



# Article A Novel Principal Component Analysis Integrating Long Short-Term Memory Network and Its Application in Productivity Prediction of Cutter Suction Dredgers

Ke Yang, Jun-Lang Yuan, Ting Xiong, Bin Wang 💿 and Shi-Dong Fan \*

School of Energy and Power Engineering, Wuhan University of Technology, Wuhan 430063, China; k\_yang@whut.edu.cn (K.Y.); junlangyuan@whut.edu.cn (J.-L.Y.); cxj\_xt@whut.edu.cn (T.X.); wang\_bin@whut.edu.cn (B.W.)

\* Correspondence: sdfan@whut.edu.cn

Abstract: Dredging is a basic construction for waterway improvement, harbor basin maintenance, land reclamation, environmental protection dredging, and deep-sea mining. The dredging process of cutter suction dredgers is so complex that the operational data show strong characteristics of dynamic, nonlinearity, and time delay, which make it difficult to predict the productivity accurately via basic principles models. In this paper, we propose a novel integrating PCA-LSTM model to improve the productivity prediction of cutter suction dredger. Firstly, multiple variables are reduced in dimension and selected by PCA method based on the working mechanism of cutter suction dredger. Then the productivity is predicted via mud concentration in long short-term memory network with relevant operational time-series data. Finally, the proposed method is successfully applied to an actual case study in China. Also, it performs well in the cross-validation and comparative study for several important characteristics: (i) it involves the operational parameters based on the mechanism analysis; and (ii) it is a deep-learning-based approach that can deal with operation series data with a special memory mechanism. This study provides a heuristic idea for integrating the data-driven method and supervision of human knowledge for application in practical engineering.

**Keywords:** operational data; productivity prediction; cutter suction dredger; long short-memory network; working mechanism

## 1. Introduction

Marine-based transportation has always played a critical role in the national economics of China [1], while rivers may suffer from sediment accumulation that obstructs riverways and reduces the carrying activity [2,3]. Cutter suction dredgers are common and useful machines that can remove the mud deposited at the bottom of water and keep transportation routes in good condition [4]. Dredging productivity is one of the most important indexes to evaluate the dredging performance, which is affected by many factors such as soil properties, the power of the pump, the cutter structural parameters, and so on [5]. The process of sand being cut into a mixture of mud and water by a rotating cutter is very complicated. Most of the parameters are dynamically influenced by the uncertain working environment and human operation [6]. Due to the limitations of dredging technology, there are indeed some obstacles for parameters monitoring and real-time prediction, which is a challenge for constructing digital models to describe this process and dredging productivity accurately [7].

Due to the developed sensor technology, more operational data have been gained for analyzing dredging performance. In the literature review, machine learning methods were recently adopted to model the complex and dynamical construction process of CSD (cutter suction dredger) for their excellent learning and mining ability [8]. Generally, the learning-based prediction models can be mainly divided into two main types based



Citation: Yang, K.; Yuan, J.-L.; Xiong, T.; Wang, B.; Fan, S.-D. A Novel Principal Component Analysis Integrating Long Short-Term Memory Network and Its Application in Productivity Prediction of Cutter Suction Dredgers. *Appl. Sci.* 2021, *11*, 8159. https://doi.org/10.3390/ app11178159

Academic Editor: Dimitris N. Pagonis

Received: 22 July 2021 Accepted: 30 August 2021 Published: 2 September 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). on the depth of structure: shallow learning models and deep learning models [9]. The shallow learning methods mainly cover neural-network-based methods such as RBF (radial basis function), ELMs (extreme learning machines), and SVM (support vector machine). As traditional learning models used in productivity prediction, Wang et al. adopted an RBF neural network to deal with the different working conditions and established an accurate nonlinear mathematical model for the instantaneous output prediction with control variables [10]. Guan et al. developed the model of cutter operation parameters using improved ELMs to simulate and predict the productivity distribution in actual construction [11]. Yang et al. predicted the cutter suction dredger production with a double hidden layer BP neural network [12].

The prediction methods on deep learning mainly include: DNN (deep neural network), DBN (deep belief network), CNN (convolutional neural network network), and RNN (recurrent neural network network). DNNs are adopted through multiple auto-encoder (AE) or denoising auto-encoder (DAE) stacking networks, wherein high-dimensional input data are extracted from classless data as the distribution of the original data is represented by deep learning neural networks [13]. Wang et al. developed DNN models for production forecasting wherein the data-driven method was well used to deal with hydraulic fractures and intrinsic complexity [14]. DNN architecture transfers sigmoid function to ReLu and maxout with the purpose of overcoming gradient disappearance, but requires small batch training, which leads to over-fitting and local optimal solution problems. DBN is also a deep network that is stacked with multiple Restricted Boltzman Machines (RBMs) and a classification layer or regression layer. Xu et al. designed a DBN-based model to approximate the function type coefficients of a state-dependent autoregressive model in nonlinear system and realize predictive control [15]. Hu et al. adopted DBN to extract deep hidden features behind the monitoring signals and predict the bearing remaining useful life [16]. Researchers have improved DBN by combining a feed-forward neural network (FNN) to make prediction more accurate [17]. Zhang proposed a multi-target DBN collection method in which the output of multiple DBNs has a certain weight that reveals the final output of the network set; this method performed well on NASA aero-engine data [18]. Furthermore, convolutional neural networks (CNNs) are greatly developed with the excellent characteristics of parameter sharing and spatial pooling, which make it more advantageous in computing speed and accuracy [19]. However, all these ML methods have a limitation in situations that involve a time-series input.

Then recurrent neural networks (RNNs) are proposed with adding a twist where the output from previous time step is fed as input to the current step. The most important feature of RNN is that the hidden state can remember all information calculated from the previous sequence [20]. Thus, it can generate output through prior input (the past memory) and learning in training. It is common to complete the parameter learning of recurrent neural networks by learning over back-propagation, wherein the error is passed forward step by step in the reverse order of time. In [21], a learning-based method is applied to improve the RNN training process while the number of prediction time-steps increases. However, RNNs still suffer from the long-term dependencies, and long short-term memory (LSTM) fills the gap by setting a gate control unit that can choose and keep some useful information in the long-term sequential data. Being different from the traditional RNN, the model is trained by both the stored information of last time step and new input of the current moment, which enhances the prediction accuracy and stability greatly [22].

However, for the practical application in CSD, it is equally significant to analyze the interrelated influencing factors as the productivity prediction. LSTM lacks effective processing for high-dimensional characteristics of large-scale data; it should be used integrating with other methods. Principal component analysis (PCA) is one of the most widely used algorithms for feature reduction, which reconstructs the main k-dimensional feature based on the original n-dimensional feature by deep learning. While PCA is a pure data-driven method that cannot consider the casual relationship and correlation between variables, the procedure of variable analysis based on a working mechanism and human experience is necessary. Yang et al. described a HEPCA model, which made variables supplement based on expert knowledge after the PCA process and generate a more accurate input for the predictive model [23].

Therefore, combining the advantages and characteristics of the different methods described above, this paper presents the long short-term memory integrating principal component analysis model (PCA-LSTM) to predict productivity using the monitoring sensors data. The PCA-LSTM is strutted into four phases. In the first phase, monitoring sensors are analyzed to select related variables according to the working mechanism and knowledge. In the second phase, PCA method is applied to extract the deep feature from the high-dimension dataset and to obtain the correlation of variables. In the third phase, a prediction model is built and trained by the LSTM network. Finally, cross validation and comparative analysis are conducted with a generated model from "Chang Shi 10" in China.

## 2. Preliminaries

In this section, the related preliminaries regarding PCA and LSTM will be introduced briefly on the basis of the practical application in this study.

## 2.1. Principal Components Analysis (PCA)

PCA is an important technique that can transform multiple variables into a few main components (comprehensive variables) by means of dimensional reduction, increasing interpretability while minimizing information loss [24]. These main components are usually expressed as linear combinations of the original variables, which can represent most of the information of the whole dataset.

For original data  $X = (x_1, x_2, ..., x_i, ..., x_n)$  and  $X \in \mathbb{R}^{k \times n}$ , we can get the covariance matrix  $C_x$ :

$$C_{\mathbf{x}} = E[(\mathbf{X} - E[\mathbf{X}])(\mathbf{X} - E[\mathbf{X}])^{\mathrm{T}}]$$
(1)

After centralizing the data, the mean function E[X] is zero and:

$$C_{\rm x} = \frac{1}{n} X X^{\rm T} \tag{2}$$

Assuming that there is a matrix  $P(P \in \mathbb{R}^{k'*k})$ , through which we can transform the original sample data matrix X into a dimensionality-reduced matrix  $Y(Y \in \mathbb{R}^{k'*n})$ :

$$Y = PX \tag{3}$$

Then the original data dimension is successfully reduced from k to k', wherein the first k' principal components can explain most of the variance.

For matrix *Y*, its covariance matrix can be expressed through original matrix *X* as:

$$\boldsymbol{C}_{y} = \frac{1}{n} \boldsymbol{Y} \boldsymbol{Y}^{T} = \frac{1}{n} (\boldsymbol{P} \boldsymbol{X}) (\boldsymbol{P} \boldsymbol{X})^{T} = \frac{1}{n} \boldsymbol{P} \boldsymbol{X} \boldsymbol{X}^{T} \boldsymbol{P}^{T} = \boldsymbol{P} \frac{1}{n} \boldsymbol{X} \boldsymbol{X}^{T} \boldsymbol{P}^{T} = \boldsymbol{P} \boldsymbol{C}_{x} \boldsymbol{P}^{T}$$
(4)

It is obvious from Equation (4) that  $C_x$  is guaranteed to be a non-negative definite matrix and thus is diagonalizable by some unitary matrix. Then the objective optimization is transformed to find an orthonormal transformation matrix P. Normally, we can use eigenvalue decomposition or singular value decomposition to solve the matrix P, and the first k'-dimensional new features corresponding to k' eigenvalues can represent the whole data best.

## 2.2. Long Short-Term Memory Network (LSTM)

In this paper, an integrating model of long short-term memory network based on principal components analysis (PCA-LSTM) is explored to analyze the operational timeseries data generated from the dredging process. The proposed model is developed on the basis of long short-term memory network (LSTM), which is a special form of recurrent neural networks (RNN) that can address long-distance dependencies and delay in timeseries modeling.

The LSTM architecture was firstly proposed by Sepp Hochreiter and Jürgen Schmidhuber in 1997 [25]. There is a special memory cell unit added to the original hidden layer in classic RNN architecture. The cell state is controlled by three gates: Input gate  $I_t$ , Forget gate  $F_t$ , and Output gate  $O_t$ , as shown in Figure 1.



Figure 1. The architecture of classical LSTM.

The forget gate  $F_t$  decides which information needs to be kept and which can be forgotten. The information consists of the current input  $X_t$  and previous hidden state/short-term memory  $h_{t-1}$ .

$$F_t = \sigma(W_{Foreget} \cdot [h_{t-1}, X_t]) + bias_{Forget}$$
(5)

For every time step, the sigmoid function generates values between 0 and 1 that indicate whether the old information is necessary. 0 denotes forget, and 1 means keep.  $W_{Forget}$  is the weight matrix between forget gate and input gate. *bias*<sub>Forget</sub> is the connection bias.

The input gate decides what should be stored in the long-term memory in the new information. It works with the current input  $X_t$  and previous short memory  $h_{t-1}$  through two layers. In the first layer, the short-term memory and current input is passed through a sigmoid function that values from 0 (not important) to 1 (important):

$$\mathbf{i}_{t} = \sigma(\mathbf{W}_{Input} \cdot [\mathbf{h}_{t-1}, \mathbf{X}_{t}]) + \mathbf{bias}_{Input}$$
(6)

where  $W_{Input}$  is the weight matrix of sigmoid operator between input gate and output gate. *bias*<sub>Input</sub> is the bias vector.

The second layer uses the tanh function to regulate the network. The tanh operator creates a vector  $\tilde{C}_t$  with all the possible values between -1 and 1:

$$\widetilde{C}_{t} = \tanh(W_{Cell} \cdot [h_{t-1}, X_{t}]) + bias_{Cell}$$
(7)

where  $W_{cell}$  is the weight matrix of tanh operator between cell state information and network output. *bias*<sub>Cell</sub> is the bias vector.

With these two layers input, the cell state updates a new cell state (long-term memory):

$$\boldsymbol{C}_t = \boldsymbol{F}_t \odot \boldsymbol{C}_{t-1} + \boldsymbol{i}_t \odot \widetilde{\boldsymbol{C}}_t \tag{8}$$

where  $\odot$  is the Hadamard product.

When it comes to the output gate, the current input  $X_t$ , the previous short-term memory  $h_{t-1}$ , and the newly obtained cell state  $C_t$  determine the new short-term memory (hidden state) that will be passed on to the cell in the next time step.

$$O_t = \sigma(W_{output} \cdot [h_{t-1}, X_t]) + bias_{output}$$
(9)

$$h_t = O_t \odot \tanh(C_t) \tag{10}$$

 $W_{output}$  is the weight matrix of the output gate. This hidden state is used for prediction. Both new cell state and hidden state are carried over to the next time step.

#### 3. The Proposed PCA-LSTM Model

As described above, the basic knowledge regarding PCA and LSTM networks was introduced to set up the proposed PCA-LSTM model in this section. Considering mechanism and human experience, the PCA procedure will bring about a more accurate variable analysis for the practical multi-sensors system. The time-series data of effective variables are subsequently learned in LSTM network to output the target prediction.

## 3.1. PCA Based on Mechanism

The traditional PCA was introduced in Section 2.1. Due to the purely data-driven process, historical data is analyzed in PCA without any prior knowledge, which may bring about that some redundancies may be considered despite the causal relationship. Therefore, human experience will be introduced to interfere with the variable selection procedure ahead of PCA, based on the known mechanism.

The monitoring system always contains a broad range of sensors data related to the target object. Some of the data is on control variables, while some of the data is just on display variables that visualize the parameters.

Assuming the sensor system obtains an initial dataset:

$$\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{i-m}, \mathbf{x}_{i-m+1}, \dots, \mathbf{x}_i, \dots, \mathbf{x}_k)$$
(11)

where  $x_i$  represents the *i*-th sensor equipped in the system.

$$\mathbf{x}_i = [\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{ij}, \dots, \mathbf{x}_{in}]^T \tag{12}$$

where  $x_{ij}$  represents the *j*-th data obtained by *i*-th sensor.

When studying the working mechanism of the target, variables causal relationship will be analyzed and some of the redundant variables will be deleted, as well as some meaningless display parameters. This creates a new sample set:

$$\mathbf{X}' = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{i-m}, \mathbf{x}_i, \dots, \mathbf{x}_k) \tag{13}$$

PCA based on human experience method obtains a hyperplanar representation of all samples through recent reconstruction, realizing the dimension reduction from k to k' with the least loss.

The samples are centralized firstly as:

$$\sum_{i} x_{i} = 0 \tag{14}$$

Then a new coordinate system can be obtained after projection transformation:

$$W = (w_1, w_2, \dots, w_{i-m}, w_i, \dots, w_k) \tag{15}$$

where  $w_i$  is the standard orthogonal basis vector.

$$w_i \|_2 = 1$$
 (16)

$$\boldsymbol{w}_i^T \boldsymbol{w}_i = 0, (i \neq j) \tag{17}$$

If a portion of the coordinate is abandoned, namely the dimension is reduced from k to k' (k' < k), the projection of samples  $x_i$  in the low-dimensional coordinate system will be:

$$z_{ij} = (z_{i1}, z_{i2}, \dots, z_{in'}) \tag{18}$$

$$z_{ij} = \boldsymbol{w}_j^T \boldsymbol{x}_i \tag{19}$$

where  $z_{ij}$  is the *j*-th coordinate of  $x_i$  in low-dimensional space; and  $x_i$  can be reconstructed as:

$$\hat{\boldsymbol{x}}_i = \sum_{j=1}^{k'} \boldsymbol{z}_{ij} \boldsymbol{w}_j \tag{20}$$

For the whole training dataset, the distance between original samples  $x_i$  and the reconstructed samples  $\hat{x}_i$  can then be determined as:

$$\sum_{i=1}^{k} \left\| \sum_{j=1}^{n'} z_{ij} \boldsymbol{w}_j - \boldsymbol{x}_i \right\|_2^2 = \sum_{i=1}^{k} \boldsymbol{z}_i^T \boldsymbol{z}_i - 2 \sum_{i=1}^{k} \boldsymbol{z}_i^T \boldsymbol{W}^T \boldsymbol{x}_i + const$$

$$\propto -tr(\boldsymbol{W}^T(\sum_{i=1}^{k} \boldsymbol{x}_i \boldsymbol{x}_i^T) \boldsymbol{W})$$
(21)

where *const* is the constant item, and *W* can be obtained by Equation (15).

Because  $\sum_{i=1}^{k} x_i x_i^T$  is a covariance matrix, the target distance function can be minimized as:

$$\begin{cases} Min_W - tr(W^T X' X'^T W) \\ s.t. W^T W = I \end{cases}$$
(22)

where *I* is the identity matrix.

With the Langerin multiplier method [26], it can be calculated as:

$$\boldsymbol{X}'\boldsymbol{X}'^{T}\boldsymbol{w}_{i} = \lambda_{i}\boldsymbol{w}_{i} \tag{23}$$

After eigenvalue decomposition, the eigenvalue  $\lambda$  can be obtained as follows:

$$\lambda = \{\lambda_1, \lambda_2, \dots, \lambda_{i-m}, \lambda_i, \dots, \lambda_k\}$$
(24)

According to the practical demand, reconstruction threshold  $\mu$  is set to satisfy the condition:

$$\frac{\sum_{i=1}^{k'} \lambda_i}{\sum_{i=1}^{k} \lambda_i} \ge \mu \tag{25}$$

When the maximum threshold  $\mu$  is satisfied, the eigenvalues can be obtained in turn:

$$\lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_{k'} \tag{26}$$

And the eigenvectors corresponding to the first k' eigenvalues constitute the PCA solution:

$$W^* = (w_1, w_2, \dots, w_{i-m}, w_i, \dots, w_{k'})$$
(27)

The variables corresponding to the eigenvectors are:

$$X'' = (x_1, x_2, \dots, x_{i-m}, x_i, \dots, x_{k'})$$
(28)

Based on the variables obtained by PCA procedure above, the correlation matrix can be calculated as:

$$\boldsymbol{R} = (\boldsymbol{r}_{ij})_{k'*k'} \tag{29}$$

Then the most positively relevant variables to the target will be proceeded in the subsequent prediction model:

$$X^{**} = (x_1, x_2, \dots, x_p)$$
(30)

### 3.2. The Proposed Methodology

The variables most related to the target were obtained by PCA based on human experience that can be used as inputs in the next LSTM network to obtain prediction results. Namely, with current input being  $X_t^{**}$ , the current cell state and hidden state are updated as described in Section 2.2.

$$C_{t} = F_{t} \odot C_{t-1} + i_{t} \odot \widetilde{C}_{t}$$

$$= F_{t} \odot C_{t-1} + i_{t} \odot \{ \tanh(W_{Cell} \cdot [h_{t-1}, X_{t}^{**}]) + bias_{cell} \}$$

$$= \{ \sigma(W_{Forget} \cdot [h_{t-1}, X_{t}^{**}]) + bias_{Forget} \} \odot C_{t-1} + \{ \sigma(W_{Input} \cdot [h_{t-1}, X_{t}^{**}]) + bias_{Input} \} \odot \{ \tanh(W_{Cell} \cdot [h_{t-1}, X_{t}^{**}]) + bias_{Cell} \}$$

$$(31)$$

$$\begin{aligned} \mathbf{h}_t &= \mathbf{O}_t \odot \tanh(\mathbf{C}_t) \\ &= \left\{ \sigma(\mathbf{W}_{output} \cdot [\mathbf{h}_{t-1}, \mathbf{X}_t^{**}]) + \mathbf{bias}_{output} \right\} \odot \tanh(\mathbf{C}_t) \end{aligned}$$
(32)

Based on the new cell state and hidden state, we define gradient  $\delta_{h}^{(t)}$  and  $\delta_{c}^{(t)}$  to calculate the back propagation error layer by layer:

$$\delta_{h}^{(t)} = \frac{\partial L_{(t)}}{\partial h_{(t)}}$$
(33)

$$\delta_{c}^{(t)} = \frac{\partial L_{(t)}}{\partial C_{(t)}} \tag{34}$$

where L(t) is the loss function, and at the last sequence index  $\tau$ , the gradient can be described as follows:

$$\delta_{h}^{(\tau)} = \frac{\partial L_{(\tau)}}{\partial O_{(\tau)}} \frac{\partial O_{(\tau)}}{\partial h_{(\tau)}} = W_{Output}{}^{T}(\hat{O}_{(\tau)} - O_{(\tau)})$$
(35)

$$\delta_{\boldsymbol{c}}^{(\tau)} = \frac{\partial \boldsymbol{L}_{(\tau)}}{\partial \boldsymbol{h}_{(\tau)}} \frac{\partial \boldsymbol{h}_{(\tau)}}{\partial \boldsymbol{C}_{(\tau)}} = \delta_{\boldsymbol{h}}^{(\tau)} \odot \boldsymbol{O}_{(\tau)} \odot (1 - \tanh^2(\boldsymbol{C}_{(\tau)}))$$
(36)

Therefore, for any moment t,  $\delta_h^{(t+1)}$  and  $\delta_c^{(t+1)}$  can be obtained, deriving from  $\delta_h^{(t)}$  and  $\delta_c^{(t)}$  as follows:

$$\boldsymbol{\delta}_{\boldsymbol{h}}^{(t)} = \frac{\partial \boldsymbol{L}_{(t)}}{\partial \boldsymbol{O}_{(t)}} \frac{\partial \boldsymbol{O}_{(t)}}{\partial \boldsymbol{h}_{(t)}} + \frac{\partial \boldsymbol{L}_{(t+1)}}{\partial \boldsymbol{h}_{(t+1)}} \frac{\partial \boldsymbol{h}_{(t+1)}}{\partial \boldsymbol{h}_{(t)}} = \boldsymbol{W}_{Output}{}^{T}(\hat{\boldsymbol{O}}_{(t)} - \boldsymbol{O}_{(t)}) + \boldsymbol{W}^{T} \boldsymbol{\delta}_{\boldsymbol{h}}^{(t+1)} diag(1 - (\boldsymbol{h}_{(t+1)})^{2})$$
(37)

where *W* is the coefficient matrix.

Then the reverse gradient error of  $\delta_c^{(t)}$  can be obtained through the gradient error of the current layer returned from  $h_{(t)}$  and the gradient error of the previous layer  $\delta_c^{(t)}$ :

$$\boldsymbol{\delta}_{\boldsymbol{c}}^{(t)} = \frac{\partial \boldsymbol{L}_{(t)}}{\partial \boldsymbol{C}_{(t+1)}} \frac{\partial \boldsymbol{C}_{(t+1)}}{\partial \boldsymbol{C}_{(t)}} + \frac{\partial \boldsymbol{L}_{(t)}}{\partial \boldsymbol{h}_{(t)}} \frac{\partial \boldsymbol{h}_{(t)}}{\partial \boldsymbol{C}_{(t)}} = \boldsymbol{\delta}_{\boldsymbol{c}}^{(t+1)} \odot \boldsymbol{F}(t+1) + \boldsymbol{\delta}_{\boldsymbol{h}}^{(t)} \odot \boldsymbol{O}_{(t)} \odot (1 - \tanh^{2}(\boldsymbol{C}_{(t)}))$$
(38)

Then, the gradient of all parameters can be calculated easily using  $\delta_h^{(t)}$  and  $\delta_c^{(t)}$ , and all the parameters can be updated iteratively with the lowest error.

As mentioned above, the proposed method can be run in terms of Figure 2. It mainly consists of two parts: PCA and LSTM. The variables most related to the target are firstly obtained by PCA based on expert knowledge and used as inputs for LSTM network to get the prediction results.



Figure 2. The flowchart of proposed novel PCA-LSTM based on mechanism.

## 4. Case Study

Cutter Suction Dredger is a special kind of ship that is widely used in dredging engineering. In this section, the proposed method is validated in a real case study of well-equipped 4500 m<sup>3</sup>/h cutter suction dredger "Chang Shi 10" that serves in the Yangzi River region.

Mud and sand is cut to mix with water by a rotary cutter during the construction operation of the dredger. Meanwhile, the dredge pump works and creates vacuum pressure at the suction mouth of the cutter. Under the great pumping force, mud is sucked into the dredger pipeline and finally discharged to the dumping area. The primary system according to the dredging procedure was highlighted in Figure 3.



Figure 3. The primary system highlighted in cutter suction dredger (1:150).

## 4.1. Principal Components Analysis Based on Mechanism and Knowledge

During the construction, mud formation is influenced by many factors such as soil type, mechanical parameters and rotation speed of the cutter, traverse speed of the dredger, dredge pump parameters, and so on. To monitor and control the dredging process, up to 255 specific real-time sensors were arranged to collect the operational data [27]. Figure 4 shows some of the related monitoring parameters and relationship in automatic control system.



Figure 4. The monitoring parameters related to dredging process and productivity.

As shown in Figure 4, some of the parameters are control variables, while some of them are only display variables that make the data visualization.

Soil property is an important factor affecting the construction process and efficiency of cutter suction dredgers. For different solidity and water-solubility, the mud concentration is limited by cutting performance and silt mixing. The cutter structure, pipeline diameter, and the motor power of pumps are all specific variables (constant), which were demonstrated according to the demand for rated productivity at the beginning of the design. However, the cutter speed, trolley trip, cutter ladder movement, and dredge pump rotation are all control variables that can be adjusted during the operation process in specific construction. When digging hard soil, the dredging depth should be reduced while enhancing the cutter speed to prevent the formation of large diameter mud balls and pipe blocking. When the dredging soil is sediment or silt, the pump velocity should be appropriately increased to reduce the mud concentration and avoid sedimentation or clogging in the pipeline.

According to the actual sensor system of cutter suction dredger "Chang Shi 10", we firstly select 20 variables from the initial operational dataset as shown in Table 1.

Variable	Variable Description			
S <sub>8</sub>	Angle of the cutter ladder			
S <sub>9</sub>	Depth of the dredging	m		
S <sub>12</sub>	Rotation speed of the submersible pump	rpm		
S <sub>13</sub>	Rotation speed of the cutter	rpm		
S <sub>20</sub>	Flow	m <sup>3</sup> /h		
S <sub>23</sub>	Soil density	kg/m <sup>3</sup>		

Table 1. Initial variables from the operational dataset.

Variable	Description	Unit	-
S <sub>79</sub>	Distance of the swing movement	m	
S <sub>80</sub>	Angle of the swing		
S <sub>100</sub>	Rotation speed of the No.1 dredge pump	rpm	
S <sub>101</sub>	Rotation speed of the No.2 dredge pump	rpm	
S <sub>108</sub>	Power of the cutter	kw	
S <sub>164</sub>	Mud density	kg/m <sup>3</sup>	
S <sub>165</sub>	Flow rate	m/s	
S <sub>182</sub>	Trolley trip	m	
S <sub>198</sub>	Discharge pressure of the submersible pump	kPa	
S <sub>199</sub>	Discharge pressure of the No.1 dredge pump	kPa	
S <sub>200</sub>	Discharge pressure of the No.2 dredge pump	kPa	
S <sub>201</sub>	Vacuum	kPa	
S <sub>223</sub>	Water density	kg/m <sup>3</sup>	
S <sub>21</sub>	Mud concentration	- %	

Table 1. Cont.

Traditionally, the instantaneous productivity of the cutter suction dredger is the product of the flow and mud concentration.

$$P = C_m \cdot Q = C_m \cdot (v \cdot \pi r^2) \tag{39}$$

where  $C_m(\%)$  represent the mud concentration; Q is the flow amount per hour; and v is the flow rate.

As shown in Table 1, we choose  $S_{21}$  (mud concentration) as the target variable. In actual dredging construction process, the change of flow rate in the sludge pipeline is one of the important factors affecting the flow. Thus, we delete the redundant variable  $S_{20}$  flow in the first step.

Meanwhile, the mud concentration is determined by the density of soil, water, and mud.

$$C_m = \frac{\gamma_m - \gamma_w}{\gamma_s - \gamma_w} \tag{40}$$

where  $\gamma_m$  is the mud density;  $\gamma_w$  is the water density; and  $\gamma_s$  is the soil density.

Then we drop three redundant variables  $S_{223}$ ,  $S_{23}$ , and  $S_{164}$  in the second step.

For the study period in this case, the ship works with just No.1 dredge pump. Thus, the variables related to No.2 dredge pump are non-meaningful to the productivity. Namely,  $S_{101}$  and  $S_{200}$  are dropped according to human analysis and finally we obtain the related variable set as:

$$X' = \{S_8, S_{182}, S_{108}, S_{13}, S_9, S_{201}, S_{12}, S_{198}, S_{100}, S_{199}, S_{165}, S_{79}, S_{80}, S_{21}\}$$
(41)

As described in Section 3.1, the selected variables X' dataset based on human experience will be processed by PCA in this section. In addition, the contribution result is shown in Figure 5.

It is obvious that the top 10 principal components can represent more than 97% of the overall data. For the top two principal components, the dataset can be plot as Figure 6.



Figure 5. The percentage of explained variance for principal components.



Figure 6. Scatter plot of the dataset after PCA.

From the scatter plot of dataset in 2D figure, the correlation of the variables related to target  $S_{21}$  can be set as shown in Equation (28):

$$X'' = \{S_8, S_{182}, S_{108}, S_{13}, S_9, S_{201}, S_{12}, S_{198}, S_{100}, S_{199}, S_{165}, S_{79}, S_{80}\}$$
(42)

The most positively relevant variables to target can be further determined by the correlation matrix, as shown in Figure 7.

As the correlation matrix shows, the correlation between  $S_{21}$  and  $S_{199}$  is 0.48677, which means the discharge pressure of No.1 dredge pump affects the concentration most. It is consistent with the practical production. Pressure will influence the mud and water proportion that is pumped into pipeline. The variable  $S_{165}$  shows a correlation of 0.34628, which is also mentioned by other researchers [5,27]. The flow rate may determine the mud sedimentation during pipeline transportation. Furthermore, the vacuum correlation is 0.34152, since the vacuum gauge is installed on the upper part of the cutter, which is sensitive to the change of the mud concentration in the pipeline. Additionally, the angle of cutter ladder, the depth of dredging, and the trolley trip are all factors that affect the mud formation by operation controllers. However, for the discharge of the submersible pump, it is just the indirect factor to display the vacuum condition.

<b>S8</b>	- 1	0.1018	-0.2070	-0.6743	0.9729	0.0818	0.0211	0.2070	-0.3452	0.2087	-0.0954	-0.5597	-0.5522	0.3133 —	
S182	- 0.1018	1	-0.0376	-0.1018	0.0835	0.0456	-0.0080	0.0365	0.2925	0.2476	-0.0132	-0.3350	-0.3582	0.1646 —	- 0.8
S108	— -0.2070	-0.0376	1	0.2733	-0.2631	0.6775	-0.0965	-0.4976	0.0795	-0.1909	-0.1856	0.3116	0.3005	0.0557 —	
<b>S13</b>	0.6743	-0.1018	0.2733	1	-0.6864	0.0539	-0.0034	-0.0981	0.2395	-0.0368	-0.0077	0.4052	0.4035	-0.1581 —	- 0.6
<b>S9</b>	- 0.9729	0.0835	-0.2631	-0.6864	1	0.0158	0.0177	0.2441	-0.3526	0.2239	-0.0795	-0.5768	-0.5683	0.3091 —	- 0.4
S201	- 0.0818	0.0456	0.6775	0.0539	0.0158	1	-0.0549	-0.4979	0.0740	0.0558	-0.0615	0.1664	0.1531	0.3415 —	0.4
S12	- 0.0211	-0.0080	-0.0965	-0.0034	0.0176	-0.0549	1	0.0597	-0.0134	0.0106	0.0025	0.0048	0.0059	-0.0022 —	- 0.2
S198	- 0.207	0.0365	-0.4976	-0.0980	0.2440	-0.4978	0.0597	1	-0.1422	0.6496	0.3650	-0.1974	-0.1899	0.0588 —	
<b>S100</b>	0.3452	0.2925	0.0795	0.2394	-0.3525	0.0740	-0.0134	-0.1422	1	0.3104	0.4641	-0.0692	-0.0821	0.0019 —	- 0
S199	- 0.2087	0.2476	-0.1909	-0.0368	0.2239	0.0558	0.0106	0.6496	0.3104	1	0.6176	-0.2384	-0.2433	0.4868 —	0.2
S165	0.0954	-0.0132	-0.1856	-0.0077	-0.0795	-0.0615	0.0025	0.3650	0.4641	0.6176	1	-0.0185	-0.0223	0.3463 —	-0.2
<b>S79</b>	0.5597	-0.3350	0.3116	0.4052	-0.5768	0.1664	0.0048	-0.1974	-0.0692	-0.2384	-0.0185	1	0.9989	-0.1356 —	0.4
<b>S80</b>	0.5522	-0.3582	0.3005	0.4035	-0.5683	0.1531	0.0059	-0.1899	-0.0821	-0.2433	-0.0224	0.9989	1	-0.1400 —	
<b>S21</b>	- 0.3133	0.1646	0.0557	-0.1581	0.3091	0.3415	-0.0022	0.0588	0.0019	0.4868	0.3463	-0.1355	-0.1400	1 -	-0.6
	S.	5182	5108	513	3	5201	512	5198	5100	5199	5165	579	598	521	

Correlation matrix

Figure 7. The correlation matrix of the variables.

Meanwhile, it is indicated in the correlation matrix that there are five negative variables to the target. Thus, we obtain the final variables set most positively relevant to the target  $S_{21}$  as:

$$X^{**} = \{S_{182}, S_{108}, S_8, S_9, S_{201}, S_{198}, S_{100}, S_{199}, S_{165}\}$$
(43)

In general, the mud concentration is mainly inter-influenced by the dredge pump pressure, flow rate, vacuum, cutter ladder angle, dredging depth, and trolley trip.

## 4.2. Modeling Prediction Analysis

In this section, we choose the first segment series data and follow the steps given in Sections 2.2 and 3.2 to train the proposed model. This segment series data is collected from the monitoring system with a frequency of 100 sample points per minute. We intercept a dataset of 18,000 for 3-hour working time zone and obtain final 16,764 data after the pre-process.

## 4.2.1. Learning Results Analysis

In terms of the variable's selection process in Section 4.1, we use the most positively related nine parameters as input to predict the target output concentration  $C_m$  (%). Considering the data amount effects on the learning ability in data-driven models, we divide the input with a proportion of 6:4 and 7:3 to test the model twice. The learning results are shown in Figures 8 and 9 respectively.



Figure 8. The learning results of 6:4 proportion.







Figure 9. The learning results of 7:3 proportion.

Concentration changes with the working conditions. As shown in the learning results, the normal range of the concentration is from 0 to 45%, which is a comprehensive result of the multiple factors interaction. High concentration is not necessarily good for the production since it may cause sedimentation or clogging in the pipeline. The results in this case are all normal and satisfactory. However, when it comes to the detailed comparison, the learning process with 60% of the dataset performs better than 70% dataset. For the 60% data training, its maximum and minimum error are 0.3091 and 0.0149 respectively. However, the maximum error is 0.526 in the 70% data training process. Also, as shown in Figure 10, for 60% dataset, the loss value decreases and then keeps steady in the training process. The error in testing falls and then keeps steady. However, for the 70% dataset, both training and testing error are less stable and consistent.



Figure 10. Loss curve of the different datasets.

## 4.2.2. Cross Validation

Considering the necessary adaptability to dynamical changes, we use another dataset of 36,000 for 6-hour working time zone and obtain a final of 31,304 data samples for further cross validation to illustrate the proposed method's effectiveness and generality. The learning results are shown in Figure 11.



Figure 11. The learning result of cross validation.

It is obvious that the proposed method performs well in both training and testing processes. The average error in cross validation is 1.021%, which decreases since the data amount is becoming larger. In other words, data volume is essential for the deep learning method to function properly. This is just the advantage we explore in this novel model for prediction with operational "big data". Especially, the model can be updated by upcoming new data for more accurate results.

## 4.3. Comparative Study

This paper presents the novel PCA-LSTM method, which combines advantages of PCA and deep learning algorithm LSTM to manage big time-series data in operation monitoring systems. We compare the proposed method with other prediction methods including traditional PCA-LSTM and LSTM using the same dataset as Section 4.2 for further analysis. The results are shown in Figures 12 and 13.



Figure 12. The training result of comparative study.



Figure 13. The testing result of comparative study.

It is obvious in Figure 12 that the proposed method works better with a satisfactory error range. LSTM shows the maximum deviation because there is no variable selection before of the prediction process. Although it is a powerful tool to deal with the big series data for its special gate control function, it cannot give consideration to the variable analysis.

In Figure 13, it is easy to find that the novel PCA-LSTM performs better than both traditional PCA-LSTM and LSTM in the test. The yellow line in the figure marks the proposed PCA-LSTM, which has the lowest mean average error of 0.9213%. The green line marks the traditional PCA-LSTM, which shows a mean average error of 1.5301%. However, LSTM shows the worst result with a MAE (mean average error) of 2.0269%. The results differences are caused mainly because of the variable selection for the prediction model. As the input of the data-driven model, variable selection should be more focused with human knowledge and experience.

From a practical point of view, the comparative results are also analyzed by different evaluating indicators such as MAE (mean average error),  $R^2$  (coefficient of determination), and RMSE (root mean square error). As shown in Table 2, all of the models show a good performance with the coefficient of determination, which can explain the LSTM effectiveness. However, for the root mean square error, the proposed method takes on a better stability in the prediction result. The comparative results indicate that the control of input is essential to the machine learning methods.

Table 2. Results analysis of the comparative study.

	MAE	<b>R</b> <sup>2</sup>	RMSE
Proposed PCA-LSTM	0.0424	0.9999	0.0925
Traditional PCA-LSTM	0.3063	0.9863	0.4054
LSTM	0.3352	0.9828	0.5010

### 5. Conclusions

This paper proposes the novel PCA-LSTM method for the productivity prediction of cutter suction dredger, wherein the deep learning process makes good use of the real-time operational monitoring data. The PCA method based on mechanism and knowledge is proposed to analyze the multiple parameters and select relevant variables from the operation process. Then the results are used as input into LSTM model to obtain the target prediction. This approach is also successfully validated by comparison against other methods on a real-world case in China. The productivity of cutter suction dredger is influenced by many correlated factors such as the soil characteristics, cutter parameters, mud pump

performance, and pipeline layout. Thus, the mud concentration should be stabilized at a suitable value by comprehensive adjustment to improve its efficiency and productivity.

However, this is still a workable extension of deep learning application in the productivity prediction of cutter suction dredger. In the future, we will further construct the dynamical predictive models according to the changing working condition. When the operational parameters change dynamically under different conditions, the generated data should be classified into status space to study how the operation influences the dredging performance. Additionally, considering the sensors distance in the system, more factors on time-delay should be put up to improve the prediction accuracy.

**Author Contributions:** Conceptualization, K.Y.; methodology, K.Y.; software, K.Y.; validation, K.Y.; formal analysis, K.Y. and T.X.; data curation, K.Y., J.-L.Y. and B.W.; writing—original draft preparation, K.Y.; writing—review and editing, K.Y. and J.-L.Y.; visualization, K.Y.; supervision, S.-D.F.; funding acquisition, S.-D.F. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by National Natural Science Foundation of China under Grant: 51679178 and 52071240.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

#### References

- 1. Kawakatsu, H.; Watada, S. Seismic evidence for deep-water transportation in the mantle. Science 2007, 316, 1468–1471. [CrossRef]
- 2. Kuehl, S.; DeMaster, D.; Nittrouer, C. Nature of sediment accumulation on the Amazon continental shelf. *Cont. Shelf Res.* **1986**, *6*, 209–225. [CrossRef]
- Walsh, J.; Nittrouer, C. Contrasting styles of off-shelf sediment accumulation in New Guinea. *Mar. Geol.* 2003, 196, 105–125. [CrossRef]
- 4. Tang, H.; Wang, Q.; Bi, Z. Expert system for operation optimization and control of cutter suction dredger. *Expert Syst. Appl.* 2008, 34, 2180–2192. [CrossRef]
- Wang, B.; Fan, S.; Jiang, P.; Xing, T.; Fang, Z.; Wen, Q. Research on predicting the productivity of cutter suction dredgers based on data mining with model stacked generalization. *Ocean. Eng.* 2020, 217, 108001. [CrossRef]
- 6. Sierhuis, M.; Clancey, W.; Seah, C.; Trimble, J.; Sims, M. Modeling and simulation for mission operations work system design. *J. Manag. Inf. Syst.* **2003**, *19*, 85–128.
- Blazquez, C.; Adams, T.; Keillor, P. Optimization of mechanical dredging operations for sediment remediation. J. Waterw. Port Coast. Ocean. Eng.-ASCE 2001, 127, 229–307. [CrossRef]
- Lai, H.; Chang, K.; Lin, C. A Novel Method for Evaluating Dredging Productivity Using a Data Envelopment Analysis-Based Technique. *Math. Probl. Eng.* 2019, 2019, 5130835. [CrossRef]
- 9. Pei, H.; Hu, C.; Si, X.; Zhang, J.; Pang, Z.; Zhang, P. Review of machine learning based remaining useful life prediction methods for equipment. *J. Mech. Eng.* **2019**, *8*, 1–13. [CrossRef]
- 10. Wang, L.; Chen, X.; Wang, W. Research and analysis on construction output prediction of cutter suction dredger based on RBF neural network. *China Harb. Eng.* **2019**, *39*, 64–68.
- 11. Guan, F.; Wang, W. Application of extreme learning machines in productivity prediction of trailing suction hopper dredger. *Sci. Technol. Innov.* **2020**, *8*, 58–61.
- 12. Yang, J.; Ni, F.; Wei, C. Prediction of cutter suction dredger production based on double hidden layer BP neural network. *Comput. Digit. Eng.* **2016**, *44*, 1234–1237.
- 13. Ren, L.; Sun, Y.; Cui, J.; Zhang, L. Bearing remaining useful life prediction based on deep autoencoder and deep neural networks. *J. Manuf. Syst.* **2018**, *48*, 71–77. [CrossRef]
- 14. Wang, S.; Chen, Z.; Chen, S. Applicability of deep neural networks on production forecasting in bakken shale reservoirs. *J. Pet. Sci. Eng.* **2019**, *179*, 112–125. [CrossRef]
- 15. Xu, W.; Peng, H.; Tian, X.; Peng, X. DBN based SD-ARX model for nonlinear time series prediction and analysis. *Appl. Intell.* 2020, 50, 4586–4601. [CrossRef]
- Hu, C.; Pei, H.; Si, X.; Du, D.; Wang, X. A prognostic model based on DBN and diffusion process for degrading bearing. *IEEE Trans. Ind. Electron.* 2019, 67, 8767–8777. [CrossRef]
- 17. Deutsch, J.; He, D. Using deep learning-based approach to predict remaining useful life of rotating components. *IEEE Trans. Syst. Man Cybern. Syst.* **2017**, *48*, 11–20. [CrossRef]

- Zhang, C.; Lim, P.; Qin, A.K.; Tan, K.C. Multiobjective deep belief networks ensemble for remaining useful life estimation in prognostics. *IEEE Trans. Neural Netw. Learn. Syst.* 2016, 28, 2306–2318. [CrossRef] [PubMed]
- Wang, Y.; Zhang, Y.; Wu, Z.; Li, H.; Christofides, P. Operational Trend Prediction and Classification for Chemical Processes: A Novel Convolutional Neural Network Method Based on Symbolic Hierarchical Clustering. *Chem. Eng. Sci.* 2020, 225, 115796. [CrossRef]
- Chow, T.; Fang, Y. A recurrent neural-network-based real-time learning control strategy applying to nonlinear systems with unknown dynamics. *IEEE Trans. Ind. Electron.* 1998, 45, 151–161. [CrossRef]
- 21. Malhi, A.; Yan, R.; Gao, R. Prognosis of defect propagation based on recurrent neural networks. *IEEE Trans. Instrum. Meas.* 2011, 60, 703–711. [CrossRef]
- 22. Li, D.; Huang, D.; Yu, G.; Liu, Y. Learning Adaptive Semi-Supervised Multi-Output Soft-Sensors with Co-Training of Heterogeneous Models. *IEEE Access* 2020, *8*, 46493–46504. [CrossRef]
- 23. Yang, K.; Liu, Y.; Yao, Y.; Fan, S.; Ali, M. Operational time-series data modeling via LSTM network integrating principal component analysis based on human experience. *J. Manuf. Syst.* **2021**, in press. [CrossRef]
- D'Agostino, R.B. Principal Components Analysis. In Handbook of Disease Burdens and Quality of Life Measures; Preedy, V.R., Watson, R.R., Eds.; Springer: New York, NY, USA, 2020. [CrossRef]
- 25. Hochreiter, S.; Schmidhuber, J. Long short-term memory. Neural Comput. 1997, 9, 1735–1780. [CrossRef] [PubMed]
- 26. Jafari, H.; Alipoor, A. A New Method for Calculating General Lagrange Multiplier in the Variational Iteration Method. *Numer. Methods Partial. Differ. Equ.* **2011**, *27*, 996–1001. [CrossRef]
- Bai, S.; Li, M.; Kong, R.; Han, S.; Li, H.; Qin, L. Data mining approach to construction productivity prediction for cutter suction dredgers. *Autom. Constr.* 2019, 105, 102833. [CrossRef]