comparability between publications. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.2. K-nearest neighbors metric

To take the thought experiment into account, we introduce a data point relevancy metric based on the euclidian distance of the training, test and validation set, the k-NN distance and demonstrate a correlation between model tracking accuracies and the suggested metric. In regression tasks, the basic goal of machine learning algorithms, is to enable computer systems to learn to discover patterns, relationships, or insights, such as parameter estimations, from data and use them to make accurate predictions. To cover all relevant patterns to a specific task, the training data set should contain samples over the given task's entire range of characteristics. If data points are not equally distributed over all behavior domains, a data set imbalance arises [13,14]. In case of an imbalanced data set, the data points can be sampled into an overrepresented majority group and an underrepresented minority group. Generally, the information of minority data points is less covered and thus more relevant. The difficult part for multi-dimensional data points and often unknown pattern distributions is to formulate a data point relevancy function over all dimensions.

We suggest handling the heliostat calibration as a regression over the parameter space, sun position, heliostat motor positions and target point to determine and correct the orientation of the heliostat. While the aimpoint is rather static and most often only swaps between operational and calibration target, the solar position changes continuously and thus is the more determining input. The solar position can be quantified either as date and time or as an Euler vector.

Since the solar position directly correlates with the heliostat's sun tracking and thus its actuator movements, solar positions relevant to the heliostat's operation can be regarded as indicators for behavioral patterns.

In this publication we apply the Nearest Neighbor (NN) distance metric to split our data set into training, validation and test set (compare Fig. 3). The NN [15] is computed as the minimum euclidean distance between a measurement's solar position, represented by azimuth and elevation to another measurements solar position. For evaluation, we also employ k-NN with $k = 1, 2, 3, \dots, k \in \mathbb{N}$, since they suite different purposes. For example, the 1-NN metric measures the distance from each measurements sun position to its nearest neighbor (Fig. 2, red line). The 2-NN metric also considers the second nearest neighbor and adds it to the 1-NN distance (red and green line). Higher k-NN orders take into account that, while several measurement points may cluster at one location, the cluster itself may be separated from the other measurement points. Here, the k-NN behaves like a regional density probability metric. A data point's probability is reversely proportional to its relevancy. The "distance" of two points to each other is computed over the sum of the k points i of minimal distance to the regarded point.

For multi-dimensional data points, the squared distance is obtained by the data points' scalar product. By applying the k-NN Metric on a data set, it can be sorted by the solar positions to each other. A k-NN based data set sampling sorts the existing data by the sun distances and, by first selecting data points of smaller distance as training data, makes it easier to learn the features in the training data set, harder to predict the validation data and hardest to predict the test data set. Since the test data set contains data with the highest range of inter- and/or extrapolation it can also be treated as a conservative assumption of the heliostat's accuracy over the year.

To create the k-NN data set sampling, first all data points are sorted by their k-Nearest Neighbor sun distance to any other available sun position. Then, the N_{test} data points of highest distance are selected as testing data points. Thereafter, out of the remaining data points, again the n_{val} data points of highest distance are chosen as validation data. The training data set is constructed from the data that is not selected as testing or validation data. If not all remaining data is selected to construct the training data set, data points of high distance are prioritized. This should ensure, that testing data is hardest to predict and thus is a reliable indicator for the trained models performance on all other data points that have a shorter distance to the training data. The algorithm for creating the data set is shown in Fig. 3.

$$k\text{-NN}(x, D) = \left(\sum_{i=1}^k \|x - d_i\|^2 \right)^{\frac{1}{2}} \quad d_i \in D \quad (1)$$

Whenever a "distance" is referenced within this paper the equation formulated in Eq. (1) between data point x and data set D is applied. This metric is closely related to the Split data set sampling by Joseph and Vakayil [16]. However, we use a simplistic approach, as e.g. we do not include angles between nearest neighbors, location in the distribution (center or edge) or weighting differences of azimuth and elevation.

2.3. Comprehensive heliostat model

The actual alignment of a heliostat is a combination of different static and dynamic errors (see upper half of Fig. 4). Common models begin with an orientation represented by f^* , which is derived from an ideal model based on the angles of the motors used (ν and τ). Deviations from this model are typically described by static parameters (f). Additionally, errors dependent on the heliostat's orientation are denoted by (f). Apart from these inherent factors, external influences like wind (g), temperature (k), time (k), or other disturbance factors (ϵ) can also be considered.

We test our method with a gradient-based regression of well-known machine learning techniques. For this purpose, we have created a comprehensive differentiable heliostat model (see lower half of Fig. 4). The model was first proposed by Pargmann et al. [17]. The heliostat's physical behavior is represented by an alignment model. The alignment model's foundation is a rigid-body kinematic system as it is commonly used in literature [18–26]. This model takes actuator positions $\vec{ax}$ as inputs and computes the corresponding concentrator plane normal orientation $\vec{n}$ and origin $\vec{o}$ as outputs. These outputs can then be combined with solar directions to compute the heliostat's reflective behavior. Our model is able to map all two axis heliostats and thus all mentioned models. It is inspired by (differentiable) robotic operating systems (ROS), where each joint forms the origin of a new coordinate system. Each coordinate system can be rotated and translated around all three axes but is part of the coordinate systems chain and thus has six degrees of freedom. Furthermore, manipulating a parent coordinate system affects all its child systems. Using this approach not only allows computation of a heliostat's alignment in global coordinates from its actuator configuration but also to use the coordinate system's inverse principle to gain the actuator configuration from a given target alignment. Therefore, each heliostat has two actuated coordinate systems at each joint plus an additional coordinate system at its concentrator's center point. Each actuator is modeled with 5 optimizable parameters,

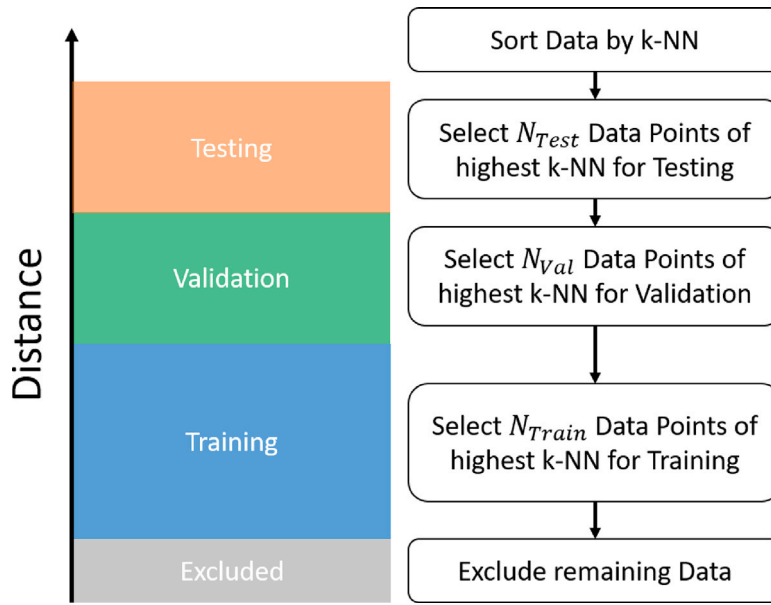


Fig. 3. Visual representation of the algorithm to apply the k-NN data set sampling.

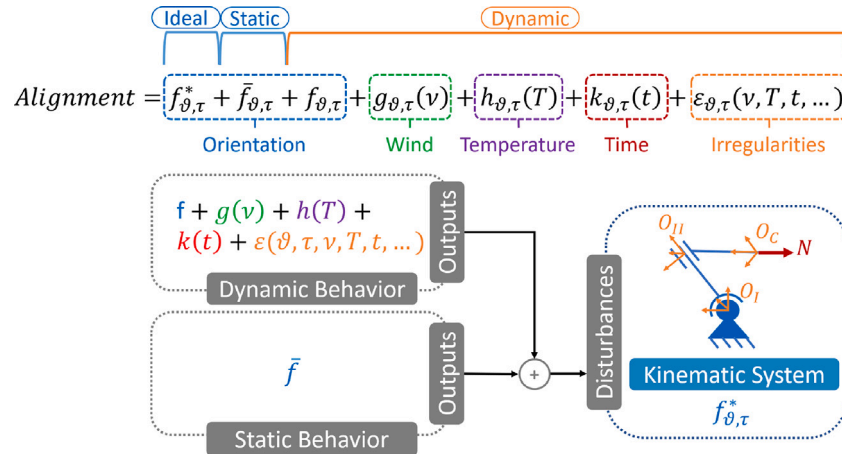


Fig. 4. Our kinematic model consists of an ideal heliostat behavior, which can be distorted by up to 28 deficiency parameters (static behavior). Each of these parameters can additionally be modeled by dynamic functions (dynamic behavior).

adding a total of ten more degrees of freedom to the model. Taking all degrees of freedom into account, this results in 28 possible parameters for optimization. Moreover, each parameter can be modeled by its own function to simulate dynamic heliostat behavior. This can be e.g. a polynomial function, a differential equation or even a neural network. The latter was used in [17] to achieve very high accuracies on small data sets. The model is to our best knowledge universally applicable for all kind of two-axes heliostats and deficiency sources. It is written entirely in Pytorch, so the training of the model profits from highly GPU optimized linear algebra. The entire model will be made publicly available.²

3. Measurement data

To analyze the effect of different data set splitting methods on reported results, a calibration data set from the solar tower in Jülich was used, which was collected using the camera-target method. It was chosen because, at the time of this work, it was the largest data set

available in the field, with almost 500 measurements, and thus offered the most flexibility. All data points were collected during fully automated daily power plant operations, and no additional measurement campaigns were conducted for this publication. However, there are some measurements taken outside of the daily routine. These measurements were obtained during multiple days of continuous heliostat tracking observations in July 2021. As no further information about this measurement campaign is available, the data must be considered separately. The data set is shown in Fig. 5.

In the upper panel, the measurement data is structured chronologically, and a training (yellow stars), test (red squares) and validation (blue dots) data set splitting is applied, as it would be used at the solar tower and according to literature (Refs. [5,27], and Smith and Ho [28]). A significant number of data points were obtained during the summer months in 2021 due to favorable weather conditions. The sampling assumes that the difficulty of the prediction increases with time. While the validation set should still be well predictable, the test set should already cause significant difficulties, due to the temporal distance of over three months.

In the lower panel, the same data is shown in Azimuth-Elevation representation using the same temporal sampling as the upper panel.

² <https://github.com/ARTIST-Association/ARTIST>

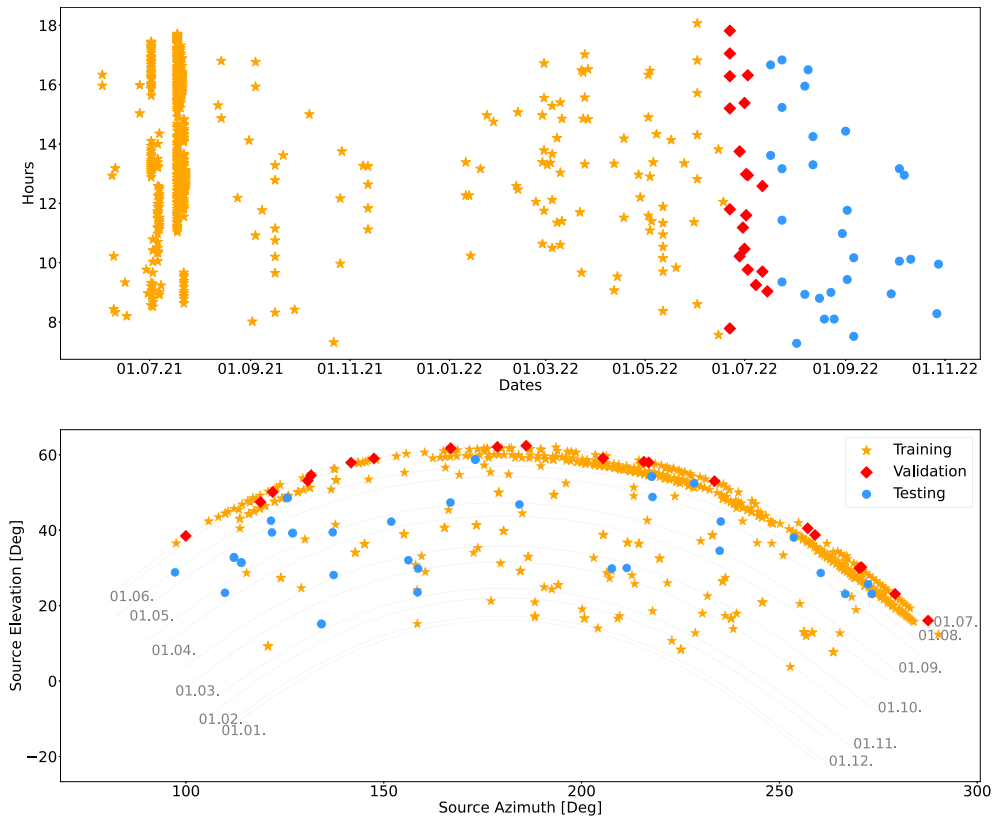


Fig. 5. Measurement data from the solar tower in Jülich, gathered by the camera-target method. The upper panel shows the measured data set chronologically using a temporal training–evaluation–test sampling. The lower panel shows the same sampling, but plots the measurement data depending on the sun position.

The measurements from summer 2021 cluster here as well, but it becomes apparent in this representation that the measurements taken in July have a high overlap with the test and validation set. The information on how the heliostat should be aligned to the sun positions in the test set is therefore (at least partially) already contained in the training data.

Due to the size of our data set, we will test the effect of time and euclidian (k-NN) splitting as well as the effect of the size of the training data set. The dataset itself will be available, so additional hypotheses can be tested in further research.

4. Results

4.1. Proof of concept using a single data point

First we want to test whether time or solar distance has the greater influence on the test accuracy. To achieve this, we used the geometric model described before and trained it on a single data point and validated the remaining points based on their NN distance to the training point. Because this paper's focus lies in evaluating the impact of calibration point distribution on the achieved tracking error, the applied model's impact on the result is regarded as a constant factor and the tracking error change between different data set splits is discussed. For evaluation the tracking error is used, which is defined as the angle δ_{track} between the predicted $\vec{n}_{pred}$ and the measured concentrator alignment vector $\vec{n}_{actual}$ of the heliostat [17] by $\delta_{track} = \arccos \frac{\vec{n}_{pred} \cdot \vec{n}_{actual}}{\|\vec{n}_{pred}\| \cdot \|\vec{n}_{actual}\|}$.

Fig. 6 shows the measured data set collected in Jülich. The upper graph shows the data points sorted by their time of measurement. The graph in the middle shows a representation in Euler angles. The color in all plots indicates the time of the measurement. The radius of the circles is directly proportional to the tracking error achieved for each individual measurement in the test data set.

Both the time-dependent and Euler-angle plots in the upper section indicate a steady increase in error with increasing temporal or spatial distances, which is consistent with e.g. [5] claim that time is a crucial parameter for calibration. However, some predictions in the time-dependent plots are very accurate despite the huge time distance. Moreover, in the Euler representation very close distant sun positions result in similar predictions, despite the temporal distance. The lower plot supports this observation and contradicts [5]'s claim. It suggests that time has only a marginal impact on the results. This plot illustrates the relationship between the distance to the training point and the prediction accuracy. As expected, the accuracy reduces with increasing distance to the training point. Moreover, the dependency can be assumed as linear. Measurements separated in time do not show any noticeable trend of increasing inaccuracies due to increasing absolute temporal differences.

4.2. Statistical evaluation on data set sampling with constant splits

When the training set includes more than one data point the evaluation can only be assessed statistically, since local distribution effects become of higher relevance to the prediction accuracy.

To conduct our analysis, we used a subset of 110 data points. The first 60 data points, arranged in chronological order, were selected for training the model, followed by the next 30 data points for model validation, and the subsequent 20 data points for testing. After completing the initial round of training and testing, we modified the data set by removing the oldest day and adding data from the following days until the data set once again contained 110 data points. By repeating this process, we were able to assess the performance of the heliostat across different seasons, while ensuring that the data set size remained constant throughout the analysis. Then, the model was trained with different training–validation–test partitioning using a time-dependent, and 1-3NN sampling. In Fig. 7 the achieved accuracies of all subsets

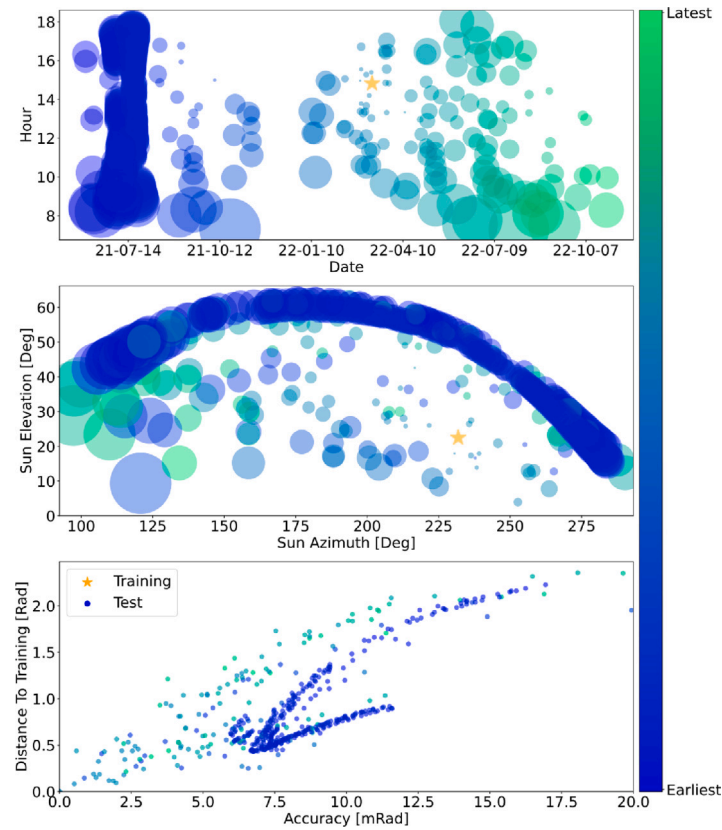


Fig. 6. Training with a randomly chosen data point (yellow star). The upper plot exhibits the validation data arranged chronologically, whereas the middle graph showcases their Euler coordinate representation. Both plots demonstrate an increasing prediction error with greater distances in time or space (indicated by the circle size). The color indicates the time of the measurement in between the earliest and the latest data point in the set. The lower plot exhibits a significant correlation between spatial distance and prediction accuracy, characterized by a linear relationship, with the training data being placed at both axes' origins. Conversely, no discernible temporal pattern can be observed in the data. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

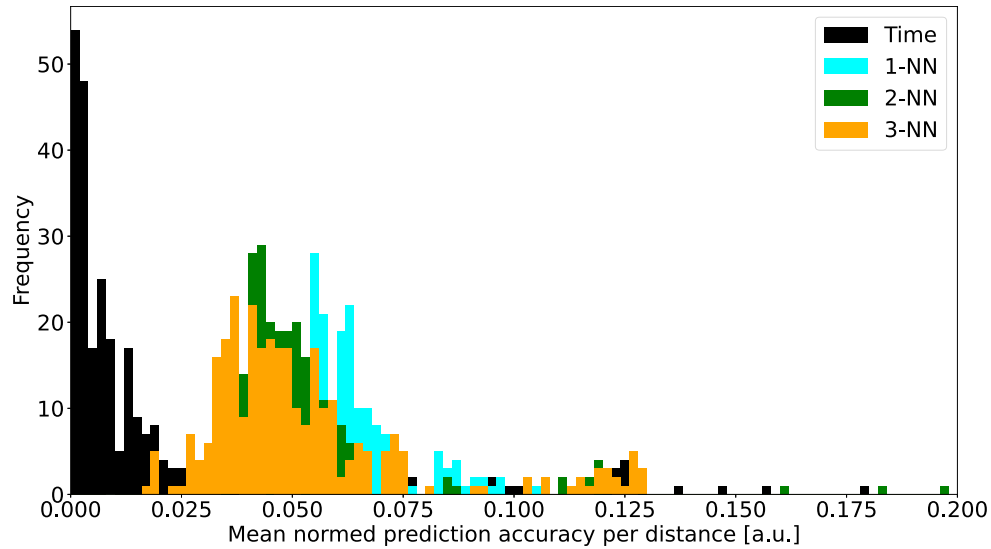


Fig. 7. Frequency of achieved accuracies in dependence to the test set distance. For *Time* the distance is measured temporal, for the other sets spatial.

divided by the mean distance of the subsets validation set to the training set is plotted against their frequency. For the spatial data set sampling (1-3NN) this is done by the spatial- and for the time depended sampling by the temporal-distance.

As can be seen, all three spatial data set sampling achieve a normal distribution different from zero very close to each other. This indicates a linear slope with some scattering due to measurement uncertainties

and local minima in the regression. The time dependent sampling centers around zero, so we do not see any temporal dependency.

Based on the preceding analysis, it can be inferred that the sun position distribution has a greater relevance to the tracking accuracy than time. However, the influence of time cannot be ruled out. In order to further explore the time influence, the first 60 data points were used as training data, the following 20 points as validation data and

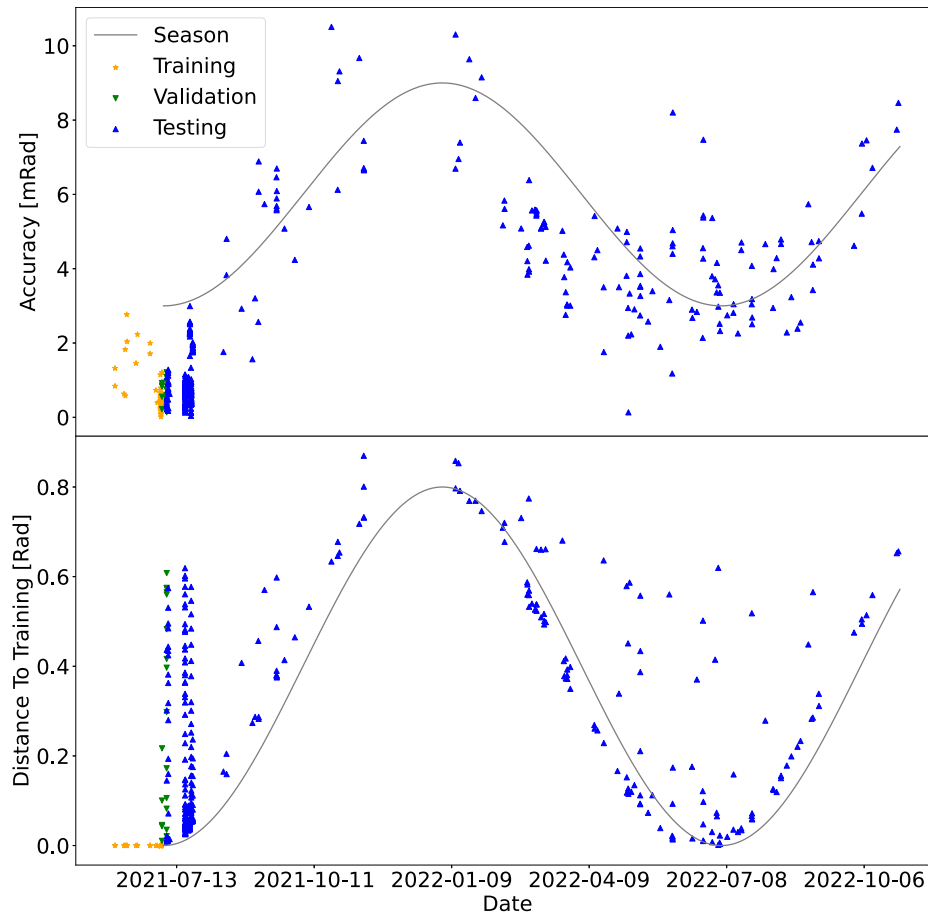


Fig. 8. Training with a time-continuous amount of 60 data points for training and 20 for validation. All other data points are then used for testing. An arbitrary cosine function with a frequency of one year is plotted, referencing seasonal changes.

all other points as testing data. Thus all training and validation data was selected from summer data, while the testing data is distributed over the entire year. The upper plot of Fig. 8 illustrates the trained model's prediction accuracies. As expected, validation and training data points are predicted most accurately. The prediction results on the test data set, in contrast are of higher variance. One distinctive feature of the resulting accuracies is a sinusoidal behavior over time with time spans of high and low prediction accuracies. This behavior can be explained by the sun's relative position over the year. For comparison, we added a cosine function plot with frequency of one year and an arbitrary amplitude to Fig. 8. The cosine was shifted to match the position of its negative amplitude to the first day within the testing data set. As can be seen, the prediction accuracies' behavior resembles the seasonal curve. Furthermore, prediction accuracies for summer data are generally better than those for data in winter months, due to the fact that training took place in summer. Data points that are separated by a difference of one year show similar results, except for the aforementioned measurement campaign, that performs slightly better. Due to winter months having lower solar elevations and thus greater distances to the training data, the seasonal behavior can also be observed within the data points distances. This is shown in the lower plot of Fig. 8.

It should be noted, that independent of the partitioning metric, accuracies between 0.1 and about 10 mrad were achieved. Other samplings of the data set may result in even higher variances. A published accuracy can therefore only be reasonably evaluated in connection with a corresponding spatial distance to the training data set.

4.3. Model behavior on different training set sizes

We conducted further analysis to investigate the impact of data set size on heliostat calibration quality, after excluding time as the main parameter. For this, we trained different models on various data set sizes using 3-NN data set sampling. We used a model with 6 parameters (*Static 6*), which utilized the same parameters as those used in Jülich. Additionally, we used a model with 20 free parameters (*Static 20*) and a model with intermediate number of parameters (*Static 14*).

Fig. 9 shows the mean test accuracy plotted against the data set size for these models. The prediction accuracy increases rapidly until 25 data points and then continues to increase until around 100 data points. Beyond that, the accuracy remains relatively constant or decreases slightly. The best accuracy of 1.7 mrad is achieved by the *Static 20* Model at 100 data points.

The plot below also depicts the mean NN distance from the test to the training data set plotted against the data set size. The curve's shape is strongly correlated with the prediction accuracies. When the k-NN value stops decreasing, the accuracy also plateaus. Furthermore, adding more data without reducing the NN distance makes the data set more imbalanced, ultimately reducing accuracy.

From this we can deduce, that the data set size is secondary. A well balanced data set, which reduces the NN-Distance the most, can achieve far better results than a larger unbalanced data set with close packed data points of low NN distance and thus relevancy.

Fig. 9 only showcases the mean value of the complete test data set. To present the full information of how well the algorithm behaves each data point should be depicted individually as it is done in Fig. 10. Here, the best result from Fig. 9 using 100 training data points using the 3-NN sampling are shown. The transparent circles with gray borders

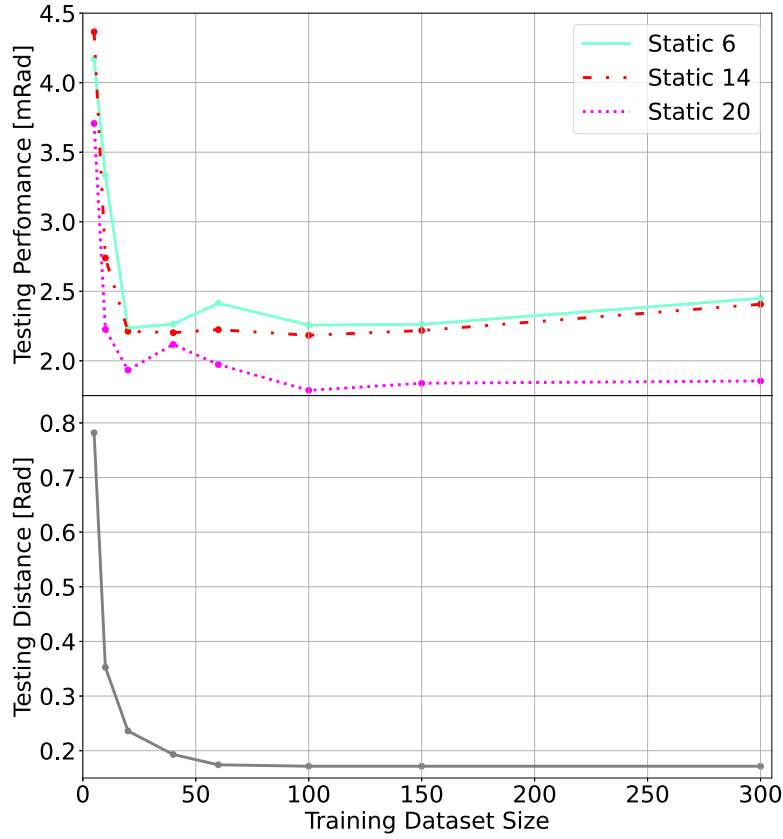


Fig. 9. Training accuracies of geometric models using different amounts of free parameters (upper plot). Every stated accuracy is connected with a distance to the evaluation data set (lower plot). This link allows a quantitative comparison between stated accuracies from different data sets. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

indicate an accuracy of 1 mrad as reference, while the colored circles, changing in size indicate the predicted accuracies. The circles' radii linearly increase with the prediction error. It easily becomes apparent, that the data points, inside the region of the training data distribution are very well covered, while the prediction accuracy is strongly reduced at the edges. This is to be expected, due to the higher amount of extrapolation. Despite the rather high mean prediction error of 1.7 mrad (compare Fig. 9) the plot indicates clearly, that the heliostat behavior will be reliable throughout the year, maintaining or exceeding the reported accuracy, except for minority data points (e.g. in early mornings). This can also be reduced by intentionally adding more training data to a minority domain.

4.4. Data point selection based on distance

In Fig. 6 an approximately constant behavior of the heliostat up to a radial distance between training and testing data of 0.17 rad can be seen. Thus, for the heliostats in Jülich a prediction in this region is as good as the accuracy reached in training. The same characteristic can be concluded from Fig. 9, which indicates that most performance improvements were gained until an average distance below $R_{region} = 0.25$ rad was achieved between training and testing data. Given R_{region} we can compute an upper limit for the required amount of calibration points for each heliostat by $N_{calib} = \frac{D_{azim}}{R_{region} \cdot 2} \cdot \frac{D_{elev}}{R_{region}}$ for a heliostat's operation ranges in azimuth and elevation D_{azim} and D_{elev} . A partial overlap of the created regions is required which is indicated by the elevation only being sampled by the regions' radius while the azimuth is sampled by their diameter. For Jülich this would result in approx. $N_{calib,J} = \frac{4.2}{0.25 \cdot 2} \cdot \frac{1.05}{0.25} = 36$ data points. The thus required solar positions can furthermore be pre-determined and calibration campaigns scheduled throughout the year. The time span between required calibration

days D_{days} can be computed by $D_{days} = \frac{R_{region}}{D_{elev}} \cdot 365$ days for a known maximum solar elevation EL_{max} at the given location and between the hours per day $D_{hours} = \frac{R_{region}}{\pi} \cdot 24$ h. For the given R_{region} this would result in approximately 87 days between calibration days and every 114 min per day.

The real operational range of a heliostat however is more restricted due to the sun's movements not covering the square assumed above. To estimate the required data points an algorithm was applied that selected a calibration point as training data point when the distance to its nearest neighbor within the already selected training data points exceeded the given region size of R_{region} . Fig. 11 visualizes the resulting data set split. All training data is shown in blue with surrounding blue regions of 0.25 rad radius in the azimuth-elevation plot. As far as possible, given that the available data set was not collected according to the suggested algorithm, the entire angle range is covered by the surrounding 27 training regions. The validation data, shown in green, is selected equally. All other data is selected as testing data, shown in orange. Two validation data points at the outer edges of the azimuth range appear show significant deviations from the average prediction performance and might be outliers. All testing data is predicted at an accuracy below 1.5 mrad. The achieved tracking error appears to outperform the results indicated in Fig. 9, where 100 calibration points achieved an accuracy of 1.7 mrad. The algorithm applied in Fig. 9 however ensured that the distance between training and testing data was maximized, thus making it harder for the model to achieve good prediction accuracies. The region based training data point distribution in comparison is designed to restrict the maximum distances between the training and testing datasets to a limit that was found to enable the model to make accurate predictions and therefore must result in an apparently better performance due to the easier to predict testing data.

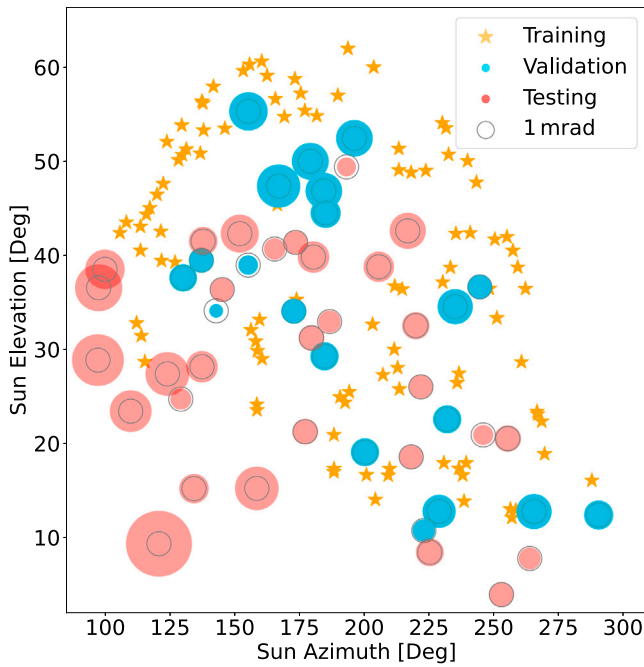


Fig. 10. Transparent illustration of the used train-test-validation sampling while indicating the prediction error of the test-set error. The size of the circles is proportional to the tracking error. The colorless circle represents a tracking error of 1 mrad and is intended to provide a simple estimate of how good the prediction was. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5. Implications

In this work, we investigated different data set sampling methods by applying them to a data set containing calibration data from 2 years of operation. In this context, we conducted the following claims

1. The proposed metric based on the euclidian distance of the training and test data set to each other can be taken as a good performance indicator for evaluation tracking accuracies. In particular it is better suited than any tested time dependent metric
2. A time dependent split can be misleading when presenting tracking accuracies.
3. A comparison between different calibration methods without detailed knowledge about the data set and the sampling methods used can lead to misleading conclusions (as observed in [12]).

From the before stated claims we can also derive potential criteria for the selection of calibration points per heliostat:

1. The solar distance (e.g. Nearest Neighbor) to already collected calibration points
2. The heliostat movement operating range coverage with calibration points, that may also be indicated by solar position regions (e.g. 0.25 rad) for the heliostats in Jülich.
3. While we cannot rule out temporal impacts for other heliostats, we cannot observe any on the data available to us. Further investigations should be considered by the community who has access to data of other heliostats.

In the next section, we assume that all of these points, including the last two, hold true for other heliostats as well, even if the approximation of independence in time and the trust region might only apply to smaller intervals.

5.1. Implications on literature

Depending on the distance between training and test data set, more or less information about the prediction task is already included in the optimization process and influence reported accuracies. In the following we investigate already published heliostat models and use the proposed metric to analyze the reliability of the stated accuracies. All stated calibration point positions are estimations reconstructed from the provided information within the given publications, using an acknowledged solar positioning algorithm [17] where necessary to transform time data into solar positions. Stated distances and visualizations are therefore subject to estimation errors due to a lack of provided data within the publications, but should be sufficiently reconstructed to indicate a lack of reliable data set selection methods within the publications. The results of this comparison are summarized in Table 1

1. Khalsa et al. [27] trained their model on data collected between 12:30pm–4:40pm on June 16 and 9:30 am and 3:55 pm on June 17, 2011 at an interval of 15–30 min and evaluated on data of July 15, 2011 12:53pm to 4:30pm at an interval of 3 to 4 min. The authors also stated environmental conditions during their measurements. For plotting, they chose a data representation using hours as the x -axis. A training data set size of (250 min + 385 min) / (15–30 points/min) was stated, equaling roughly 30 data points. We used the given information to derive a representation in Euler angles (Fig. 12(a)). Although the training and evaluation set is clearly separated in time, both training and testing data was taken close to the summer solstice and at the same time of day. The mean distance of the test set to the training set is 0.072 rad.
2. Smith and Ho [28] evaluated their model on a major time difference of six months. Training data was collected between August 9–27 2012, October 22–29 2012 and February 4–5 2013. Two testing campaigns were completed between October 30 to November 1 2012 and February 12–27 2013. Data per heliostat was collected every other day at 1 h intervals. The corresponding Euler representation is shown in Fig. 12(b) The authors waited a very long time to evaluate their model, that could have resulted in a mean distance between test and training data of 0.132 rad. However, it is evident that despite a significant time difference between October and February, the solar angles remain similar due to their proximity to the equinox. Due to the one training day in February, the mean distance to the training data set is greatly reduced to 0.082 rad, also affecting the reliability of the method.
3. The study conducted by Nadia AL-Rousan and Desa [5] utilized the amount of data as the x -axis. However, due to the absence of time information, an azimuth/elevation representation could not be derived. It can be inferred that the measurements were conducted between December and June, over a period of approximately 15 days. The data set consisted of 153 training samples, and a train/test sampling ratio of 70:30 was employed. Nonetheless, it was not clear how the separation between the sets was performed, as the different subsets were not labeled in the accompanying plots. Although the authors provided a precise description of their experimental setup, the presentation of their data by considerations of this paper's indications was insufficiently transparent, which hindered the reproducibility of their findings. Considering the use of degrees (instead of mrad) in the y -axis and the limited information about the data set, further investigation is required to assess the suitability of the reported accuracy for practical applications on solar towers.
4. Pargmann et al. [29] used a similar representation as [5]. However, the x -axis also states the exact date of measurement. Data was used from the end of July to the mid of August for training, and tested on data starting at the end of August to early

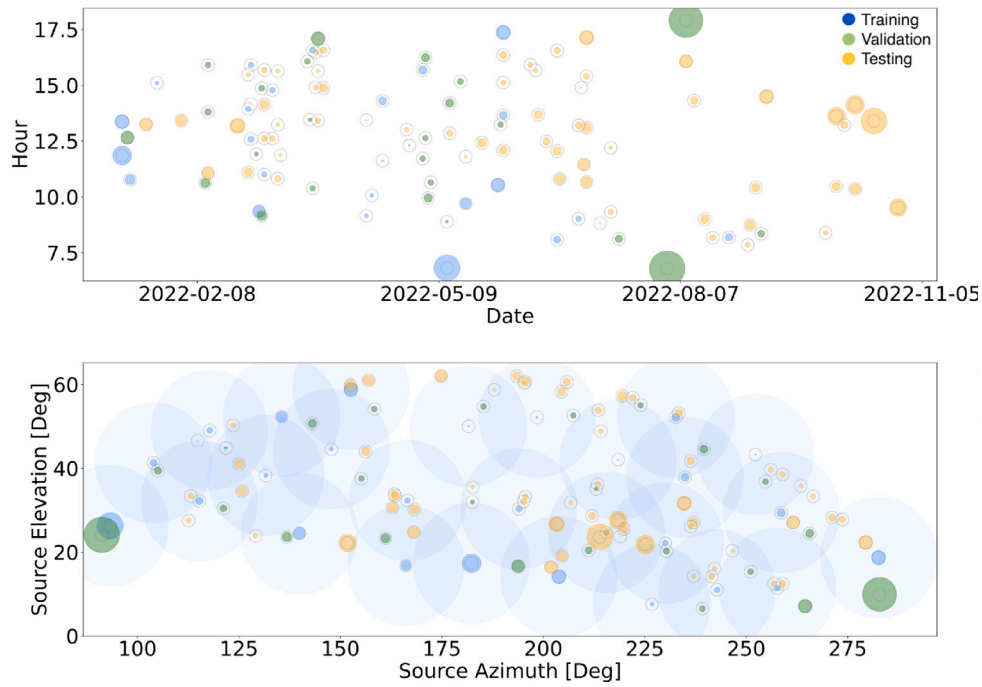


Fig. 11. Achieved pre-aligned model predictions on the 2022 AJ.23 after being trained and validated on data points selected by the data point sampling based on regions. The upper subplot shows the data points' time of creation distribution plotted by hour over date. The lower subplot shows the data points' solar position distribution plotted by solar elevation over solar azimuth. The sampling regions are indicated as transparent blue areas. Prediction accuracies are indicated by the data points marker sizes. The larger the marker, the higher the prediction error. Reference markers (gray circles) are added to each data point indicating a reference prediction accuracy of 1 mrad. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

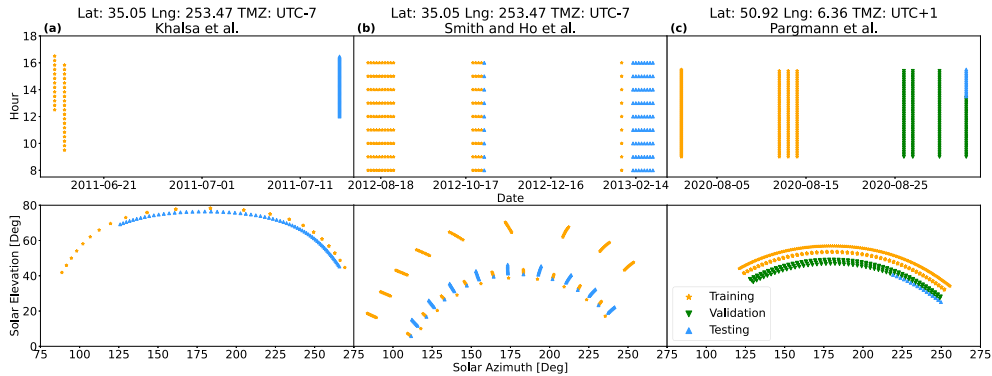


Fig. 12. Sun angles selected for training, validation and testing derived from different publications. The upper row represents the temporal representation, the lower the representation in Euler coordinates. Data was taken from the publications of Khalsa et al. [27], Smith and Ho [25] and Pargmann et al. [29].

Table 1

Tracking accuracies with high precision from different publications, which will be discussed later. Even lower accuracies can be found in literature [12].

	Khalsa et al (2011)	Smith and Ho (2014a)	Pargmann et al (2021)
Accuracy	0.1 mrad–0.4 mrad	0.78 mrad–1.5 mrad	0.42 mrad
Dataset size (Train/Validation/Test)	approx. (30/60/–)	Split not clear	for best accuracy approx. (300/175/25)
1-NN Distance	0.072 rad to validation set	0.082 rad to validation set	0.09 rad to test set

February. In Fig. Fig. 12(c), the sun positions used for training are determined on the basis of the information given in the paper. The mean distance between testing and training data set is 0.121 rad, between validation and training data set is 0.09 rad and between testing and validation data set is 0.026 rad.

These publications did not publish their datasets, did not provide complete information about dataset size or splitting methods, or did not use a three-part dataset, which are important considerations in meeting machine learning standards. However, they provided enough information to reconstruct at least parts of the missing data. Additionally, some

publications (as summarized in [7]) provide limited information, which poses challenges for reconstructing their methods and fully interpreting their results.

The mean distance between the training and test sets of the mentioned publications falls between 0.072 rad and 0.0121 rad. This indicates that the test distances are significantly lower than the range associated with constant behavior in our findings, which may affect the validity of their reported accuracies. To illustrate this by another example. If we take our data set and enlarge the training data set by adding more points until the kNN distance to the test set hat fallen to 0.07 rad, we can completely dispense with a geometry model. Instead,

to predict a point in the test data set, we take the next two data points in the training set in elevation and azimuth direction and interpolate the motor steps linearly. With this very simple method, we obtain a prediction accuracy of 0.35 mrad. While this simple method yields better results than those reported in [25,27], and [29], it may not maintain performance over an entire year and could face challenges in practical power plant operations. For different types of heliostats, this linear range can be smaller; Based on our analysis of the Jülich heliostats, the quality conditions we derived highlight areas where these publications may not fully align with our suggested standards.

An interesting case occurs when examining the data set used by Smith and Ho [25]. The kNN metric for this publication produces comparable results to those observed in the other publications. However, an examination of the azimuth and elevation plots reveals that a training–test split with larger distances could have been conducted by clearly delineating the winter and summer data sets. Instead, individual points in winter fall within the training area. By selecting this split, the tracking accuracy could be significantly higher than what would be achieved over the course of the year. To address such special cases, a representation similar to Fig. 10 would be beneficial and could enhance future publications.

5.2. Implications on upcoming publications

From the results presented in this paper and the analysis of previous publications, 3 conclusions can be drawn.

First, we highly recommend to present calibration data in upcoming publications using a presentation alike Fig. 10 using the proposed azimuth/elevation representation.

Second, accuracies should be linked to a corresponding distance to the training set as in Fig. 9. Although the distance might not be the only domain coverage criteria, it is a good, easy to implement quantitative indicator.

Third, we also recommend utilizing a 3-sampling strategy for training/evaluation/testing, which is not extensively covered in the literature on evaluating the accuracies of calibrated heliostat models.

5.3. Implications on the calibration

Taking into account the stated criteria for heliostat calibration point selection from subsection,

a data set for each heliostat can be gathered, spanning over all possible sun positions on just 4–5 measurement days under the assumption of a approximately constant behavior region of 0.25 rad, which can be planned in advance. Considering such a data set and our results in this publication, we expect a rather constant heliostat behavior for the next years. In our investigation, as depicted in Fig. 11, implementation of this approach has yielded a year-round accuracy below 1.5 mrad using only 27 calibration points. Notably, this level of precision surpasses that of alternative data set partitions tested, despite employing significantly more data. It should be noted, that the amount of calibration measurements can still be reduced by collecting the data applying utilizing the kNN Metric as follows:

- Initiate calibration by recording the current sun position.
- Determine the next heliostat for calibration based on the relevancy score of available data points to the current sun position.

Although clustering of measurement points remains feasible, its incidence diminishes. Moreover, establishing a minimum nearest neighbor (NN) distance serves to mitigate clustering further.

5.4. Implications on calibration at the solar tower in Jülich

The kNN algorithm was implemented at the solar tower in Jülich for the purpose of collecting new calibration points. Prior to the implementation of the algorithm, approximately 230,000 data points were collected at the solar tower and utilized for calibration purposes. With this considerable number of measurements (approximately 100 per heliostat), the field exhibited a tracking error of approximately 2 mrad. Following the introduction of the new algorithm, this value decreased to around 1.7 mrad within a month, without the necessity for an additional measurement campaign.

5.5. Implications on daily power plant efficiency estimation

Our study shows that the accuracy of heliostats is mainly affected by the position of the sun, rather than time. This crucially affects the prediction accuracy of heliostats for upcoming years. Since the time was mistakenly considered as the primary determinant, the accuracy of a heliostat for the following year could only be estimated coarsely. However, since heliostat accuracy is primarily reliant on the sun positions, precise predictions can be made for any given day of the (next) year, by inserting corresponding sun angles as lines in Fig. 10 and interpolating the values of the evaluation data set. Approaches to this already exist in [30] The proposed approach yields more accurate predictions than using the mean accuracy.

6. Conclusion & outlook

In this research, we introduced the widely used k-NN algorithm to the field of heliostat calibration and explored the influence of different data set sampling strategies on the achieved accuracies. Our study has revealed the crucial role of solar position coverage in determining the relevancy of heliostat calibration data points, surpassing the previously presumed significance of time or data set size. Notably, we found for the heliostats in Jülich that time only has a marginal influence on their precision, while an excessively imbalanced data set can be detrimental in worst-case scenarios. Moreover, we found a sphere around training points where the heliostat behaves linearly, so linear interpolation can achieve accuracies below 0.35 mrad as long as the prediction is inside this range.

In the second part of the paper we assume that other heliostats behave of different data sets similar to those of the heliostats in Jülich. Under this assumption, many publications in literature do not provide reliable results, since their test data set is inside the sphere of linear behavior. While previous publications have advanced the understanding of heliostat calibration, there are areas where additional transparency regarding data sets and sampling methods could enhance reproducibility and comparability. By providing detailed information about data set sizes, splits, and adopting standard machine learning practices, future studies can build upon this foundation more effectively. To prevent these issues in the future, we recommend using the k-NN-distance as a quality-assurance metric, which should always be disclosed alongside published tracking accuracies. This, combined with the presentation of data set and results as demonstrated in this work, facilitates comprehensible and comparable outcomes across different calibration approaches and helps avoid errors in subsequent studies.

We found k-NN to be a reliable and simple to understand metric to estimate a calibration points accuracy. However, already from our observations it became clear, that there is no linear dependency between k-NN and the prediction error if more than one training data point is applied. This indicates that k-NN on solar positions might be a good basis for future research into heliostat calibration point relevancy scoring algorithms. More sophisticated classification algorithms could potentially outperform the k-NN metric as a relevancy indicator.

The potential to calibrate a heliostat with minimal data – around 30 points – while maintaining high accuracy and long-term stability presents a significant opportunity for the future of CSP technology. However, the question that remains open is which model will ultimately deliver the highest accuracy.

CRedit authorship contribution statement

Max Pargmann: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Moritz Leibauer:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Vincent Nettelroth:** Supervision. **Daniel Maldonado Quinto:** Supervision, Funding acquisition. **Robert Pitz-Paal:** Funding acquisition.

Declaration of competing interest

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

References

- [1] S. Karellas, T.C. Roumpedakis, Chapter 7 - solar thermal power plants, in: F. Calise, M.D. D'Accadia, M. Santarelli, A. Lanzini, D. Ferrero (Eds.), *Solar Hydrogen Production*, Academic Press, 2019, pp. 179–235, <http://dx.doi.org/10.1016/B978-0-12-814853-2.00007-2>, URL <https://www.sciencedirect.com/science/article/pii/B9780128148532000072>.
- [2] C. Corsi, M.J. Blanco, V. Grigoriev, J. Pye, Upper limits to the mean annual optical efficiency of solar mono-tower systems, *Sol. Energy* 236 (2022) 88–99, <http://dx.doi.org/10.1016/j.solener.2022.02.038>, URL <https://www.sciencedirect.com/science/article/pii/S0038092X22001438>.
- [3] K.W. Stone, Automatic heliostat track alignment method, 1986, US Patent 4, 564, 275.
- [4] L. Oberkirsch, D. Zanger, B. Hoffschmidt, Validation of a closed-loop aim point management system at the jülich solar tower, *Sol. Energy* 264 (2023) 111951, <http://dx.doi.org/10.1016/j.solener.2023.111951>, URL <https://www.sciencedirect.com/science/article/pii/S0038092X23005844>.
- [5] N.A.M. Nadia AL-Rousan, M.K.M. Desa, Efficient single and dual axis solar tracking system controllers based on adaptive neural fuzzy inference system, *J. King Saud Univ., Eng. Sci.* 32 (7) (2020) 459–469.
- [6] J. Armendariz, C. Ortega-Estrada, F. Mar-Luna, E. Cesaretti, Dual-axis solar tracking controller based on fuzzy-rules emulated networks and astronomical yearbook records, in: *Proceedings of the World Congress on Engineering*, Vol. 1, 2013, pp. 3–5.
- [7] J.C. Sattler, J. Götsche, Progress in the development of a laser and camera system for the calibration of heliostat fields of central receiver systems, in: *AIP Conference Proceedings*, Vol. 2445, AIP Publishing LLC, 2022, 070011.
- [8] M. Burisch, L. Gomez, D. Olasolo, C. Villasante, Heliostat kinematic system calibration using uncalibrated cameras, in: *AIP Conference Proceedings*, Vpl. 1850, AIP Publishing LLC, 2017, 030007.
- [9] Hines, Heliostat characterization using starlight, 2017, US Patent US20180299264A1.
- [10] G. Bern, M. Bitterling, P. Schöttl, A. Ferriere, Y. Volut, A. Heimsath, P. Nitz, Experimental assessment of simultaneous in-situ heliostats calibration methodology HelioControl at themis facility, in: *AIP Conference Proceedings*, Vol. 2303, AIP Publishing LLC, 2020, 030005.
- [11] J.E. Gouws, Calibration of Heliostats Using a Drone (Ph.D. thesis), Stellenbosch: Stellenbosch University, 2018.
- [12] J.C. Sattler, M. Röger, P. Schwarzbözl, R. Buck, A. Macke, C. Raeder, J. Götsche, Review of heliostat calibration and tracking control methods, *Sol. Energy* 207 (2020) 110–132.
- [13] Y. Yang, K. Zha, Y. Chen, H. Wang, D. Katabi, Delving into deep imbalanced regression, in: *International Conference on Machine Learning*, PMLR, 2021, pp. 11842–11851.
- [14] P. Branco, L. Torgo, R.P. Ribeiro, Pre-processing approaches for imbalanced distributions in regression, *Neurocomputing* 343 (2019) 76–99.
- [15] T. Cover, P. Hart, Nearest neighbor pattern classification, *IEEE Trans. Inf. Theory* 13 (1) (1967) 21–27.
- [16] V.R. Joseph, A. Vakayil, Split: An optimal method for data splitting, *Technometrics* 64 (2) (2022) 166–176.
- [17] M. Pargmann, M. Leibauer, V. Nettelroth, D. Maldonado Quinto, R. Pitz-Paal, Enhancing heliostat calibration on low data by fusing robotic rigid body kinematics with neural networks, *Sol. Energy* 264 (2023) 111962, <http://dx.doi.org/10.1016/j.solener.2023.111962>, URL <https://www.sciencedirect.com/science/article/pii/S0038092X23005960>.
- [18] K.W. Stone, Automatic heliostat track alignment method, 1986, URL <https://patents.google.com/patent/US4564275A/en>.
- [19] Y.T. Chen, A. Kribus, B.H. Lim, C.S. Lim, K.K. Chong, J. Karni, R. Buck, A. Pfahl, T.P. Bligh, Comparison of two sun tracking methods in the application of a heliostat field, *J. Solar Energy Eng.* 126 (1) (2004) 638–644, <http://dx.doi.org/10.1115/1.1634583>, URL <https://asmedigitalcollection.asme.org/solarenergyengineering/article/126/1/638/451491/Comparison-of-Two-Sun-Tracking-Methods-in-the>.
- [20] Y.T. Chen, B.H. Lim, C.S. Lim, General sun tracking formula for heliostats with arbitrarily oriented axes, *J. Solar Energy Eng.* 128 (2) (2006) 245–250, <http://dx.doi.org/10.1115/1.2189868>, URL <https://asmedigitalcollection.asme.org/solarenergyengineering/article/128/2/245/478177/General-Sun-Tracking-Formula-for-Heliostats-With>.
- [21] K.-K. Chong, C.-W. Wong, F.-L. Siaw, T.-K. Yew, S.-S. Ng, M.-S. Liang, Y.-S. Lim, S.-L. Lau, Integration of an on-axis general sun-tracking formula in the algorithm of an open-loop sun-tracking system, *Sensors* 9 (10) (2009) 7849–7865, <http://dx.doi.org/10.3390/s91007849>, URL <http://www.mdpi.com/1424-8220/9/10/7849>.
- [22] M. Guo, Z. Wang, W. Liang, X. Zhang, C. Zang, Z. Lu, X. Wei, Tracking formulas and strategies for a receiver oriented dual-axis tracking toroidal heliostat, *Sol. Energy* 84 (6) (2010) 939–947, <http://dx.doi.org/10.1016/j.solener.2010.02.015>, URL <https://linkinghub.elsevier.com/retrieve/pii/S0038092X10001052>.
- [23] C.E. Andraka, C.K. Ho, S.S.S. Khalsa, An Automated method to correct heliostat tracking Errors, Tech. Rep., Sandia National Lab.(SNL-NM), Albuquerque, NM (United States), 2011.
- [24] M. Guo, F. Sun, Z. Wang, J. Zhang, Properties of a general azimuth-elevation tracking angle formula for a heliostat with a mirror-pivot offset and other angular errors, *Sol. Energy* 96 (2013) 159–167, <http://dx.doi.org/10.1016/j.solener.2013.06.031>, URL <https://linkinghub.elsevier.com/retrieve/pii/S0038092X13002612>.
- [25] E. Smith, C. Ho, Field demonstration of an automated heliostat tracking correction method, *Energy Procedia* 49 (2014) 2201–2210, <http://dx.doi.org/10.1016/j.egypro.2014.03.233>, URL <https://linkinghub.elsevier.com/retrieve/pii/S1876610214006870>.
- [26] J. Freeman, B. Shankar, G. Sundaram, Inverse kinematics of a dual linear actuator pitch/roll heliostat, *AIP Conf. Proc.* 1850 (1) (2017) 030018, <http://dx.doi.org/10.1063/1.4984361>, URL <https://aip.scitation.org/doi/abs/10.1063/1.4984361>, Publisher: American Institute of Physics.
- [27] S. Khalsa, C. Ho, C. Andraka, An automated method to correct heliostat tracking errors, *Proc. SolarPACES* (2011) 20–23.
- [28] E. Smith, C. Ho, Field demonstration of an automated heliostat tracking correction method, *Energy Procedia* 49 (2014) 2201–2210.
- [29] M. Pargmann, D. Maldonado Quinto, P. Schwarzbözl, R. Pitz-Paal, High accuracy data-driven heliostat calibration and state prediction with pretrained deep neural networks, *Sol. Energy* 218 (2021) 48–56, <http://dx.doi.org/10.1016/j.solener.2021.01.046>, URL <https://www.sciencedirect.com/science/article/pii/S0038092X21000621>.
- [30] V. Grigoriev, K. Milidonis, M.J. Blanco, G. Tsouloupas, Polyharmonic splines for interpolation over sun path, *Sol. Energy* 235 (2022) 209–218.