Analysis and Research on the Performance of Solar Concentration Based on Big Data and Machine Learning

Shaobing Wu¹², Changmei Wang* and Runsheng Tang*

¹ Education Ministry Key Laboratory of Advanced Technology and Preparation for Renewable Energy Materials, Yunnan Normal University, Kunming 650500, China
² Institute of Information Security, Yunnan Police College, Kunming 650223, China

Email: 823804919@qq.com; kingtang01@126.com

Abstract. Empirical formulas for PTC optical efficiency calculation are difficult and costly to obtain from rigorous comparative experiments, whereas simpler optical modeling methods inadequately incorporate realistic optical effects. In this article, algorithms are respectively developed to calculate the geometric concentration ratio \(C_g\) of linear Cassegrainian solar concentrators (CSC) with a secondary flat mirror based on the way of edge rays from solar sources to a flat-plate receiver. On the basis of the large amount of data generated, machine learning and Python language programming methods are used to analyze and process the data, and the functional relationship between the concentration ratio and each parameter is obtained. The learning and training effect is good, and the ideal result is achieved.

Keywords. Solar concentration; big data; machine learning.

1. Introduction

Generally, empirical formulas for PTC optical efficiency calculation are difficult and costly to obtain from rigorous comparative experiments, whereas simpler optical modeling methods inadequately incorporate realistic optical effects. Thus, new computing methods that are accurate, fast, and cost-effective for PTC optical efficiency calculation or empirical fitting formulas are urgently needed. As an accurate optical efficiency computing method for geometric designing, energy analyses, and optimization solar concentration collectors, Cheng Z D et al. [1] proposed fitting formulas for PTCs by combining Monte Carlo ray tracing (MRCT) and population-based particle swarm optimization algorithm. The optical efficiency fitting formulas were obtained by analyzing the Monte Carlo ray tracing data through a variable reduction technique, while the fitting formula coefficients were acquired from the particle swarm optimization algorithm. Wu S B et al. [2] proposed big data security framework based on encryption. Some scholars have conducted research on big data and machine learning. Bode G et al. [3] proposed real-world application of machine-learning-based fault detection trained with experimental data. Rodriguez S D et al. [4] developed the methods for Improving the concentration ratio of parabolic troughs using a second-stage flat mirror. Wu S B et al. [5] analyzed the crime data in YD county and proposed crime prediction using data mining and machine learning. In this paper, Machine learning and python language programming methods are used to analyze the performance of the secondary plane reflection solar concentrator, and perform machine learning on the generated data to obtain a higher accuracy of solar concentrator and the functional relationship between related parameters and solar concentration ratio.
2. Background

2.1. Linear Regression

Linear regression is a simple prediction model in which the error terms are assumed to follow a Gaussian distribution $\epsilon \sim N(0, \sigma^2)$ [6]. The linear regression formulation of the problem in the present study is expressed as follows:

$$\tilde{y} = a_1x_1 + a_2x_2 + \cdots + a_5x_5 + \epsilon$$

where $x_i$ represents nodal features such as reflectance, transmittance, absorptivity, intercept factor, and incidence angle distribution, while $a_i$ is the coefficient. On the right-hand side of Eq. (1), $\epsilon$ is the only stochastic variable, while other terms are fixed observation values.

2.2. R-square Value ($R^2$)

In statistics, the coefficient of determination $R^2$ in the model is used to measure the predictive ability of the model. The specific expression of $R^2$ is as follows [7]:

$$R^2 = \frac{\sum_{k=1}^{n}(y_k - \tilde{y}_k)^2}{\sum_{k=1}^{n}(y_k - \bar{y}_k)^2}$$

In the formula, $\bar{y}_k$ represents the mean value of $y_k$. The value range of $R^2$ is 0-1. The larger the value of $R^2$, the better the prediction effect. If the value is 1, it can be said that the regression model perfectly fits the actual data.

2.3. Machine Learning-Based Linear Regression

The best way to solve machine learning or analysis problems in the real world is to use a machine learning framework, from collecting data to using machine learning techniques to convert it into valuable information or knowledge. Supervised machine learning techniques are used in conjunction with supervised, labeled data to train the model and then predict the results of unseen test data samples. Certain processes (such as feature scaling, extraction, and selection) must remain the same in order to use the same features to train the model and extract the same features from invisible test data samples to test the model in the prediction phase [8].

Simplification of a complex problem, followed by simulation using software tools is the main approach used by researchers for studying concentration ratio for a secondary flat concentrator. In this study, machine learning is employed to solve a complex concentration ratio calculation problem. Linear regression is then utilized to generate a simple fitting expression. This expression is convenient for calculating optical efficiency and studying the optical performance of PCTs. Therefore, linear regression provides simple relationships that can be exploited for optics-related calculations.

First, we used this approach for designing the secondary reflection hyperbolic concentrator presented in Sections 3.1 and 3.2 after selecting the best performance parameter. Then, we calculated the concentration ratio introduced in Section 3.2, and employed the Python program for designing the concentration ratio by calculating and obtaining a dataset of parameters related to the concentration ratio. Finally, by considering the concentration ratio as target, a linear regression method in machine learning is used for its calculation. The learning results were evaluated using the evaluation index accuracy rate ($R^2$) in machine learning, with the effect getting better as $R^2$ approached 1, with the regression coefficients concurrently obtained.

3. Algorithms to Correlate the CR and Receiver’s Position of Linear CSCs

To be convenient for analysis, a coordinate system with the origin at bottom of parabola trough and x-axis normal to the aperture is used in this work, as shown in figure 1. In the suggested coordinate system, the primary parabola is expressed by

$$y^2 = 4fx$$

(3)
where \( \phi, D \) and \( f \) are the rim angle, width and focal length of primary parabolic trough, respectively.

### 3.1. Design Description of CSC with a Flat Mirror

The secondary reflector as flat mirror is called Newton structure system. Its concentrating principle is consisted of a main mirror and a secondary mirror. The primary mirror adopts a rotating paraboloid and the secondary mirror adopts a plane. It utilizes the parabolic and planar reflection properties: a parabolic concave mirror can converge all the light parallel to the optical axis to its focus; the planar reflection changes the focus of the concentrating light.

#### 3.2. Definition

The definition of the Concentration Ratio by \( C_g \) is strictly geometric, as ratio of aperture area to absorber surface area

\[
C_g = \frac{A_{ap}}{A_{abs}}
\]

(5)

Considering the shading, we denote this ratio by \( C_e \) and call it the efficient concentration ratio which is

\[
C_e = \frac{A_{ap} - \text{Shading}}{A_{abs}}
\]

(6)

where the \( \text{Shading} \) is that \( \text{Shading} \) occurs if direct sunlight rim angels to reach a mirror because the mirror is in the shadow of another mirror.

Another important definition by IF and call it the intercept factor which is

\[
IF = \frac{C}{C_e}
\]

(7)

**Figure 1.** Secondary reflection surface structure for the flat.

**Figure 2.** The rays \( r_1 \) and \( r_2 \) making an angle \( \theta \) to the axis of the parabola intersect at a point \( E \).

### 3.3. Linear CSC with a Secondary Flat Reflector (CSC-F)

Considering another two rays \( r_1 \) and \( r_2 \) making an angle \( \theta \) to the axis of the parabola, as shown in figure 2. After reflection, they intersect at a point \( E \) of the intersection of the straight lines passing through points \( A_1 \) and \( A_2 \) with the directions of rays \( r_1 \) and \( r_2 \) after reflection. It is given by
\begin{align}
&x_E = f - \frac{2f \sin^2 \theta}{\cos \phi + \cos^2 \phi} \\
y_E = \frac{f \sin 2\theta}{\cos \phi + \cos^2 \phi}
\end{align}

(8)

\begin{align}
&x_M = (1-m) \cdot f \\
y_M = k \cdot (x_M - x_E) + y_E
\end{align}

(9)

where \( k = -\tan(\phi - \theta) \).

The focal belt location of secondary reflection is:

\begin{align}
&x_B = 2x_M - x_E \\
y_B = y_E
\end{align}

(10)

From what have been discussed above, the efficient geometrical concentration ratio is computed by following:

\begin{align}
&C_e = \frac{D}{2|y_M|}, \text{ or } ECR = \frac{D}{2|y_M|} \\
&C_g = \frac{D}{2|y_B|}, \text{ or } CR = \frac{D}{2|y_B|}
\end{align}

(11)

(12)

4. Results and Discussion

4.1. Optimal Geometry of CSCs for Maximizing ECR

4.1.1. Optimal Geometry of CSCs for Maximizing ECR Based on Flat. As shown in figure 3, the relation between ECR and Shading percentage of flat mirror with rim angle increasing. The ECR of flat mirror increases with rim angle increasing, the rim angles vary from 27 to 87 at \( \theta=0.269^\circ \) and the Ce of flat mirror increases from 82.306 to 104.906 and then decreases to 13.777. With the rim angle increasing, the Shading percentage keeps below 5% all along.

Figure 3. The relationship between concentration ratio and shading percentage based on rim angle.

Figure 4. The relationship between the concentration ratio and focal plane position and rim angle.
4.1.2. Effects of Ce and Spot Location of CSC_F. As shown in figure 4, the relation between ECR and focal belt location of flat mirror with rim angle increasing. The ECR of flat mirror increases with rim angle increasing, the rim angles vary from 27 to 87 at $\theta=0.269$ and the ECR of flat mirror increases from 82.306 to 104.906 at rim angle =45 and then decreases to 13.777. With the rim angle increasing, the focal belt location decreases from 107.207 to 26.557.

4.1.3. Effects of ECR and Relative Location $m/f$ of CSC_F. As shown in figure 5, the efficient concentration ratio of flat mirror increases with rim angle increasing, while the relative position of focal point of flat keeps at about 0.005$f$ (f is focal distance) all along. The efficient concentration ratio of parabola increases from 82.306 to 104.906 at rim angle =45 and then decreases to 13.777 with rim angle increasing from 27° to 89°.

Figure 5. The relationship between the concentration ratio and the relative position of the plane mirror and the rim angle.

4.2. Big Data Visualization Based on Python
In order to analyze the relationship between the data and study the relationship between rim angle and Ce, f, Blocking, Spot, Ce/C0, Cg/C0, Ce/Cg, we derived the design of the secondary reflection plane concentrator from the previous the data obtained by multiple iterations is filtered and the relevant data is obtained as shown in figure 6 ($\theta=0.5^\circ$) and figure 7 ($\theta=1.0^\circ$).

4.3. The Functional Relationship between Solar Concentration Ratio and Various Factors Based on Big Data and Machine Learning

4.3.1. Data Collection and Preprocessing. The data used in this article comes from the previously built model, through multiple iterations, generated under the programming implementation of Python language programming, a total of 37,521 pieces of data, of which 19,869 pieces are based on $\theta=0.5^\circ$, and those based on $\theta=1.0^\circ$ There are 17,652.

4.3.2. Function Relationships between Efficient Concentration Ratio and All Sorts of Parameters
1. Function relationship based on $\theta=0.5^\circ$
   (1) Relationship between rim angle, f, m, blocking, spot and Ce.
   Based on big data, using machine learning linear regression model, with the help of Python language programming, the following functional relationships are obtained.
   \[ y_c = 188.1089 - 1.1525x_1 - 0.9228x_2 + 0.000000046x_3 - 66.8580x_4 + 74.7771x_5 \]
   The mean square error of train and test are: 0.73,0.39  
The coefficient of determination is: $R^2=1.00$
   Where x1 means $\phi$, x2 means f, x3 means m, x4 means Blocking, x5 means spot, the same below.
   (2) Relationship between rim angle, f, m, blocking, spot and Cg.
Figure 6. Relationships between Ce, Cg, f, Blocking, Spot, Ce/C0, Cg/C0, Ce/Cg and Φ based on θ=0.5°. (a) Φ - Ce, (b) Φ - f, (c) Φ - blocking, (d) Φ - Ce/C0, (e) Φ - Cg/C0, (f) Φ - Ce/Cg.

\[
y_{Ce} = 189.7517 - 1.1527x_1 - 0.9257x_2 + 0.000000046x_3 - 67.1349x_4 + 75.0747x_5
\]
The mean square error of train and test are: 0.73, 0.40
The coefficient of determination is: R^2=1.00
(3) Relationship between rim angle, f, m, blocking, spot and Ce/C0.
\[
y_{Ce/C0} = 3.2821 - 0.02007x_1 - 0.01610x_2 + 0.000000008x_3 - 1.1712x_4 + 1.3\times10^{-6}x_5
\]
The mean square error of train and test are: 0.00, 0.00
The coefficient of determination is: R^2=1.00
(4) Relationship between rim angle, f, m, blocking, spot and Cg/C0.
\[
y_{Cg/C0} = 3.3120 - 0.0201x_1 - 0.01616x_2 + 0.000000081x_3 - 1.1714x_4 + 1.3099x_5
\]
The mean square error of train and test are: 0.00, 0.00
Figure 7. Relationships between Ce, Cg, f, Blocking, Spot, Ce/C0, Cg/C0, Ce/Cg and $\phi$ based on $\theta=1.0^\circ$. (a) $\phi$ - Ce, (a') $\phi$ - Ce, (b) $\phi$ - f, (c) $\phi$ - blocking, (d) $\phi$ - Spot, (e) $\phi$ - Ce/C0, (f) $\phi$ - Ce/Cg.

The coefficient of determination is: $R^2=1.00$

(5) Relationship between rim angle, f, m, blocking, spot and Ce/Cg.

$$y_{Ce/C0} = 1.0008 - 0.000027x_1 - 0.0000058x_2 + 0.00000095x_3 - 0.00723x_4 - 0.00319x_5$$

The mean square error of train and test are: 0.00, 0.00

The coefficient of determination is: $R^2=1.00$

2. Function relationship based on $\theta=1.0^\circ$

(1) Relationship between rim angle, f, m, blocking, spot and Ce.

$$y_{Ce} = 79.0412 - 0.1449x_1 - 0.2097x_2 + 0.00000403x_3 - 49.5094x_4 + 49.6217x_5$$

The mean square error of train and test are: 0.04, 0.03

The coefficient of determination is: $R^2=1.00$

(2) Relationship between rim angle, f, m, blocking, spot and Cg.
\[ y_{c_8} = 80.3092 - 0.1435x_1 - 0.2100x_2 + 0.00000413x_3 - 49.7050x_4 + 47.7996x_5 \]

The mean square error of train and test are: 0.04, 0.03

The coefficient of determination is: \( R^2 = 1.00 \)

(3) Relationship between rim angle, \( f \), \( m \), blocking, spot and \( Ce/C_0 \).
\[ y_{c_e/C_0} = 2.7581 - 0.00508x_1 - 0.00732x_2 + 0.00000139x_3 - 1.7240x_4 + 1.6580x_5 \]

The mean square error of train and test are: 0.00, 0.00

The coefficient of determination is: \( R^2 = 1.00 \)

(4) Relationship between rim angle, \( f \), \( m \), blocking, spot and \( Cg/C_0 \).
\[ y_{c_g/C_0} = 2.8048 - 0.00506x_1 - 0.00735x_2 + 0.00000142x_3 - 1.7323x_4 + 1.6659x_5 \]

The mean square error of train and test are: 0.00, 0.00

The coefficient of determination is: \( R^2 = 1.00 \)

(5) Relationship between rim angle, \( f \), \( m \), blocking, spot and \( Ce/C_g \).
\[ y_{c_e/C_g} = 1.0008 - 0.00005x_1 - 0.00002x_2 + 0.000000112x_3 - 0.00378x_4 - 0.00625x_5 \]

The mean square error of train and test are: 0.00, 0.00

The coefficient of determination is: \( R^2 = 1.00 \)

5. Conclusions
The main objective of this work was to demonstrate how to use machine learning methods to derive empirical formulas and the functional relationships between parameters and dependent variables in the context of big data. In this paper, the data generated by a Secondary flat concentrators system is used in machine learning through machine learning training and testing methods, achieving very good learning and training effects, and the determination coefficient as an evaluation standard has reached 1.0, that is \( R^2 = 1.00 \). It provides a useful method and guidance for big data machine learning. This method can be used in all kinds of big data machine learning training.

Acknowledgements
This work is in partial fulfilment of the funded research program 51466016, which was funded by the National Natural Science Foundation of China.

References
[1] Cheng Z D, Zhao X R and He Y L 2018 Novel optical efficiency formulas for parabolic trough solar collectors: computing method and applications Appl. Energy 224 682-97 https://doi.org/10.1016/j.apenergy.2018.05.033.
[2] Wu S B and Wang C M 2018 Big data security framework based on encryption LNCS 11065 528-540 https://doi.org/10.1007/978-3-030-00012-7_48.
[3] Bode G, Thul S, Baranski M and Müller D Real-world application of machine-learning-based fault detection trained with experimental data Energy 198 (2020) 117323.
[4] Rodriguez S D and Rosengarten G 2015 Improving the concentration ratio of parabolic troughs using a second-stage flat mirror Appl. Energy 159 620-32.
[5] Wu S B, Wang C M, Cao H S, Jia X M 2018 Crime prediction using data mining and machine learning CENet 2018 pp 360-375 https://doi.org/10.1007/978-3-030-14680-1_40.
[6] Mason K, Reno M J, Blakely L, et al. 2020 A deep neural network approach for behind-the-meter residential PV size, tilt and azimuth estimation Sol. Energy 196 260-9 https://doi.org/10.1016/j.solener.2019.11.100.
[7] Song H and Liu X Q 2018 Data Science Technology and Application (Publishing House of Electronics Industry) p 89.
[8] Subasi A H 2020 Practical Machine Learning for Data Analysis Using Python (Academic Press) p 63.