Data-Driven Intelligent Recognition of Flatness Control Efficiency for Cold Rolling Mills

Xiaomin Zhou [1,2,*], Liqi Li [1], Xinglong Ma [1] and Tao Xu [2]

[1]   School of Mechanical Engineering, University of Science and Technology Beijing, Beijing 100083, China
[2]   Shunde Innovation School, University of Science and Technology Beijing, Foshan 528399, China
*   Correspondence: zhouxiaomin@ustb.edu.cn; Tel.: +86-1851-0257-595

Abstract: In the production process of strip tandem cold rolling mills, the flatness control system is important for improving the flatness quality. The control efficiency of actuators is a pivotal factor affecting the flatness control accuracy. At present, the data-driven methods to intelligently identify the flatness control efficiency have become a research hotspot. In this paper, a wavelet transform longitudinal denoising method, combined with a genetic algorithm (GA-WT), is proposed to handle the big noise of the measured data from each signal channel of the flatness meter, and Legendre orthogonal polynomial fitting is employed to extract the effective flatness features. Based on the preprocessed actual production data, the adaptive moment estimation (Adam) optimization algorithm is applied, to intelligently identify the flatness control efficiency. This paper takes the actual production data of a 1420 mm tandem cold mill as an example, to verify the performance of the new method. Compared with the control efficiency determined by the empirical method, the flatness residual MSE 0.035 is 5.4% lower. The test results indicate that the GA-WT-Legendre-Adam method can effectively reduce the noise, extract the flatness features, and achieve the intelligent determination of the flatness control efficiency.

Keywords: control efficiency; data noise reduction; Adam optimization; data-driven

1. Introduction

Cold-rolled strips are widely used in the fields of automobile manufacturing, aerospace, household appliance panels, and so on [1], because of their high geometric accuracy and great mechanical properties. With the development of technology, cold-rolled strips have become thinner and thinner, and the geometric shape and dimensional accuracy requirements are becoming higher and higher [2,3]. In order to improve the quality of the strips, modern rolling mills adopt an automatic gauge control system (AGC) to solve the gauge problem, and an automatic flatness control system (AFC) to solve the flatness problem. At present, the gauge control basically meets the requirements, but the flatness problem is becoming more and more complicated [4]. Due to the large number of actuators required for flatness control, and the strong coupling and hysteresis between the actuators [5], flatness control is still a great challenge in practice. The cold tandem rolling mill contains multi-stand flatness controls, and the final stand is a closed-loop feedback control. The control efficiency reflects the adjustment ability of the actuators to the flatness in a closed-loop feedback control. Distributing the flatness deviation to each actuator reasonably can improve the flatness control accuracy [6]. In actual production, the control efficiency of each actuator varies with different process parameters, especially, when the ratio of the width to thickness is larger, it is more unstable. The control efficiency is often obtained by an empirical method which has been used since setting up the factory. However, there are errors, inevitably, for specific working conditions, and the update speed is slow. Therefore, obtaining an accurate control efficiency is significant in the research on flatness control.

To the best of our knowledge, there are two main methods used to obtain the control efficiency: the experimental method and the finite element method [7]. The experimental

method obtains the control efficiency by changing the setting value of the actuator one by one to obtain the change of the flatness. Although it is more accurate, the high cost and the limited working conditions make it infeasible. Compared with the experimental method, the finite element simulation has the advantage of being more economical, having a wider application, and more flexible simulation of various rolling conditions [8]. However, the finite element method needs many assumptions in the analysis of the rolling process, which leads to errors between the calculation results and the actual situations. Meanwhile, the calculation process is slow, so it is not suitable for the requirements of online real-time control [9]. Thanks to a large amount of rolling process data stored by steel companies, a lot of information related to the production process can be revealed with the development of big data and sensor technology [10]. However, the use of the rolling process data is not sufficient. Therefore, based on the rolling process data, mining the information with the goal of analyzing the relationship between the flatness and the process, provides not only a new way to improve the flatness control accuracy but also a new approach for the intelligent recognition of the flatness control efficiency.

The actual rolling production data has the characteristics of noise, outliers, deviations, and so on. In order to calculate the control efficiency, it is necessary to analyze and process the data. Thus, it is extremely important to reduce the noise of the production data. Wavelet transform, a key noise reduction method at home and abroad, has the advantages of multi-resolution, fast operation speed and small storage space. In addition, it has been widely used in the fields of image compression, data denoising, and edge detection.

Y. Kim et al. proposed cepstrum-assisted empirical wavelet transform (CEWT) to decompose the signal in the fault diagnosis of planetary gear trains. The CEWT solved the problem of empirical wavelet transform (EWT), which required physical understanding to isolate fault-related signals and improve performance of fault diagnosis [11]. F. Dengand et al. proposed a new peak detection algorithm based on continuous wavelet transform and image segmentation, which has made progress in identifying weak peaks, overlapping peaks, and removing false peaks [12].

M. Zolfaghari found that the electricity production data in the prediction is non-stationary and non-linear, and traditional forecasting methods display a poor robustness. Therefore, they proposed an AWT-LSTM-RF hybrid model, combining adaptive wavelet transform (AWT), long short-term memory network (LSTM), and random forest (RF) algorithms for the prediction. The results showed that the hybrid model of AWT-LSTM-RF is better than the benchmark model. At the same time, the wavelet transform algorithm for the treatment of the input data can improve the predictive ability of RF [13]. Denoising the actual data under the wavelet transform alone leads to signal distortion, therefore, the threshold of wavelet coefficients can be optimized in combination with the genetic algorithm, to remove noise as much as possible while retaining the characteristics of the signal. The flatness meter has dozens of sensor channels and these sensor channels can measure the flatness along the width direction at the same time. Figure 1 shows the measurement signal of the strip flatness coil. Because of the noise interference in the actual signal, the characteristics of the flatness cannot be expressed directly. Therefore, data processing is required. Simply denoising each signal channel can only solve the noise problem of a single sensor channel, and the strip flatness represented by the width direction sensor still has no obvious strip flatness features. So it is necessary to fit in the width direction with a Legendre orthogonal polynomial to obtain the strip flatness represented by the width direction sensor signal. Based on the strip flatness features extracted after denoising, Legendre fitting, and the corresponding actual production process data, an optimized algorithm is used to intelligently identify the strip flatness control effect. The Adam optimization algorithm is an extension of the stochastic gradient descent method. It has the advantages of simplicity, directness, high efficiency, small memory, and handling sparse gradients in solving non-convex optimization problems.
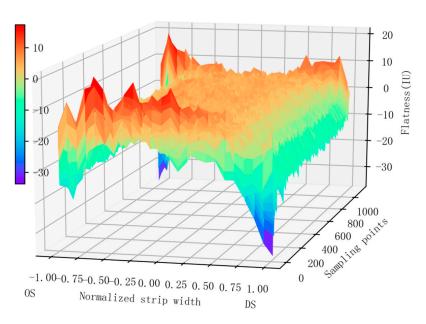
Figure 1. Three−dimensional strip flatness.

This work chooses the Adam optimization method to identify the flatness control efficiency. Based on the actual production process data of a tandem cold rolling mill, the influence of the different actuators on the flatness is analyzed. The GA-WT-Legendre-Adam intelligent identification method for flatness adjustment and control effects is proposed and compared with the control efficiency determined by the empirical method. The analysis results show that the method proposed in this paper can effectively reduce the influence of noise, and the flatness residual MSE calculated by the recognized flatness control efficiency is reduced by 5.4% compared with the empirical method. This noise reduction method can use the industrial data of online production to realize the online identification of the flatness control efficiency.

2. Flatness Closed-Loop Control System

A cold tandem rolling mill consists of a five-stands six-high UCM (Universal Crown Mill). Among them, the first four stands use a preset model to make preliminary adjustments to the thickness and flatness of the strip. The fifth stand uses a closed-loop feedback control strategy to adjust the flatness, which is the most critical part in the flatness control system. The flatness control system is shown in Figure 2. The flatness closed-loop feedback control system is established in the rolling stabilization stage. The closed-loop feedback control model calculates the adjustment of the actuator through the deviation between the measured flatness and the target flatness. Then a control signal is sent to the actuator and the new flatness deviation is detected at the same time. Continuous dynamic adjustments for the actuator in real time are essential for obtaining a stable and good flatness [14,15]. At present, the five-stands cold tandem rolling mill includes intermediate roll bending (IRB), intermediate roll shifting (IRS), work roll bending (WRB), tilting and segment cooling, and other actuators [16,17]. Among them, tilting can eliminate the first order flatness deviation. WRB, IRB, and IRS can eliminate the secondary and the fourth order flatness deviations. Elimination of the higher order flatness deviation requires segmented cooling [18]. The control performance of each type of actuator can be reflected on the control efficiency curve. The general segmented cooling of work rolls cannot be solved by formulas, and the response speed is slow. It is often used to eliminate the remaining flatness deviation after the adjustment of other actuators. Therefore, this paper only takes the influence of tilting, WRB, IRB, and IRS into account.
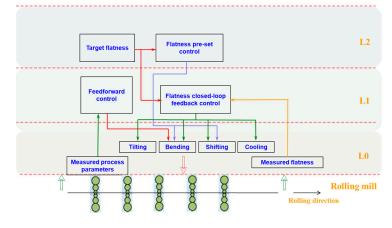
Figure 2. Flatness control system.

The basis of the flatness control is the control efficiency, which can effectively reflect the control performance of the actuators on the flatness. Because of its simplicity and fast calculation process, it can play a key role in a unit with a variety of control methods. The control efficiency can be optimized by making full use of the rolling production process data to obtain a more precise flatness control [19]. The control efficiency, described by the following Formula (1), means the change of the flatness deviation caused by the change of each unit of the actuator:

$$eff_{ij} = \frac{\Delta flat_{ij}}{\Delta Act_j} \tag{1}$$

where eff_{ij} is the actuator efficiency of the jth actuator in the ith segment of the width direction, ΔAct_j is the adjustment of the jth actuator and $\Delta flat_{ij}$ is the flatness change caused by the jth actuator in the ith segment of the width direction.

The multi-variable optimal flatness control algorithm, based on control efficiency, studies the effect of the actuator on the flatness deviation of each measurement section in the width direction. The control amount of each actuator can be solved in connection with the flatness deviation of each measurement section to get the optimal control for the entire system. Therefore, as long as the control efficiency of the flatness is obtained, the adjustment of each actuator under the optimal flatness control can be calculated using Formula (2), to guarantee the minimum value of the performance index J.

$$J = \sum_{i=1}^{n} \left[Tarflat_i - \left(\sum_{j=1}^{m} eff_{ij} \cdot \Delta Act_j \right) \right]^2 \tag{2}$$

$Tarflat_i$ is the target flatness deviation in the ith segment of the width direction, m is the number of flatness control technologies, and n is the number of flatness measurement channels in the width direction.

Among the required actuators, the response speed of the tilting is the fastest, which mainly eliminates the asymmetric flatness deviation, so the control strategy mainly refers to the control of the roll bending and roll shifting. Due to the large number of different control methods, various control goals and the coupling effects of them, choosing either the sequential or parallel calculation needs to be considered for adjustment in the flatness control. The cold tandem rolling mill that provides the data source for this paper adopts a sequential solution strategy, also known as a relay solution, which is based on the least squares principle of the priority sequence table. The execution mechanism is calculated hierarchically based on the sequence in the priority sequence table. The objective function of an executive agency can be expressed as

$$J_j = (Rflat_{ij} - \Delta Act_j \times eff_{ij})^2 \tag{3}$$

where J_j is the objective function of the *j*th actuator; $Rflat_{ij}$ is the remaining flatness deviation after the adjustment of the first $j-1$ actuators.

In order to minimize the objective function J_j, the partial derivative of the objective function is calculated, as in Equation (4), to obtain the adjustment of the *j*th actuator.

$$\frac{\partial J_j}{\partial eff_j(x)} = 0 \tag{4}$$

The priority is determined by the sensitivity of the actuator. Generally, it has the sequence of WRB, IRB, and IRS. The sequential solution strategy makes the best use of the actuators to eliminate the flatness deviation. When the front actuator cannot completely eliminate the flatness deviation, the remaining flatness deviation is eliminated by the latter actuator. Because this kind of control strategy calculates the control quantity of each actuator in sequence, according to the artificially prescribed sequence, there is no unsolvable situation. It successfully avoids the matrix solving process and the limit problem of the actuator, which are not suitable for online and real-time use. From the above calculation formula, it is obvious that accurate control efficiency is the core content of the flatness closed-loop control system, which is different from the empirical method that has a long update cycle and is not suitable for specific working conditions. The data-driven method can dynamically obtain the flatness control functions of different steel specifications from the data more accurately and with a greater flexibility, thereby the accuracy of the closed-loop feedback control of the current steel specifications can be improved.

3. Establishment and Test of Noise Reduction Model

Using actual production data to establish a data-driven model for the intelligent recognition of the flatness control efficiency can reflect the real performance of the flatness control mechanism, and realize the adjustment of the control efficiency with the change of working conditions. However, the data actually collected at the production site contains noise, has a wide range of fluctuations, and thus the true situation of the data is obscured. It is necessary first to extract meaningful flatness features to establish a data-driven intelligent identification model of flatness control efficiency.

Wavelet transform is a time-frequency localization analysis method, in which the wavelet window area is constant and the time and frequency domain window ranges are variable. It solves the shortcoming of the short-time Fourier transform, which adopts a fixed sliding window function and cannot change the time resolution. Wavelet transform uses a short-time window when analyzing high-frequency signals and a long-time window when analyzing low-frequency signals. This achieves the adaptive change of the time-frequency window and has the abilities of local analysis and refinement.

The wavelet of continuous wavelet transform is a short-duration time function. Suppose that $\psi(t)$ is an integrable function and $\psi(t) \in H^2(R)$. $\psi(t)$ is a basic wavelet or wavelet mother function, if the Fourier transform satisfies:

$$C_\psi = \int_{-\infty}^{+\infty} \left| \hat{\psi}(\omega) \right|^2 w^{-1} d\omega < \infty \tag{5}$$

By shifting and scaling the wavelet $\psi(t)$, a set of wavelet basis functions $\{\psi_{a,b}(t)\}$ can be obtained:

$$\psi_{a,b}(t) = |a|^{-\frac{1}{2}} \psi\left(\frac{t-b}{a}\right) \tag{6}$$

where $a, b \in R$ and $a > 0$. b is the translation factor indicating the position where the function is translated along the t axis and a is the scale factor of a wavelet basis function.

For a signal $f(t) \in L^2(R)$, its continuous wavelet transform is:

$$W_f(a,b) = \langle f, \psi_{a,b} \rangle = |a|^{-\frac{1}{2}} \int_R f(t) \psi\left(\frac{t-b}{a}\right) dt \tag{7}$$

where $W_f(a, b)$ is the coefficient after wavelet transform of the signal $f(t)$. Through the inverse transform of $W_f(a, b)$, we can obtain:

$$f(t) = \frac{1}{C_\psi} \int\limits_{R+} \int\limits_{R} \frac{1}{a^2} W_f(a, b) \psi \frac{t - b}{a} da db \tag{8}$$

In practical applications, the wavelet transform needs to be discretized. Generally, the scale factor and translation factor of the continuous wavelet transform are discretized. The scale factor is $a = a_0^j$ and the translation factor is $b = k b_0 a_0^j$, $j, k \in Z$. The step value is a fixed value other than 1, and usually $a_0 > 1$, then the function can be written as

$$\psi_{j,k}(t) = a_0^{-\frac{j}{2}} \psi \left(\frac{t - k b_0 a_0^j}{a_0^j} \right) = a_0^{-\frac{j}{2}} \psi \left(\int\limits_{-\infty}^{+\infty} a_0^{-j} t - k b_0 \right) \tag{9}$$

The corresponding discrete wavelet transform can be written as

$$\left(W_\psi f \right)(a, b) = \langle f, \psi_{a,b} \rangle = a^{-\frac{j}{2}} \int\limits_{-\infty}^{+\infty} f(t) \overline{\psi \left(a_0^{-j} t - k b_0 \right)} dt \tag{10}$$

Because the wavelet transform has the characteristics of multi-resolution, it can effectively remove the noise while preserving the characteristic information of the original signal as much as possible. The formula for wavelet denoising is as follows:

$$s(n) = f(n) + \sigma e(n) \tag{11}$$

where $s(n)$ is the actual noisy signal, $f(n)$ is the effective signal, n is the sampling time, σ is the noise intensity and $e(n)$ is the noise component.

Wavelet denoising requires processes such as multi-layer decomposition of noisy signals, threshold processing of wavelet coefficients, and reconstruction of the wavelet inverse transform. The more decomposition layers of a noisy signal there are, the easier it is to eliminate the noise signal, but the error is greater after reconstruction. So it is very important to choose an appropriate number of decomposition layers. The appropriate number of decomposition layers can be calculated using Formula (12):

$$max\_level = \log_2 \left(l_{data} / l_{filter} \right) \tag{12}$$

where max_level is the number of decomposition levels, l_{data} is the length of the data, and l_{filter} is the length of the filter.

The processing of the wavelet coefficient threshold is a key factor to improve the effect of wavelet denoising. Common threshold functions include the hard threshold and the soft threshold. Since the soft threshold function can shrink large coefficients, it can reduce the discontinuity problems in the hard threshold noise reduction process. At the same time, it has a good adaptability and an excellent noise reduction effect. Therefore, the soft threshold function is selected in this paper as follows:

$$w_\lambda = \begin{cases} [\text{sgn}(w)](|w| - \lambda) & |w| \geq \lambda \\ 0 & |w| < \lambda \end{cases} \tag{13}$$

where w is the wavelet coefficient, λ is the threshold, and w_λ is the wavelet coefficient after threshold processing.

From (13), it can be found that, how to choose an appropriate threshold is the key to wavelet denoising. In order to obtain the optimal wavelet coefficients, this paper combines a genetic algorithm to optimize the threshold. After the wavelet coefficients are decomposed by the wavelet transform, the genetic algorithm is used to select an appropriate threshold,

remove the noise in the high frequency, and retain the useful information. Then the wavelet inverse transform is reconstructed to obtain the noise-reduced data. Figure 3 shows the noise reduction process of GA-WT (Genetic Algorithm-Wavelet).
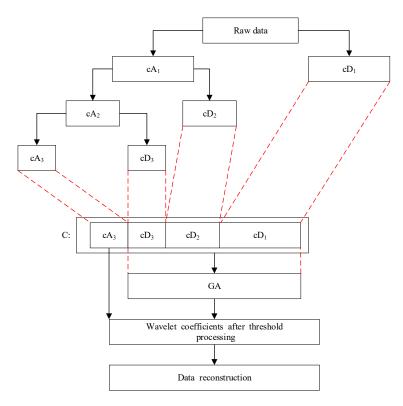

Figure 3. GA-WT noise reduction process.

To establish the GA-WT model, it is necessary to select an appropriate wavelet basis function. In this paper, the db8 basis function is selected, and the number of decomposition layers calculated by Formula (12) is six. The low-frequency wavelet coefficients are reserved, and the optimal threshold is selected by the genetic algorithm for high-frequency wavelet coefficients. The initial population of the genetic algorithm is 200, the crossover probability is 0.8, the number of iterations is 50, and the mutation probability is 0.003.

In order to verify the performance of the model, a standard data set is established. White noise is added to obtain simulation data. The commonly used EEMD (Ensemble Empirical Mode Decomposition) noise reduction method, and the wavelet transform method without genetic algorithm optimization, are compared with the GA-WT method proposed in this paper. The Signal to Noise Ratio (SNR) and Root Mean Square Error (RMSE) are used to measure the noise reduction performances of these three methods. The SNR is defined as the ratio of the effective component of the reconstructed signal to the noise. A larger ratio indicates a better noise reduction performance. The calculation formula is:

$$\text{SNR} = 10\lg\left(\frac{\sum\limits_{n=1}^{N} x^2(n)}{\sum\limits_{n=1}^{N} [x(n) - y(n)]^2}\right) \tag{14}$$

where $y(n)$ is the original signal, $x(n)$ is the signal after denoising, and N is the signal length.

The RMSE reflects the similarity between the reconstructed signal and the original signal. A smaller value of RMSE indicates that the reconstructed signal is closer to the original signal. It is defined as:

$$\text{RMSE} = \sqrt{\frac{\sum\limits_{n=1}^{N}[x(n)-y(n)]^2}{N}} \tag{15}$$

According to the GA-WT model, the appropriate wavelet basis function needs to be selected firstly. This paper chooses the db8 basis function, and the number of decomposition layers is 6, from Formula (12). The low frequency part is retained, and the genetic algorithm is used to select the optimal soft threshold for the high frequency part. Figure 4 shows the comparison of these three methods before and after denoising. It is obvious that these three methods can basically restore the signal trend. In detail, the wavelet transform based on the genetic algorithm performs better and reconstructs the original data more precisely. The unoptimized wavelet transform and EEMD fluctuate greatly, and the noise is not removed as much as possible. Even some valid information is deleted at the same time.
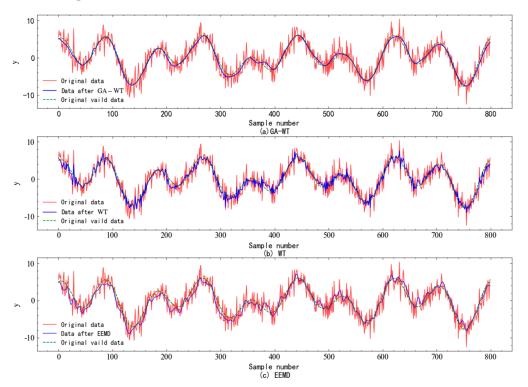

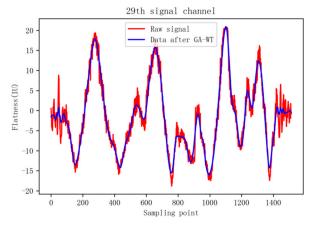**Figure 4.** Comparison of the three methods.

The SNR comparison of these three methods is shown in Table 1. The SNR of GA-WT is 18.98 db, which is higher than the 14.41 db of the unoptimized wavelet transform and 12.42 db of EEMD. At the same time, the RMSE of GA-WT is 0.39 IU. This is lower than 0.67 IU of the wavelet transform without optimization and 0.87 IU of EEMD. It shows that the noise reduction performance of GA-WT is the best, so the GA-WT method is chosen in this paper to reduce the noise of the measured flatness data.
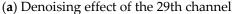
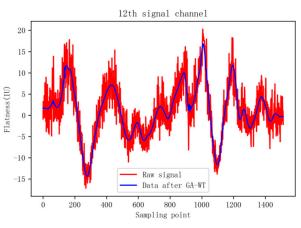**Table 1.** Comparison of the three methods.

Method	SNR/db	RMSE/IU
GA-WT	18.98	0.39
WT	14.41	0.67
EEMD	12.42	0.87

4. Intelligent Recognition of Flatness Control Efficiency Based on Data-Driven Model

4.1. Data Preprocessing

Modern rolling mills are equipped with a variety of actuators with different control capabilities. This paper mainly studies the control effects of tilting, WRB, IRB, and IRS. The experimental data is the actual data produced by a 1420 mm UCM cold tandem rolling mill. The steel grade is AQ0510B1, the strip width is 1150 mm, and the exit thickness is 0.8 mm. A total of 1514 sampling points are selected in the stable stage of the rolling process, and each sampling point includes variables such as 40 sections of flatness measurement data, rolling force, tilting, roll bending force, and roll shifting, etc. After a preliminary observation of the data, it is found that the width direction of the flatness fluctuates sharply, so it is difficult to extract the effective flatness features. There exists noise in the rolling direction of each signal channel, so GA-WT is adopted for each signal channel in the rolling direction of the strip to obtain the flatness after noise reduction. The denoising flatness data of some certain signal channels, such as the 29th channel, which is near the strip edge channel, and the 12th channel, which is near the strip center channel, are shown in Figure 5.
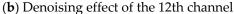

(**a**) Denoising effect of the 29th channel

(**b**) Denoising effect of the 12th channel

Figure 5. Denoising effect of certain signal channels.

After the noise reduction, the fluctuation of the flatness composed of all signal channels in the width direction at the same time, is reduced. However, the flatness features are still not obvious. In general, Legendre polynomials or Chebyshev polynomials are used to describe the shape defects. The integral value of the even-degree polynomial of the Chebyshev polynomial is not zero along the strip width direction, which cannot satisfy the self-equilibrium condition of the residual stress of the strip [20]. The Legendre polynomial can satisfy the self-equilibrium condition of the residual stress in the plate width direction.

Therefore, it is necessary to perform Legendre orthogonal polynomial fitting on the flatness in the width direction, which can be effective to extract the flatness features of the strip. Due to the limitation of the control ability of the actuator, the fourth-order Legendre orthogonal polynomial is used for fitting, which can remove the influence of higher-order terms. The Legendre orthogonal polynomial is shown as:

$$y_1 = \pm x \tag{16}$$

$$y_2 = \pm \left(\frac{3}{2} x^2 - \frac{1}{2} \right) \tag{17}$$

$$y_3 = \pm \frac{1}{2} \left(5x^3 - 3x \right) \tag{18}$$

$$y_4 = \pm \frac{1}{8} \left(35x^4 - 30x^2 + 3 \right) \tag{19}$$

where x is the sampling points in the width direction of the strip, y_1 is the first order flatness, y_2 is the second order flatness, y_3 is the third order flatness, and y_4 is the fourth order flatness.

Figure 6 shows that the original signal width direction is severely oscillating, which masks the characteristics of the flatness shape. After longitudinal denoising, the oscillation amplitude is weakened, but there are still fluctuations. Then after the Legendre orthogonal polynomial fitting, the data transition is smoother, in line with the current control performance of the actuator. At the same time, the influence of the high-order shape deviation is excluded, so the flatness control efficiency can be calculated by a Legendre orthogonal polynomial.
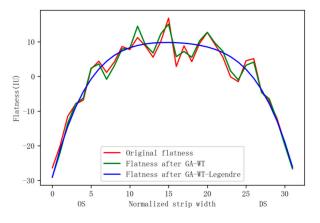

Figure 6. Comparison of the three flatnesses.

4.2. Optimization Method

Gradient descent is one of the most commonly used methods to optimize model parameters, and Adam is an optimization method that can be used to improve the gradient descent method. The Adam algorithm has a fast calculation speed, uses the momentum method to select the optimization direction, and uses the RMSProp algorithm to select the learning rate [21]. With the idea of moment estimation, the Adam optimization algorithm dynamically adjusts the learning rate by modifying the first-order and second-order moment estimation of the gradient. The Adam algorithm calculates the current time gradient as follows:

$$g_t = \nabla J(w) \tag{20}$$

where $J(w)$ is the objective function to be optimized.

The first moment estimation at the current moment is shown as follows:

$$M_t = \beta_1 \cdot M_{t-1} + (1 - \beta_1) \cdot g_t \tag{21}$$

where t is the time step, M_{t-1} is the first moment estimate at the previous moment, and $\beta_1 \in [0,1)$ is the hyperparameter, which is the exponential decay rate of the instantaneous estimation.

The second moment at the current moment is estimated as follows:

$$V_t = \beta_2 \cdot V_{t-1} + (1 - \beta_2) \cdot g_t{}^2 \qquad (22)$$

where V_{t-1} is the second moment estimate at the previous moment and $\beta_2 \in [0,1)$ is the hyperparameter, which is the exponential decay rate of the instantaneous estimation.

The second-order moment estimation is actually calculated according to the average value of the exponential movement. The second-order moment estimation is initialized to a 0 vector, that is, at the beginning of the iteration, V_t is close to 0, which requires the estimation of the second-order moment. The moment uses the exponentially averaged original moment of the gradient squared at the decay rate β_2, that is, the centerless variance. Let $g_1, g_2 \ldots, g_t$ be the gradient of the subsequent time steps, then Equation (22) can be written as:

$$V_t = (1 - \beta_2) \sum_{i=1}^{t} \beta_2^{t-i} \cdot g_i{}^2 \qquad (23)$$

Then the relationship between the expected value $\mathbb{E}[V_t]$ of the second-order moment estimation at the time step and the expected value of the second-order moment $\mathbb{E}[g_t^2]$ is the expectation on both sides of Equation (23):

$$
\begin{aligned}
\mathbb{E}[V_t] &= \mathbb{E}\left[(1 - \beta_2) \sum_{i=1}^{t} \beta_2^{t-i} \cdot g_i{}^2\right] \\
&= \mathbb{E}[g_t^2] \cdot (1 - \beta_2) \sum_{i=1}^{t} \beta_2^{t-i} + \zeta \\
&= \mathbb{E}[g_t^2] \cdot (1 - \beta_2{}^t) + \zeta
\end{aligned} \qquad (24)
$$

When the true second-order moment $\mathbb{E}[g_t^2]$ is a constant value, ζ is 0. Otherwise, the exponential decay rate β_2 is usually selected by adding the past moment to the current moment gradient with a small weight, so ζ is a minimum value. In the case of sparse gradients, in order to reliably estimate the second moment, it is necessary to choose a smaller β_2 to average many gradients. Since β_2 is small, the lack of an initialization bias correction will lead to a much larger initial step size.

Then the deviation correction of the second-order moment estimation V_t is shown as follows:

$$\hat{V}_t = \frac{V_t}{1 - \beta_2^t} \qquad (25)$$

In the same way, the deviation correction of the first-order moment estimation M_t is shown in Equation (26):

$$\hat{M}_t = \frac{M_t}{1 - \beta_1^t} \qquad (26)$$

Update the control amount by Formula (27):

$$w_t = w_{t-1} - \eta \frac{\hat{M}_t}{sqr(\hat{V}_t) + \varepsilon} \qquad (27)$$

where $\eta \in (0,1]$ is the learning rate and ε is a very small constant to avoid the situation where the second moment estimation V_t is zero during the calculation process.

4.3. Intelligent Recognition of Flatness Control Efficiency

The objective function that defines the intelligent recognition of the flatness control efficiency is shown in Equation (28):
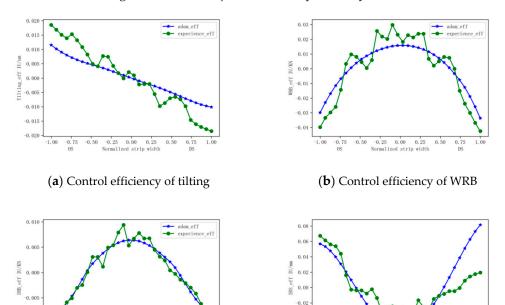
$$J(w) = \frac{\sum\limits_{i=1}^{n} (Tarflat_i - \sum\limits_{j=1}^{4} \Delta Act_j \cdot eff_{ij})^2}{2n} \tag{28}$$

In order to minimize the objective function $J(w)$, calculate its gradient at the current position:

$$g_t = \frac{\partial J(w)}{\partial \Delta Act} = -\frac{\begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \end{bmatrix} \cdot (Tarflat - \sum\limits_{j=1}^{4} \Delta Act_j \cdot eff_j)^{\mathrm{T}}}{n} \tag{29}$$

where w_1, w_2, w_3, w_4 are respectively the adjustment amounts of tilting, WRB, IRB, and IRS.

Figure 7 shows the results using the Adam algorithm and the empirical method. The results of the two methods both show that the tilting has the adjustment ability of the asymmetric term; and the work roll bending roll, the middle roll bending roll, and the middle roll shifting roll have the adjustment ability of the symmetric term.

(**a**) Control efficiency of tilting

(**b**) Control efficiency of WRB

(**c**) Control efficiency of IRB

(**d**) Control efficiency of IRS

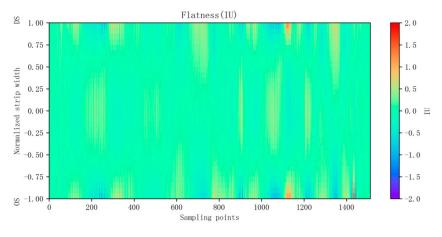**Figure 7.** Comparison of the two control efficiencies.

Substituting the control efficiency of the two methods into the control system, calculate the flatness residual after flatness control and defines the MSE of the flatness residual as shown in Equation (30):
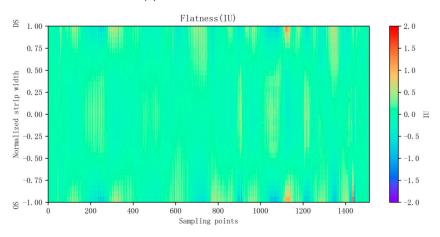
$$MSE = \frac{\sum\limits_{k=1}^{m} \left(\frac{\sum\limits_{i=1}^{n} \left(Tarflat_i - \sum\limits_{j=1}^{4} w_j \cdot eff_{ij} \right)^2}{n} \right)}{m} \tag{30}$$

where m is the number of sampling points with length, n is the number of flatness meter sensors in the width direction, and the other parameters remain unchanged.

The flatness residuals obtained by the two methods are shown in Figure 8. It can be seen that the flatness residuals calculated by the two methods are very small, especially, the control efficiency of Adam's method has smaller flatness residuals at certain sampling points. The MSEs of the flatness residuals are shown in Table 2. It can be seen that the MSE of the flatness deviation calculated by the control efficiency of the Adam algorithm is 0.035, and the MSE of the flatness deviation calculated by the control efficiency of the empirical method is 0.037. The MSE of the proposed method is reduced by 5.4% compared to the empirical method, so the regulation effect based on the Adam algorithm is more accurate.

(**a**) Error under Adam method

(**b**) Error under the empirical method

Figure 8. Error under the two methods.

Table 2. MSE under the two methods.

Method	MSE
Adam	0.035
Empirical	0.037

5. Conclusions

(1) This paper employs a GA-WT denoising method to deal with the large measurement noise in each signal channel of the flatness meter. Compared with the unprocessed wavelet transform, the SNR of GA-WT is increased by 31.7% and the RMSE is reduced by

41.8%. Compared with EEMD, the SNR of GA-WT is increased by 52.8% and the RMSE is reduced by 55.2%.

(2) GA-WT is used to perform longitudinal denoising in the strip length direction for each signal channel of the flatness meter. Legendre orthogonal polynomial fitting is used on all the flatness signal channels in the width direction, after longitudinal denoising. The method proposed can extract the flatness features, and reflect the real flatness, which can be controlled by the flatness control actuators.

(3) Based on the processed flatness data and the adjustment of the actuator, compared with the empirical method, the control efficiency calculated by the Adam algorithm is 0.035. This is 5.4% lower than the empirical method, which indicates that the GA-WT-Legendre-Adam method can effectively extract flatness features and obtain a more accurate control efficiency.

(4) Production data is characterized by high noise and low quality. In the future, noise reduction and calculation accuracies will be improved further by improving the quality of the actual production data.

Author Contributions: Conceptualization, X.Z.; methodology, X.Z. and L.L.; software, X.Z.; validation, X.Z., L.L. and X.M.; formal analysis, X.Z.; investigation, L.L. and T.X.; resources, X.Z.; data curation, X.Z.; writing—original draft preparation, X.Z.; writing—review and editing, X.Z. and L.L.; visualization, X.Z.; supervision, X.M.; project administration, X.Z. and T.X.; funding acquisition, X.Z. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Due to the nature of this research, participants of this study did not agree for their data to be shared publicly, so supporting data is not available.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jia, D.; Li, S.; Mi, N.; Shen, Z.; Li, X. Core technology of process control system for 1250 mm five stands 6-high tandem cold mill. *Steel Roll.* **2020**, *37*, 71–76.
2. Huang, C.; Zhao, M. Research progress on cold-rolled strip shape recognition and control. *J. Iron Steel Res.* **2013**, *25*, 1–7.
3. Liu, L.; Han, J.; He, Y. Development of cold rolling flatness control theory. *J. Iron Steel Res.* **1997**, *9*, 51–54.
4. Wang, P.; Zhang, D.; Liu, J.; Wang, J.; Yu, X. Analysis and improvement for flatness control system of 1450 tandem cold mill. *China Metall.* **2009**, *19*, 31–35.
5. Li, X.; Wang, Q.; Zhang, Y.; Zhang, D. Present status and future prospects of strip flatness control based on elastic-plastic finite element. *Steel Roll.* **2020**, *37*, 1–11.
6. Song, L.; Shen, M.; Yang, L.; Liu, J.; Wang, J.; Chen, X. Shape control dimensionality reduction efficiency inherited regulation method of cold rolling wide strip. *Iron Steel* **2016**, *51*, 70–75.
7. Huang, L.; Chen, X.; Zhang, Q.; Chen, S.; Wei, C.; Wu, B.; Wang, J. Study on the Flatness Controlling Actuator's Efficiency of Cold Rolling Mill. *Metall. Equip.* **2000**, *1*, 4–7.
8. Liu, X. The Simulation Research of Shape Control Efficiency Function on Cold Rolling Mill. Master's Thesis, Yanshan University, Qinhuangdao, China, 2015.
9. Wang, H. Model Optimization and Simulation Software Development of Flatness Control System for Cold Rolled Strip. Master's Thesis, Yanshan University, Qinhuangdao, China, 2020.
10. Wang, X.; Yan, Y. Promoting the technical progress of steel industry with intelligent manufacturing. *Metall. Autom.* **2019**, *43*, 1–5.
11. Kim, Y.; Ha, J.M.; Na, K.; Park, J.; Youn, B.D. Cepstrum-assisted empirical wavelet transform (CEWT)-based improved demodulation analysis for fault diagnostics of planetary gearboxes. *Measurement* **2021**, *183*, 109796. [CrossRef]
12. Deng, F.; Li, H.; Wang, R.; Yue, H.; Zhao, Z.; Duan, Y. An improved peak detection algorithm in mass spectra combining wavelet transform and image segmentation. *Int. J. Mass Spectrom.* **2021**, *465*, 116601. [CrossRef]
13. Zolfaghari, M.; Golabi, M.R. Modeling and predicting the electricity production in hydropower using conjunction of wavelet transform, long short-term memory and random forest models. *Renew. Energy* **2022**, *170*, 1367–1381. [CrossRef]
14. Liang, X.; Xu, J.; Wang, G.; Liu, X. Close loop feedback control principle of cold rolling strip flatness and its application effect. *Metall. Autom.* **2006**, *30*, 36–39.

15. He, X.; Wang, C. Study on closed-loop shape control strategy of cold colling. *Metall. Autom.* **2018**, *42*, 44–48.
16. Pin, G.; Francesconi, V.; Cuzzola, F.A.; Parisini, T. Adaptive task-space metal strip-flatness control in cold multi-roll mill stands. *J. Process Control* **2013**, *23*, 108–119. [CrossRef]
17. Bemporad, A.; Bernardini, D.; Cuzzola, F.A.; Spinelli, A. Optimization-based automatic flatness control in cold tandem rolling. *J. Process Control* **2010**, *20*, 396–407. [CrossRef]
18. Wang, X.; Yang, Q. Research on Asymmetric Shape Control Character and Generalized Overall Shape Control Strategy for Universal Crown Mill. *Chin. J. Mech. Eng.* **2012**, *48*, 58–65. [CrossRef]
19. Sun, J.; Shan, P.; Peng, W.; Zhang, D. Cold-rolled flatness actuator efficiency coefficient obtaining based on data noise reduction. *Iron Steel* **2021**, *56*, 67–74.
20. Peng, Y. Theoretical Studies and Engineering Application of Shape Preset Control for HC Cold Mill Based on Strip Element Method. Ph.D. Thesis, Yanshan University, Qinhuangdao, China, 2000.
21. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. In Proceedings of the 3rd International Conference for Learning Representations, San Diego, CA, USA, 7–9 May 2015.